PhytoFrontiers™
An Open Access Journal from The American Phytopathological Society

Resource Announcement

A Draft Genome Resource for '*Candidatus* Phytoplasma prunorum', the Agent Associated with European Stone Fruit Yellows Disease

Jose Pedro Fonseca[1,2] | Schyler Nunziata[1] | Jarred Yasuhara-Bell[1] | Assunta Bertaccini[3] | Yazmín Rivera[1] | Eric Newberry[1,†]

[1] Plant Pathogen Confirmatory Diagnostics Laboratory (PPCDL), Science and Technology, Plant Protection and Quarantine, Animal Plant Health Inspection Service, United States Department of Agriculture, Laurel, MD 20708, U.S.A.

[2] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh NC 27695, U.S.A.

[3] Department of Agricultural and Food Sciences, Alma Mater Studiorum – University of Bologna, Bologna, Italy

Accepted for publication 5 September 2024.

† Corresponding author: E. Newberry; eric.newberry@usda.gov

Author contributions: J.P.F., J.Y.-B., and E.N. conceptualized the work; J.P.F. wrote the original draft; E.N., J.Y.-B., A.B., and Y.R. reviewed and edited the manuscript; S.N. did the bioinformatic analysis; J.P.F. prepared Illumina libraries and performed sequencing; and J.P.F. and E.N. analyzed the data.

Data availability: Sequencing data for this project can be found at the National Institutes of Health (NIH) genetic sequence database of the National Center for Biotechnology Information (NCBI)-GenBank (https://www.ncbi.nlm.nih.gov/nucleotide/) under accession numbers SAMN39606366 (SRA accession SRR29553188) and SAMN39606365 (SRA accession SRR29553187) and BioProject accession number PRJNA1068968.

Funding: This research was supported by the intramural research program of the United States Department of Agriculture (USDA), Animal and Plant Health Inspection Service (APHIS).

The author(s) declare no conflict of interest.

Abstract

'*Candidatus* Phytoplasma prunorum' is associated with European stone fruit yellows disease, affecting wild and cultivated species of *Prunus* at different degrees of susceptibility, and so far is being mainly restricted to Europe. Here, we report draft genome sequences for '*Ca.* Phytoplasma prunorum' strains ESFY1 and LNS1, which represent the first available for this species. Strain ESFY1 is the causal agent of the European stone fruit yellows disease of *P. persica* in Germany, and LNS1 is the causal agent of leptonecrosis disease of *P. salicina* in Italy. Their draft genomes consist of 480,123 and 490,930 bases assembled into 28 and 16 contigs, with an average GC content of 21.1 and 20.7%, respectively. Analysis of average nucleotide identity (ANI) values between genomes further supported the delineation of '*Ca.* P. prunorum' as a species distinct from '*Ca.* P. mali', in agreement with previous phylogenetic data.

Keywords: genome, mollicutes, pathogen, phytoplasma, stone fruit yellows

'*Candidatus* Phytoplasma spp.' are associated with devastating diseases on a wide host range of monocot and dicot plants worldwide, including numerous economically important crops. They belong to the class Mollicutes, with a small genome size (averaging around 750 kb), low GC content (21 to 28%), and no cell wall. Phytoplasmas dwell in the plant phloem, are mainly insect-transmitted, and have not been obtained in pure culture (Contaldo et al. 2016). This has led their identification and taxonomy to rely on genetic relatedness as inferred primarily through restriction length polymorphism (RFLP) and sequence analysis of the 16S rRNA gene, as well as other barcoding loci (Bertaccini et al. 2022; IRPCM 2004; Lee et al. 1998).

'*Ca.* P. prunorum' is the agent associated with several economically important diseases of stone fruit trees in Europe such as peach yellows, apricot chlorotic leaf roll, and plum leptonecrosis amongst others and is referred to commonly as European stone fruit yellows (ESFY) (Danet et al. 2011; Sabaté et al. 2016). '*Ca.* P. prunorum' is a member of the apple proliferation group or 16SrX (subgroup B; Lee et al. 1998) and is closely related to 16SrX members '*Ca.* P. mali' and '*Ca.* P. pyri', which are responsible for apple proliferation disease (AP) and pear decline (PD), respectively (Seemüller and Schneider 2004). Strain ESFY1 was obtained through dodder-transmitted periwinkle from a symptomatic peach tree (*Prunus persica* var. *persica* 'South Haven') grown in

Germany, while strain LNS1 was obtained similarly from a symptomatic plum tree (*P. salicina* cultivar Ozark Premier) grown in the Udine region of Italy (Loi et al. 1995).

