Microbiology® Resource Announcements

Complete Genome Sequence of "*Candidatus* Phytoplasma asteris" RP166, a Plant Pathogen Associated with Rapeseed Phyllody Disease in Poland

Shu-Ting Cho (卓舒婷),ᵃ Agnieszka Zwolińska,ᵇ Weijie Huang,ᶜ Roland H. M. Wouters,ᶜ Sam T. Mugford,ᶜ Saskia A. Hogenhout,ᶜ Chih-Horng Kuo (郭志鴻)ᵃ

ᵃInstitute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan
ᵇVirology and Bacteriology Department, Institute of Plant Protection, National Research Institute, Poznań, Poland
ᶜDepartment of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, Norfolk, United Kingdom

ABSTRACT The complete genome sequence of "*Candidatus* Phytoplasma asteris" RP166, which consists of one 829,546-bp circular chromosome, is presented in this work. This bacterium is associated with rapeseed phyllody disease in Poland and belongs to the 16SrI-B (i.e., aster yellows) group.

Phytoplasmas are plant pathogens transmitted by phloem-feeding insects from the order *Hemiptera*. Rapeseed (*Brassica napus* L.) crops are persistently threatened by "*Candidatus* Phytoplasma asteris" (aster yellows group, 16SrI-B subgroup), which causes phyllody diseases (1, 2). Phyllodies are leafy structures that develop in place of flowers and transform the infected plants into sterile "zombies" (i.e., plants that serve only for phytoplasma propagation). These dramatic morphological changes are induced by bacterial effectors (3). To better study the epidemiology of this disease, genomic resources are in great demand for the investigation of pathogenicity genes and the development of molecular markers.

Strain RP166 was collected from a naturally infected winter rapeseed plant at the Field Experimental Station of the Institute of Plant Protection–National Research Institute (Winna Góra, Poland; coordinates, 52.208921, 17.437842). Healthy *Macrosteles laevis* leafhoppers (*Cicadellidae*) were allowed to feed on the infected plant to acquire the bacteria. Next, the insects were maintained on healthy barley for several weeks to increase the phytoplasma titers prior to DNA extraction.

Two platforms were used for shotgun sequencing. For Illumina, a cetyltrimethylammonium bromide (CTAB) buffer protocol (4) was used for DNA extraction, followed by the use of a KAPA library preparation kit (catalog number KK8234) and Invitrogen SizeSelect gels (catalog number G6610-02) for ∼600-bp fragments. The MiSeq 2 × 300-bp paired-end sequencing (v3 chemistry) produced ∼10 Gb raw reads. For Oxford Nanopore Technologies (ONT), DNA was prepared using the Illustra Nucleon Phytopure kit (catalog number RPN8510) according to Wouters et al. (5). The library was prepared using the ONT ligation kit (catalog number SQK-LSK109) without shearing or size selection. The MinION run (R9.4 chemistry) produced 1,697,567 raw reads (∼5.9 Gb; N_{50}, 12,607 bp). Guppy v2.3.1 was used for base calling with a minimum quality score of 7; no further processing of the ONT reads was conducted.

The analysis procedure was modified from those described in our previous studies (6, 7). The Illumina reads were quality (Q) trimmed with a Q20 cutoff; reads shorter than 100 bp were discarded. *De novo* assembly was performed using Velvet v1.2.10 (8) (parameters: hash_length=91, scaffolding=no, exp_cov=30, cov_cutoff=5, max_coverage=500, min_contig_lgth=2000). Putative phytoplasma contigs were identified based on BLASTX (9) searches against a custom database of protein sequences from

available "*Ca*. Phytoplasma asteris" genomes (10). Then, the ONT reads were mapped to the contigs using minimap2 v2.15 (11) to produce a circular scaffold. Finally, an iterative process was used until the assembly was completed. In each iteration, the Illumina reads were mapped using BWA v0.7.12 (12), checked using SAMtools v1.2 (13), and inspected using IGV v2.3.57 (14). For gene prediction, RNAmmer (15), tRNAscan-SE (16), and Prodigal (17) were used. The annotation was based on the homologs in other phytoplasmas (10), as identified by OrthoMCL (18), followed by manual curation using BlastKOALA (19) and GenBank (20). The chromosome was rotated to have *dnaA* as the first gene.

Strain RP166 has one 829,546-bp circular chromosome with 27.7% G+C content; no plasmids were found. The Illumina and ONT reads provided 116× and 597× coverage, respectively. The annotation contains 6 rRNA genes, 32 tRNA genes, 753 protein-coding genes, and 69 pseudogenes.

Data availability. The raw reads have been deposited at the NCBI Sequence Read Archive under the accession numbers SRR12000858 and SRR12000859. The genome sequence has been deposited at GenBank/ENA/DDBJ under the accession number CP055264.

REFERENCES

1. Zwolińska A, Krawczyk K, Klejdysz T, Pospieszny H. 2011. First report of "Candidatus Phytoplasma asteris" associated with oilseed rape phyllody in Poland. Plant Dis 95:1475. https://doi.org/10.1094/PDIS-03-11-0177.
2. Tang Q, Fu Y, Ma M, Yao Y, Qu Z, Xie J, Cheng J, Jiang D. 2019. First report of phytoplasma groups 16SrI and 16SrV infecting *Brassica napus* in China. Crop Prot 126:104921. https://doi.org/10.1016/j.cropro.2019.104921.
3. MacLean AM, Sugio A, Makarova OV, Findlay KC, Grieve VM, Tóth R, Nicolaisen M, Hogenhout SA. 2011. Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in *Arabidopsis* plants. Plant Physiol 157:831–841. https://doi.org/10.1104/pp.111.181586.
4. Mugford S, Wouters R, Mathers TC, Hogenhout S. 2020. High quality DNA extraction from very small individual insects. Protocols https://doi.org/10.17504/protocols.io.bg6wjzfe.
5. Wouters R, Mugford S, Biello R, Heavens D, Hogenhout S. 2020. Extraction of high molecular weight DNA from aphids and other sap-feeding insects for long-read sequencing. Protocols https://doi.org/10.17504/protocols.io.bhftj3nn.
6. 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. https://doi.org/10.1371/journal.pone.0062770.
7. 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:e01398-15. https://doi.org/10.1128/genomeA.01398-15.
8. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
9. 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.
10. Cho S-T, Kung H-J, Huang W, Hogenhout SA, Kuo C-H. 2020. Species boundaries and molecular markers for the classification of 16SrI phyto-

plasmas inferred by genome analysis. Front Microbiol https://doi.org/10.3389/fmicb.2020.01531.
11. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
12. 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.
13. 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.
14. Robinson JT, Thorvaldsdottir 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.
15. 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. https://doi.org/10.1093/nar/gkm160.
16. 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.
17. 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. https://doi.org/10.1186/1471-2105-11-119.
18. Li L, Stoeckert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503.
19. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
20. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. 2018. GenBank. Nucleic Acids Res 46:D41–D47. https://doi.org/10.1093/nar/gkx1094.