



# Draft Genome Sequence of a “*Candidatus* Phytoplasma asteris”-Related Strain (Aster Yellows, Subgroup 16Srl-B) from South Africa

Beatrix Coetzee,<sup>a</sup> Nicoleen Douglas-Smit,<sup>b</sup> Hans J. Maree,<sup>a,c</sup> Johan T. Burger,<sup>a</sup> Kerstin Krüger,<sup>b</sup> Gerhard Pietersen<sup>a,d</sup>

<sup>a</sup>Department of Genetics, Stellenbosch University, Stellenbosch, South Africa

<sup>b</sup>Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa

<sup>c</sup>Citrus Research International, Stellenbosch, South Africa

<sup>d</sup>Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

**ABSTRACT** Here, we report the draft genome sequence of a phytoplasma discovered in grapevine. The genome size is 600,116 nucleotides (nt), with 597 predicted open reading frames. It is most similar to a maize bushy stunt phytoplasma of group 16Srl-B (aster yellows). The possible presence of a 3,833-nt plasmid was also noted.

Aster yellows phytoplasma of the 16Srl-B group (1, 2) was first reported in South Africa associated with grapevine yellows disease (3), and the leafhopper *Mgenia fuscovaria* was identified as the vector (4).

Healthy *Catharanthus roseus* plants were placed in a phytoplasma-infected *Vitis vinifera* cv. Colombar vineyard in Vredendal, South Africa, for natural transmission. After 17 weeks, samples were collected from *C. roseus*, diseased *Vitis vinifera*, and *M. fuscovaria* leafhoppers captured in the same vineyard. DNA was extracted from each using a cetyltrimethylammonium bromide (CTAB) method (5). Next-generation sequencing libraries were prepared with TruSeq v.3 and sequenced on the Illumina HiScanSQ platform (Agricultural Research Council, Pretoria, South Africa) to produce paired-end 2 × 100-nucleotide (nt) reads.

Sequencing of the *C. roseus* sample resulted in 15,270,318 read pairs. Read pairs were merged with PEAR v.0.9.8 (6). Quality trimming was performed on merged reads with Trimmomatic v.0.36 (7). Adaptors were removed (ILLUMINACLIP:3:30:10 parameter), 3' and 5' nucleotides were removed if they had a Phred quality score below 20 (LEADING:20 and TRAILING:20 parameters), and only reads with a minimum length of 20 nt were retained (MINLEN:20 parameter).

Quality-trimmed reads were aligned to the *C. roseus* genome (GenBank assembly accession number [GCA\\_000949345](https://.ncbi.nlm.nih.gov/assembly/GCA_000949345)) with Bowtie 2 v.2.2.8 (8). Retained unmapped reads were assembled into contigs with SPAdes v.3.13.0 (9), including the reference sequences of maize bushy stunt (GenBank accession number [CP015149](https://ncbi.nlm.nih.gov/assembly/CP015149)) and onion yellows (GenBank accession number [AP006628](https://ncbi.nlm.nih.gov/assembly/AP006628)) (the complete genome sequences of group 16Srl-B), to guide the assembly (with the SPAdes “untrusted contigs” parameter). Assembled contigs were compared to the NCBI nucleotide database with BLAST v.2.4.0 (10). Only contigs with hits to phytoplasmas were retained. The unmapped reads were again assembled in SPAdes, using these contigs as reference sequences. This process was repeated 10 times, each time retaining only contigs with BLAST hits to phytoplasmas. This yielded two contigs of 600,116 nt (28.4% GC content) and 3,833 nt (25.7% GC content) in length, representing the phytoplasma genome and a putative plasmid, respectively.

Using Prokka v.1.12 (11), 561 protein-coding sequences, 32 tRNAs, and four rRNAs were identified in the genome. Two 16S rRNA-encoding genes were identified, and

**Citation** Coetzee B, Douglas-Smit N, Maree HJ, Burger JT, Krüger K, Pietersen G. 2019. Draft genome sequence of a “*Candidatus* Phytoplasma asteris”-related strain (aster yellows, subgroup 16Srl-B) from South Africa. *Microbiol Resour Annot* 8:e00148-19. <https://doi.org/10.1128/MRA.00148-19>.

