Phytopathogenic Mollicutes
Vol. 13 (1), June 2023, 23-24

doi: 10.5958/2249-4677.2023.00012.9

Omics

Complete genome sequences of phytoplasma strains in group 16SrII associated with *Parthenium* phyllody in India

Kiran Kirdat[1,2], Bhavesh Tiwarekar[1], Shivaji Sathe[2] and Amit Yadav[1]

[1]National Centre for Cell Science, University of Pune Campus, Ganeshkhind, Pune 411007, India
[2]Department of Microbiology, Tuljaram Chaturchand College, Baramati, Maharashtra 413102, India

Abstract

It was performed the whole-genome sequencing of two phytoplasma strains in the 16SrII group, PR34 and PR08, associated with phyllody and witches' broom disease in the common weed, *Parthenium hysterophorus*. These phytoplasmas are associated with diseases in numerous pulse crops in India. Complete circular genomes were obtained after multiple sequencing attempts, employing DNA pre-processing and hybrid assemblies that combined short and long-read sequences. These genomes, deposited under GenBank accession numbers CP097206 and CP097207, offer valuable insights into the pathogenicity and evolutionary molecular characteristics of phytoplasma enclosed in the 16SrII group.

Keywords: Peanut witches' broom phytoplasma, 16SrII group, phyllody, genome sequence, *Parthenium*

Introduction

Phytoplasmas are phloem-inhabiting obligate plant pathogens linked with different diseases in legumes, horticultural crops, and weed species across the globe (Duduk *et al*., 2018). In India *'Candidatus* Phytoplasma aurantifolia' and *'Ca.* P. australasia', are associated with phyllody and witches' broom diseases in various economically important crops and their associated weeds (Thorat *et al*., 2016; 2017; Kirdat *et al*., 2020b). Weeds, in particular, serve as secondary hosts for phytoplasmas (Rao *et al*., 2017; Duduk *et al*., 2018). Given the significant economic impact of these phytoplasmas, the genomes of two strains (PR08 and PR34) detected in the common weed *Parthenium hysterophorus* were sequenced. The goal was to gain a better understanding of the genome structure of this pathogen and to shed light on the factors responsible for its pathogenicity. These newly obtained genome sequences are expected to enhance comprehension of the biology of 16SrII phytoplasmas and aid in developing effective strategies for managing the diseases associated with them.

Materials and Methods

P. hysterophorus plants exhibiting typical phyllody and witches' broom symptoms were collected from the Pune region of Maharashtra state of India. The presence of phytoplasmas was confirmed by PCR amplification followed by Sanger sequencing of the 16S rRNA gene from all samples (Deng and Hiruki, 1991; Kirdat *et al*., 2022). The identity of phytoplasmas strains PR34 and PR08 was confirmed using the EzBioCloud database (Yoon *et al*., 2017). The genomic DNA was extracted from infected leaf tissue using the CTAB method and enriched for prokaryotic DNA selection using the NEBNext microbiome enrichment kit (New England BioLabs, USA). The enriched DNA was sequenced on the Illumina NovaSeq 6000 and Oxford Nanopore Technology (ONT) MinION platform following the manufacturer's instructions and described previously (Ranebennur *et al*., 2022).

The bioinformatics pipelines described by Kirdat *et al.* (2020a) was used to generate hybrid genome assemblies of strains PR08 and PR34. Briefly, all QC-passed Illumina reads were subjected to metagenomic assembly using MEGAHIT v1.1.3 (Li *et al*., 2016), followed by taxonomic binning using MetaBAT2 v2.12.1 (Kang *et al*., 2015). The raw reads obtained from both platforms were mapped to all 16SrII genomes available on NCBI and phytoplasma-specific bin using Bowtie2 v2.3.5.1 (Langmead and Salzberg, 2012) and minimap2 v2.22-r110 (Li, 2018). These mapped reads were used to generate hybrid assembly in Unicycler v0.4.8 (Wick *et al*., 2017). The order and orientation of contigs were determined, and scaffolding was done by MeDuSa v1.6 (Bosi *et al*., 2015). The assemblies were curated for low coverage bases, submitted to the DDBJ/ENA/GenBank database, and underwent annotation with PGAP (Zhao *et al*., 2012). The OGRI values were obtained using the EzBioCloud orthoANI calculator (Yoon *et al*., 2017) and GGDC (Auch *et al*., 2010). The obtained genomes were characterized for the presence of various genome features, especially related to pathogenicity.

Corresponding author e-mail: Amit Yadav (amityadav@nccs.res.in)

Results and Discussion

The phytoplasma-specific reads were assigned to generate Unicycler assemblies for strains PR34 and PR08. The 16S rRNA gene fetched from these genomes found their closest match with reference sequences of *'Ca.* P. aurantifolia' (strain WBDL, GenBank accession number U15442) and *'Ca.* P. australasia' (strain *Carica papaya*, GenBank accession number Y10097), respectively. For PR34, six contigs spanning 614,946 bp were obtained in the first attempt, while PR08 generated a single circular genome sequence of 588,746 bp. All shorter contigs lacking significant similarity to phytoplasma sequences were removed. The PR34 and PR08 genomes were inspected manually and curated for low coverage bases. The annotated protein sequences of both genomes were verified for their association with phytoplasma using BLASTx searches. The MeDuSa oriented and scaffolded contigs of the strains PR34 and PR08 yielded a single circular chromosome of size 614,574 bp and 588,746 bp (GenBank accession numbers CP097206 and CP097207), respectively.