The genomic DNA used for sequencing was extracted according to the method described by Prince et al. (1993), treated with the New England Biolab's Microbiome Enrichment Kit (NEB, Ipswich, MA) to increase the sequencing depth of microbial DNA, and purified using the Zymo Genomic DNA Clean and Concentrator Kit (Zymo Research, Irvine, CA). Libraries were prepared using the Illumina DNA Prep Kit with IDT Nextera DNA UD Indexes Set A (Illumina, San Diego, CA). All libraries were quantified using the Qubit fluorometer (Thermo Fisher, Milan, Italy) and quality-verified for the presence of a single amplicon band using an Agilent 4200 TapeStation (Agilent, Santa Clara, CA). Libraries were sequenced using a NextSeq 2000 instrument with the P1 300-cycle Reagent Kit (2- × 150-bp read length) (Illumina, San Diego, CA).

After sequencing, reads obtained from pooled libraries were demultiplexed using the onboard NextSeq 2000 bcl2fastq2 v2.20 software, and the de novo genome assembly of draft genomes was performed in two iterations. Briefly, Illumina reads were trimmed with Trimmomatic v0.39 (Bolger et al. 2014) and used as input for an initial de novo assembly using Velvet Optimizer (Zerbino 2010), which automatically determines the *k*-mer length using a range of 53 to 99. To identify putative phytoplasma contigs in the initial draft assembly, all contigs were queried against the National Center for Biotechnology Information (NCBI) nonredundant protein database using BLASTx v2.14.0 (Camacho et al. 2009). All contigs mapping to phytoplasma proteins with an e-value cutoff of $1e^{-5}$ were extracted as putative phytoplasma contigs. Putative plant contigs were identified and removed based on BLASTn searches against the NCBI standard nucleotide collection database. In the second iteration of the assembly, the remaining putative phytoplasma contigs from the initial assembly were used as the reference to extract phytoplasma reads using bowtie2 (Langmead and Salzberg 2012). The extracted reads were used as input for de novo assembly with the CLC Genomics Workbench v23.0.2 using default parameters and a minimum contig length of 200 bp. The resulting contigs were confirmed as phytoplasma by BLASTx searches of the NCBI nonredundant protein database and BLASTn searches against the NCBI-nr database as well as a local database of phytoplasma genomes. The 16S rRNA gene sequences were identified and extracted from initial assemblies using Barrnap v.0.9 (Seemann 2018).

Finally, the assembled genomes were processed to remove contaminant sequences using Kraken2 in a similar manner to Lu and Salzberg (2018). The genomes were first split into 100-bp overlapping pseudo-reads, with each pseudo-read beginning every 50 bp. These reads were compared with a database with RefSeq archaea, bacteria, viral, plasmid, human, vector, protozoa, fungi, and plant. Any pseudo-read that was classified as eukaryote or with greater than 1% classified as nonphytoplasma taxon was masked by replacing the sequence with 'Ns'. Contigs with nonmasked bases <100 bp in length were removed from the final assembly. The completeness of the draft genome was estimated by BUSCO version 4.0.6 (Seppey et al. 2019) and the mollicutes database, consisting of 151 universal single-copy orthologs. Annotation was performed by Prokka v1.14.5 (Seemann 2014). The sequencing and assembly statistics for the draft genomes generated in this study are presented in Table 1.

Average nucleotide identity (ANI) values were calculated with the JSpeciesWS webserver (Richter et al. 2016) using the ANIm algorithm between phytoplasma strains ESFY1 and LNS1, revealing that the two phytoplasmas were closely related with a pairwise ANI score of 99.68%. On the other hand, a comparison between the '*Ca*. P. prunorum' ESFY1 and LNS1 genomes with the closely related '*Ca*. P. mali' (BioSample ID SAMEA2272373) displayed a similarity of 91.12%, which is below the 95% ANI species threshold recommended by the revised guidelines for phytoplasma species delineation (Bertaccini et al. 2022). It is important to note that the difference in genomic identity observed between the released draft genomes in this work ('*Ca*. P. prunorum') and '*Ca*. P. mali' agrees with previous phylogenetic and serological data supporting the distinct circumscription of phytoplasmas from AP, PD, and ESFY as distinct species (Seemüller and Schneider 2004).