**Editor** Irene L. G. Newton, Indiana University, Bloomington

**Copyright** © 2019 Coetzee et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Beatrix Coetzee, [beatrix@sun.ac.za](mailto:beatrix@sun.ac.za).

**Received** 25 February 2019

**Accepted** 28 March 2019

**Published** 25 April 2019

their restriction digestion patterns that were generated with iPhyClassifier (12) confirmed they were in the 16Srl-B group.

This phytoplasma most closely resembles maize bushy stunt, based on genome organization and sequence identity. Using a minimum amino acid identity of 95% over at least 95% of the length, 393 (70%) of the 561 coding sequences aligned to maize bushy stunt.

The putative plasmid has five coding sequences, including a replication-associated protein with a geminivirus replication catalytic domain AL1 and a single-stranded DNA binding protein. Both these proteins share the highest amino acid identity with proteins from the rice orange leaf phytoplasma plasmid (GenBank accession numbers [ATL14544](#) and [ATL14548](#), with 78% and 96% shared amino acid identity, respectively).

Reads from the *V. vinifera* and *M. fuscovaria* samples were treated in the same manner and aligned to the generated phytoplasma genome and putative plasmid contigs with Bowtie 2. The *V. vinifera* data set covers 58% of the genome and 100% of the plasmid with at least one read. The *M. fuscovaria* data set covers 94.7% of the genome and 100% of the plasmid. This is good evidence that the *V. vinifera* and *M. fuscovaria* data sets contain the same phytoplasma strain (both the genome and putative plasmid) as the *C. roseus* data set.

**Data availability.** Sequencing data are available at NCBI under BioProject number [PRJNA522055](#). Assembled genome and plasmid sequences from the *C. roseus* data are available at NCBI GenBank under accession numbers [CP035949](#) and [CP035950](#), respectively.

## ACKNOWLEDGMENTS

We acknowledge Jeff Joubert, Vinpro, for his help with the maintenance of the *C. roseus* plants.

The work was financially supported by Winetech, the University of Pretoria, and the European Union's Horizon 2020 (EU H2020) research and innovation program under grant agreement number 727459 (TROPICSAFE). Any dissemination of the results must indicate that they reflect only the authors' view and that the EU H2020 is not responsible for any use that may be made of the information they contain.

## REFERENCES

- Carstens R, Petersen Y, Stephan D, Burger J. 2011. Current status of aster yellows disease in infected vineyards in the Vredendal grape producing area of South Africa. *Phytopathogenic Mollicutes* 1:83–85. <https://doi.org/10.5958/j.2249-4669.1.2.014>.
- Zambon Y, Contaldo N, Richards RS, Bertaccini A, Burger J. 2015. Multi-gene characterization of aster yellows phytoplasmas infecting grapevine in South Africa. *Phytopathogenic Mollicutes* 5:S21–S22. <https://doi.org/10.5958/2249-4677.2015.00008.0>.
- Engelbrecht M, Joubert J, Burger JT. 2010. First report of aster yellows phytoplasma in grapevines in South Africa. *Plant Dis* 94:373. <https://doi.org/10.1094/PDIS-94-3-0373A>.
- Krüger K, De Klerk A, Douglas-Smit N, Joubert J, Pietersen G, Stiller M. 2011. Aster yellows phytoplasma in grapevines: identification of vectors in South Africa. *Bull Insectol* 64:S137–S138.
- Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13–15.
- Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614–620. <https://doi.org/10.1093/bioinformatics/btt593>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Zhao Y, Wei W, Lee I-M, Shao J, Suo X, Davis RE. 2013. The iPhyClassifier, an interactive online tool for phytoplasma classification and taxonomic assignment. *Methods Mol Biol* 938:329–338. [https://doi.org/10.1007/978-1-62703-089-2\\_28](https://doi.org/10.1007/978-1-62703-089-2_28).