Draft Genome Sequence of ‘*Candidatus* Phytoplasma pini’-Related Strain MDPP: A Resource for Comparative Genomics of Gymnosperm-Infecting Phytoplasmas

Weili Cai,^{1,2} Jonathan Shao,³ Yan Zhao,³ Robert E. Davis,³ and Stefano Costanzo^{1,†}

¹ United States Department of Agriculture (USDA), Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Science and Technology, Beltsville, MD, U.S.A.

² Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, U.S.A.

³ Molecular Plant Pathology Laboratory, USDA, Beltsville, MD, U.S.A.

Abstract

‘*Candidatus* Phytoplasma pini’-related strain MDPP, the reference strain of subgroup 16SrXXI-B, is a pathogen associated with witches’ broom disease of *Pinus* spp. in North America. Here, we report the first draft genome sequence of ‘*Ca. Phytoplasma pini*’ strain MDPP, which consists of 474,136 bases, with a G + C content of 22.22%. This information will facilitate comparative genomics of gymnosperm-infecting phytoplasmas.

Phytoplasmas are bacterial plant pathogens that lack a cell wall and inhabit plant phloem tissue. Phytoplasmas cause a wide variety of symptoms such as yellowing, stunting, witches’ brooms, phyllody, virescence, and little leaves. The mode of transmission for phytoplasmas from plant to plant is by insect vectors (Weintraub and Beanland 2006). Currently, more than 1,000 plant species have been described as host plants of phytoplasmas, including many agriculturally important crops (Bertaccini et al. 2014; Seemüller et al. 2013). Due to the lack of in vitro culture, available genome sequences of phytoplasmas are limited. Currently, the molecular characterization and classification is mainly based on restriction fragment length polymorphism analysis of the 16S ribosomal RNA (rRNA) sequence (IRPCM 2004; Lee et al. 1998; Wei et al. 2007). In April 2015, we identified a new ‘*Candidatus* Phytoplasma pini’-related strain from a pine tree located in a wooded area in Laurel, MD (Costanzo et al. 2016). This pine tree had excessive shoot branching, resulting in a witches’-broom phenotype. The strain was designated MDPP. Here, we report the first draft genome sequence of the strain MDPP, providing a resource for comparative genomics of gymnosperm-infecting phytoplasmas present in Europe and North America.

Genomic DNA of strain MDPP was extracted from symptomatic pine needles using the DNeasy plant minikit (Qiagen, Valencia, CA, U.S.A.). The pathogen titer level was estimated using 23S rRNA-based real-time quantitative PCR, giving a threshold cycle value of 18.66 per 2 µl of total infected plant DNA (Hodgetts et al. 2009). Bacterial DNA was enriched using the NEBNext microbiome DNA enrichment kit (New England Biolabs., Ipswich, MA, U.S.A.) that was further amplified using the REPLI-g Mini Kit (Qiagen). The 2 × 300-bp Illumina paired-end sequencing was performed on an Illumina MiSeq platform (Illumina, San Diego, CA, U.S.A.). The paired-end sequencing libraries were prepared using an Illumina TruSeq PCR free DNA library prep kit. The DNA was fragmented using a Covaris M220 with an average size of 550 bp.

[†]Corresponding author: S. Costanzo; stefano.costanzo@usda.gov

The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

The author(s) declare no conflict of interest.

Accepted for publication 13 December 2019.

Funding

This research was supported by the USDA Animal and Plant Health Inspection Service.

Keywords

genome, gymnosperm, phytoplasmas, pine

In total, 5.17×10^7 reads with a mean of 269 bp/read were generated from MiSeq sequencing. These paired-end reads were aligned using bowtie2 (version 2.2.5) (Langmead and Salzberg 2012) to the *Pinus taeda* genome (loblolly pine) APFE000000000.3 and *P. lambertiana* genome (sugar pine) LMTP000000000.1. The resulting unmapped reads (31,810,002) were merged and used in de novo assembly using the SPAdes assembler (version 3.11.0) (Nurk et al. 2013) with a kmer of 51. The resulting contigs were screened for phytoplasma origin using blastx against nr and against 9 complete and 23 draft phytoplasma genomes available from the genomes database of NCBI. This generated 16 contigs that were well aligned with known phytoplasmas. The contigs ranged from 6,925 to 110,223 bp ($N_{50} = 40,003$), corresponding to 55,309 reads of phytoplasma origin. The contigs comprised a total length of 474,136 bp, with a G + C content of 22.22% and an average coverage of 28x. Gene prediction was performed by GeneMark.hmm (version 3.05) (Besemer et al. 2001) using onion yellows phytoplasma as a model (GenBank accession number AP006628). The genome included 406 predicted open reading frames, 20 tRNA, and 3 rRNA coding genes (16S, 23S, and 5S rRNA).

The draft genome sequence has been deposited under BioProject PRJNA474201 and BioSample SAMN09289773. This Whole Genome Shotgun project has been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the accession VIAE000000000. The version described in this article is version VIAE01000000. Data has also been deposited in the Sequence Read Archive (SRA) database, under SRA accession SRR9643267.

Literature Cited

- Bertaccini, A., Duduk, B., Paltrinieri, S., and Contaldo, N. 2014. Phytoplasmas and phytoplasma diseases: A severe threat to agriculture. *Am. J. Plant Sci.* 5: 1763-1788.
- Besemer, J., Lomsadze, A., and Borodovsky, M. 2001. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29:2607-2618.
- Costanzo, S., Rascoe, J., Zhao, Y., Davis, R. E., and Nakhla, M. K. 2016. First report of a new '*Candidatus* Phytoplasma pini'-related strain associated with witches'-broom of *Pinus* spp. in Maryland. *Plant Dis.* 100:1776.
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., Pribelski, A. D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J., Lasken, R., Clingenpeel, S. R., Woyke, T., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads. Pages 158-170 in: *Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science*, vol 7821. M. Deng, R. Jiang, F. Sun, and X. Zhang, eds. Springer, Berlin, Heidelberg.
- Hodgetts, J., Boonham, N., Mumford, R., and Dickinson, M. 2009. Panel of 23S rRNA gene-based real-time PCR Assays for improved universal and group-specific detection of phytoplasmas. *Appl. Environ. Microbiol.* 75: 2945-2950.
- IRPCM. 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.
- Langmead, B., and Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- Lee, I. M., Gundersen-Rindal, D. E., Davis, R. E., and Bartoszyk, I. 1998. Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. *Int. J. Syst. Bacteriol.* 48:1153-1169.
- Seemüller, E., Sule, S., Kube, M., Jelkmann, W., and Schneider, B. 2013. The AAA+ ATPases and HflB/FtsH proteases of '*Candidatus* phytoplasma mali': Phylogenetic diversity, membrane topology, and relationship to strain virulence. *Mol. Plant-Microbe Interact.* 26:367-376.
- Wei, W., Davis, R. E., Lee, I. M., and Zhao, Y. 2007. Computer-simulated RFLP analysis of 16S rRNA genes: Identification of ten new phytoplasma groups. *Int. J. Syst. Evol. Microbiol.* 57:1855-1867.
- Weintraub, P. G., and Beanland, L. 2006. Insect vectors of phytoplasma. *Annu. Rev. Entomol.* 51:91-111.