Table 1. Genome statistics of sequenced strains.

Strain ID	PR34	PR08
Contig	1	1
Genome length	614,574	588,746
Proteins	474	468
rRNAs	6	6
tRNAs	28	27
Coding density	70.61	72.74
%GC	24.65	24.36

The genome of strain PR34 has the smallest reported size among phytoplasmas in the 16SrII group, with a genome coverage of 5700X for Illumina reads and 180X for ONT. PR34 assembly had two rRNA operons, 474 protein-coding genes, 28 tRNA genes, 18 pseudogenes, and 24.65% G+C content. The genome coverage of strain PR08 for Illumina reads was 2750X, and for ONT, it was 70X. The final assembly showed 24.36% G+C content and included two rRNA operons, 27 tRNA genes, 468 protein-coding genes, and 15 pseudogenes (Table 1). The genome features of both 16SrII strains include the presence of a wide range of homologs of effector proteins, the absence of potential mobile units (PMUs), the presence of superoxide dismutase (SOD) gene, multiple copies of truncated hemolysin genes, and the presence of characteristic full-length and truncated group II introns.

Acknowledgments

The authors acknowledge the project funding and fellowships to K.K. and B.T. by the Department of Science and Technology grant number SERB/EEQ/2016/000752, the funding by Department of Biotechnology, grant number BT/COORD.II/01/03/2016 (NCMR) for in-house laboratory facilities, and the University Grant Commission for providing of CSIR-UGC NET-JRF fellowship to K.K. (Ref. No. 857/CSIR-UGC NET JUNE 2017).

References

Auch AF, von Jan M, Klenk HP and Göker M 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences*, 2(1): 117–134.

Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lió P, Crescenzi P, Fani R and Fondi M 2015. MeDuSa: a multi-draft-based scaffolder. *Bioinformatics*, 31(15): 2443–2451.

Deng S and Hiruki C 1991. Amplification of 16S rRNA genes from culturable and nonculturable mollicutes. *Journal of Microbiological Methods*, 14(1): 53–61.

Duduk B, Stepanovic J, Yadav A and Rao GP 2018. Phytoplasmas in weeds and wild plants. In: *Phytoplasmas: Plant Pathogenic Bacteria-I-Characterization and Epidemiology of Phytoplasma Associated diseases*, pp 313-345. Eds GP Rao, A Bertaccini, N Fiore and LW Liefting, Springer, Singapore.

Kang DD, Froula J, Egan R and Wang Z 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex *Microbial Communities*, 3: e1165.

Kirdat K, Tiwarekar B, Swetha P, Padma S, Thorat V, Manjula KM, Kavya N, Sundararaj R and Yadav A 2022. Nested real-time PCR assessment of vertical transmission of sandalwood spike phytoplasma ('*Ca.* Phytoplasma asteris'). *Biology*, 11(10): 1494.

Kirdat K, Tiwarekar B, Thorat V, Narawade N, Dhotre D, Sathe S, Shouche Y and Yadav A 2020a. Draft genome sequences of two phytoplasma strains associated with sugarcane grassy shoot (SCGS) and Bermudagrass white leaf (BGWL) diseases. *Molecular Plant-Microbe Interactions*, 33(5): 715–717.

Kirdat K, Tiwarekar B, Thorat V, Sathe S and Yadav A 2020b. First report of association of a 16SrII group phytoplasma with a witches' broom disease of *Croton bonplandianum*. *Phytopathogenic Mollicutes*, 10(1): 100-103.

Langmead B and Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4): 357.

Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H and Lam TW 2016. MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, 102: 3–11.

Li H 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18): 3094–3100.

Ranebennur H, Kirdat K, Tiwarekar B, Rawat K, Chalam VC, Solanke AU, Yadav R, Singh K,·Sathe S, Yadav A and Rao GP 2022. Draft genome sequence of '*Candidatus* Phytoplasma australasia', strain SS02 associated with sesame phyllody disease. *3 Biotech*, 12: 107.

Rao, G P, Madhupriya T, Manimekalai R, Tiwari AK and Yadav A 2017. A century progress of research on phytoplasma diseases in India. *Phytopathogenic Mollicutes*, 7(1): 1–38.

Thorat V, Kirdat K, Takawale P and Yadav A 2017. First report of 16SrII-D phytoplasmas associated with fodder crops in India. *Phytopathogenic Mollicutes*, 7(2): 106-110.

Thorat V, More V, Jadhav P, Mane SS, Nandanwar RS, Surayavanshi M, Shouche Y and Yadav A 2016. First report of a 16SrII-D group phytoplasma associated with witches' broom disease of soybean (*Glycine max*) in Maharashtra, India. *Plant Disease*, 100(12): 2521.

Wick RR, Judd LM, Gorrie CL and Holt KE 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *Plos Computational Biology*, 13(6): e1005595.

Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H and Chun J 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology*, 67(5): 1613.

Zhao Y, Wu J, Yang J, Sun S, Xiao J and Yu J 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics*, 28(3): 416–418.