genomeA
Journals.ASM.org

Draft Genome Sequence of a 16SrII-A Subgroup Phytoplasma Associated with Purple Coneflower (*Echinacea purpurea*) Witches' Broom Disease in Taiwan

Shu-Heng Chang (張書恆),[a] Shu-Ting Cho (卓舒婷),[b] Chung-Li Chen (陳宗禮),[c] Jun-Yi Yang (楊俊逸),[a] Chih-Horng Kuo (郭志鴻)[b,d,e]

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

The bacterial genus "*Candidatus* Phytoplasma" contains a group of insect-transmitted plant pathogens in the class *Mollicutes*. Here, we report a draft genome assembly and annotation of strain NCHU2014, which belongs to the 16SrII-A subgroup within this genus and is associated with purple coneflower witches' broom disease in Taiwan.

Received 7 October 2015 **Accepted** 8 October 2015 **Published** 25 November 2015

Citation Chang S-H, Cho S-T, Chen C-L, Yang J-Y, Kuo C-H. 2015. Draft genome sequence of a 16SrII-A subgroup phytoplasma associated with purple coneflower (*Echinacea purpurea*) witches' broom disease in Taiwan. Genome Announc 3(6):e01398-15. doi:10.1128/genomeA.01398-15.

Address correspondence to Jun-Yi Yang, jyang@dragon.nchu.edu.tw, or Chih-Horng Kuo, chk@gate.sinica.edu.tw.

Phytoplasmas are a group of phloem-limited plant pathogens vectored by sap-feeding insects (1). Because of the difficulties involved in cultivating these bacteria outside of their hosts, genome sequencing and comparative analysis have been adopted as a major tool to study them (2, 3). The strain NCHU2014 was associated with purple coneflower (*Echinacea purpurea*) witches' broom disease in Taiwan (4). Based on its 16S rRNA gene sequence, this strain is most closely related to "*Candidatus* Phytoplasma australasiae" and has been assigned to the 16SrII-A subgroup within the genus. To facilitate future investigation on the biology of this bacterium, as well as to improve the taxon sampling of available phytoplasma sequences for comparative analyses, we report a draft genome assembly of this bacterium here.

The bacterial strain NCHU2014 was collected from naturally infected purple coneflower at the agricultural experiment station of National Chung Hsing University (Taichung, Taiwan) in June 2014. Subsequently, the strain was transferred to periwinkle (*Catharanthus roseus*) through dodder (*Cuscuta australis*) and maintained by grafting. Mature leaves from artificially infected periwinkle were collected for total DNA extraction by using the DNeasy plant minikit (Qiagen). The Illumina MiSeq platform was used to generate 300-bp reads from one paired-end library (~500 bp insert, 34,321,466 reads).

The procedures for genome assembly and annotation were based on those described in our previous studies (5, 6). Briefly, the initial *de novo* assembly was performed using Velvet version 1.2.10 (7). Putative phytoplasma contigs were identified by running BLASTx (8) searches against the NCBI nonredundant database (9). Additionally, the contigs were checked against the periwinkle chloroplast genome (10) to exclude possible contamination from the plant host. The draft genome of a closely related phytoplasma associated with peanut witches' broom (PnWB) disease (5) was used as a reference for scaffolding. PCR and Sanger sequencing were used for gap filling. For final verification, the Illumina raw reads were mapped to the assembly using BWA version 0.7.12 (11), programmatically checked using the mpileup program in SAMtools package version 1.2 (12), and visually inspected using IGV version 2.3.57 (13).

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 the PnWB phytoplasma (5) and aster yellows phytoplasma (17), as identified by OrthoMCL (18). Subsequent manual curation was based on BLASTp (8) searches against the NCBI nonredundant database (9) and the KEGG database (19, 20).

The first version of this draft genome contains 28 contigs with a combined size of 545,427 bp; the average G+C content is 23.9%. The annotation includes four rRNA genes, 26 tRNA genes, and 433 protein-coding genes.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number LKAC00000000. The version described in this paper is the first version, LKAC01000000.

ACKNOWLEDGMENTS

The funding for this project was provided by the Ministry of Science and Technology (MOST-103-2313-B-005-027) and the Institute of Plant and Microbial Biology at Academia Sinica, Taiwan. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The Illumina sequencing service was provided by Genomics BioSci & Tech Ltd. (New Taipei, Taiwan).

REFERENCES

1. **IRPCM Phytoplasma/Spiroplasma Working Team—Phytoplasma Taxonomy Group.** 2004. "*Candidatus* Phytoplasma", a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. Int J Syst Evol Microbiol **54:**1243–1255. http://dx.doi.org/10.1099/ijs.0.02854-0.
2. **Sugio A, Hogenhout SA.** 2012. The genome biology of phytoplasma: modulators of plants and insects. Curr Opin Microbiol **15:**247–254. http://dx.doi.org/10.1016/j.mib.2012.04.002.
3. **Kakizawa S, Yoneda Y.** 2015. The role of genome sequencing in phyto-

plasma research. Phytopathogenic Mollicutes **5:**19. http://dx.doi.org/10.5958/2249-4677.2015.00058.4.
4. **Tseng Y-W, Deng W-L, Chang C-J, Su C-C, Chen C-L, Jan F-J.** 2012. First report of a 16SrII-A subgroup phytoplasma associated with purple coneflower (*Echinacea purpurea*) witches'-broom disease in Taiwan. Plant Dis **96:**582. http://dx.doi.org/10.1094/PDIS-10-11-0888.
5. **Chung W-C, Chen L-L, Lo W-S, Lin C-P, Kuo C-H.** 2013. Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. PLoS One **8:**e62770. http://dx.doi.org/10.1371/journal.pone.0062770.
6. **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.
7. **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.
8. **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.
9. **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.
10. **Ku C, Chung W, Chen L, Kuo C.** 2013. The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in asterids. PLoS One **8:**e68518. http://dx.doi.org/10.1371/journal.pone.0068518.
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/Map format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
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, Rodland 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, 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. **Bai X, Zhang J, Ewing A, Miller SA, Jancso Radek A, Shevchenko DV, Tsukerman K, Walunas T, Lapidus A, Campbell JW, Hogenhout SA.** 2006. Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. J Bacteriol **188:**3682–3696. http://dx.doi.org/10.1128/JB.188.10.3682-3696.2006.
18. **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.
19. **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.
20. **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.