Literature Cited

Bertaccini, A., Arocha-Rosete, Y., Contaldo, N., Duduk, B., Fiore, N., Montano, H. G., Kube, M., Kuo, C.-H., Martini, M., Oshima, K., Quaglino, F., Schneider, B., Wei, W., and Zamorano, A. 2022. Revision of the '*Candidatus* Phytoplasma' species description guidelines. Int. J. Syst. Evol. Microbiol. 72:005353.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: Architecture and applications. BMC Bioinf. 10:421.

Contaldo, N., Satta, E., Zambon, Y., Paltrinieri, S., and Bertaccini, A. 2016. Development and evaluation of different complex media for phytoplasma isolation and growth. J. Microbiol. Methods 127:105-110.

Danet, J. L., Balakishiyeva, G., Cimerman, A., Sauvion, N., Marie-Jeanne, V., Labonne, G., Laviña, A., Batlle, A., Križanac, I., Škorić, D., Ermacora, P., Serçe, Ç. U., Çağlayan, K., Jarausch, W., and Foissac, X. 2011. Multilocus sequence analysis reveals the genetic diversity of European fruit tree phytoplasmas and supports the existence of inter-species recombination. Microbiology 157:438-450.

IRPCM. 2004. '*Candidatus* Phytoplasma', a taxon for the wall-less, nonhelical prokaryotes that colonize plant phloem and insects. Int. J. Syst. Evol. Microbiol. 54:1243-1255.

Langmead, B., and Salzberg, S. L. 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. M. 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.

Loi, N., Carraro, L., Musetti, R., Pertot, I., and Osler, R. 1995. Dodder transmission of two different MLOs from plum trees affected by "leptonecrosis". Acta Hortic. 386:465-470.

Lu, J., and Salzberg, S. L. 2018. Removing contaminants from databases of draft genomes. PLoS Comput. Biol. 14:e1006277.

Prince, J. P., Davis, R. E., Wolf, T. K., Lee, I.-M., Mogen, B. D., Dally, E. L., Bertaccini, A., Credi, R., and Barba, M. 1993. Molecular detection of diverse mycoplasma like organisms (MLOs) associated with grapevine yellows and their classification with aster yellows, X-disease, and elm yellows MLOs. Phytopathology 83:1130-1137.

TABLE 1

Summary of the draft genomes from '*Candidatus* Phytoplasma prunorum' strains sequenced in this work

Strain	Reads	Contigs	Assembly size (bp)	N50	GC (%)	BUSCO completeness (%)	CDS	tRNAs	Depth (×)	rRNAs	Genome accessions	Assembly
ESFY1	25,815,990	28	480,123	30,457	21.17	89.4	431	26	721	2	JAZHCZ000000000	GCA_036924395.1
LNS1	41,561,855	16	490,930	112,565	20.72	91.4	428	21	210	2	JAZHCY000000000	GCA_036924415.1

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. 2016. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929-931.

Sabaté, J., Laviña, A., and Batlle, A. 2016. Incidence and distribution of '*Candidatus* Phytoplasma prunorum' and its vector *Cacopsylla pruni* in Spain: An approach to the epidemiology of the disease and the role of wild *Prunus*. Plant Pathol. 65:837-846.

Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.

Seemann, T. 2018. Barrnap 0.9: Rapid ribosomal RNA prediction. https://github.com/tseemann/barrnap

Seemüller, E., and Schneider, B. 2004. '*Candidatus* Phytoplasma mali', '*Candidatus* Phytoplasma pyri' and '*Candidatus* Phytoplasma prunorum', the causal agents of apple proliferation, pear decline and European stone fruit yellows, respectively. Int. J. Syst. Evol. Microbiol. 54:1217-1226.

Seppey, M., Manni, M., and Zdobnov, E. M. 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962:227-245.

Zerbino, D. R. 2010. Using the Velvet *de novo* assembler for short-read sequencing technologies. Curr. Protoc. Bioinformatics 31:11.5.1-11.5.12.