

# High-quality full genome assembly of historic *Xylella fastidiosa* strains from ICMP collection using a hybrid sequencing approach

Pragya Kant,<sup>1</sup> Natasha Brohier,<sup>1</sup> Rachel Mann,<sup>1</sup> Luciano Rigano,<sup>2</sup> Robert Taylor,<sup>2</sup> Johanna Wong-Bajracharya,<sup>3</sup> Toni A. Chapman,<sup>3</sup> Monica Kehoe,<sup>4</sup> Rebecca Roach,<sup>5</sup> Brendan Rodoni,<sup>1</sup> Fiona Constable<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 3.

**ABSTRACT** High-quality complete genomes of five *Xylella fastidiosa* strains were assembled by combining Nanopore and Illumina sequencing data. Among these, International Collection of Micro-organisms from Plants (ICMP) 8731, ICMP 8742 and ICMP 8745 belong to subspecies *fastidiosa* while ICMP 8739 and ICMP 8740 were determined as subspecies *multiplex*. The strains were further classified into sequence types.

**KEYWORDS** bacteriology, *Xylella fastidiosa* subspecies *fastidiosa*, *Xylella fastidiosa* subspecies *multiplex*, long and short reads, MinION

*Xylella fastidiosa* is a xylem-limited plant pathogenic bacterium that has caused significant epidemics and economic losses in many plant species in the Americas and Europe (1). This bacterium is the number one quarantine pathogen for Australia and New Zealand. The genus *Xylella* has complex taxonomy, having two species: *fastidiosa* and *taiwanensis*. Within the *fastidiosa*, there are six known subspecies, three of them, *fastidiosa*, *multiplex*, and *pauca*, are of significance based on the genomic analysis (2). Furthermore, *X. fastidiosa* is differentiated into sequence types (3).

We obtained DNA from ICMP (International Collection of Micro-organisms from Plants), New Zealand for sequencing. The strain ICMP 8731 was isolated from *Vitis vinifera* and ICMP 8745 from *Ambrosia*, respectively, in Florida. The strains ICMP 8740 and ICMP 8742 were collected from host *Platanus occidentalis* and *Ulmus americana*, respectively, in Washington, DC, and ICMP 8739 was collected from *Prunus dulcis* in California. All bacterial cultures were isolated in the 1980s and had not undergone prior sequencing.

Illumina sequencing library were generated using PerkinElmer NEXTFLEX Rapid XP DNA-Seq Kit at half reaction with PerkinElmer NEXTFLEX UDI Barcodes (PerkinElmer, Massachusetts, USA). Library was paired-end sequenced (2 × 150 bp) using the Illumina NovaSeq 6000. Quality control (QC) of the data were processed using fastp (4). For Nanopore sequencing, Rapid RBK004 library was prepared and sequenced using R9.4 flowcell in a MinION Mk1B device (Oxford Nanopore Technologies) following manufacturer's instructions. The sequencing was controlled with MinKNOW software. Output fast5 files were basecalled and demultiplexed using high-accuracy model of Guppy V6.4.8 to obtain fastq files. NanoPlot (<https://github.com/wdecoster/NanoPlot>) was used for reads QC.

Hybrid assemblies were performed by Unicycler version 0.4.7 using both Nanopore and Illumina reads.

All *X. fastidiosa* strains were circularized in a chromosome except for ICMP 8739 and ICMP 8745, which were visualized with gaps and nodes using Bandage V0.8.1 (5).

**Editor** Vincent Michael Bruno, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Pragya Kant, [pragya.kant@agriculture.vic.gov.au](mailto:pragya.kant@agriculture.vic.gov.au).

The authors declare no conflict of interest.

See the funding table on p. 3.

**Received** 11 July 2023

**Accepted** 7 September 2023

**Published** 17 October 2023

Copyright © 2023 Kant 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/).

TABLE 1 Description and annotation metrics of the five ICMP *Xylella fastidiosa* strains hybrid assemblies

Strain name	ICMP 8731	ICMP 8739	ICMP 8740	ICMP 8742	ICMP 8745
Subspecies	<i>fastidiosa</i>	<i>multiplex</i>	<i>multiplex</i>	<i>fastidiosa</i>	<i>fastidiosa</i>
BioSample	SAMN33037963	SAMN33037965	SAMN33037964	SAMN33037966	SAMN33037967
GenBank accession	CP117821.1	JAQQT000000000	CP117820.1	CP117823.1	JAQQT000000000
Genome size (bp)	2,409,646	2,023,822	2,548,838	2,556,755	2,589,186
Nanopore reads	230,567	281,205	440,928	180,657	70,444
Nanopore N50 (bp)	3,069	2,523	5,225	7,306	6,247
Nanopore coverage	152x	179x	436x	246x	75x
Illumina paired-reads	7,816,992	7,121,802	7,954,171	8,174,227	8,198,756
Total Illumina bases	1,174x	1,283x	1,137x	1,154x	1,147x
GC content	51.53%	51.67%	51.65%	51.49%	51.54%
CDS	2,084	1,711	2,227	2,252	2,308
Genes	2,136	1,765	2,286	2,311	2,367
ncRNA	3	3	3	3	3
Regulatory	1	1	1	1	1
rRNA	3	6	6	6	6
tmRNA	1	1	1	1	1
tRNA	45	44	49	49	49
Plasmid name	pXFAS01			pXF879-41	
Plasmid size (bp)	38,298			41,752	

Annotations were generated by National Center of Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (6) (Table 1).

A phylogenetic analysis and a tree were generated using an up-to-date bacterial core gene (UBCG) set, a core genome analysis pipeline that extracts and concatenates 92 core

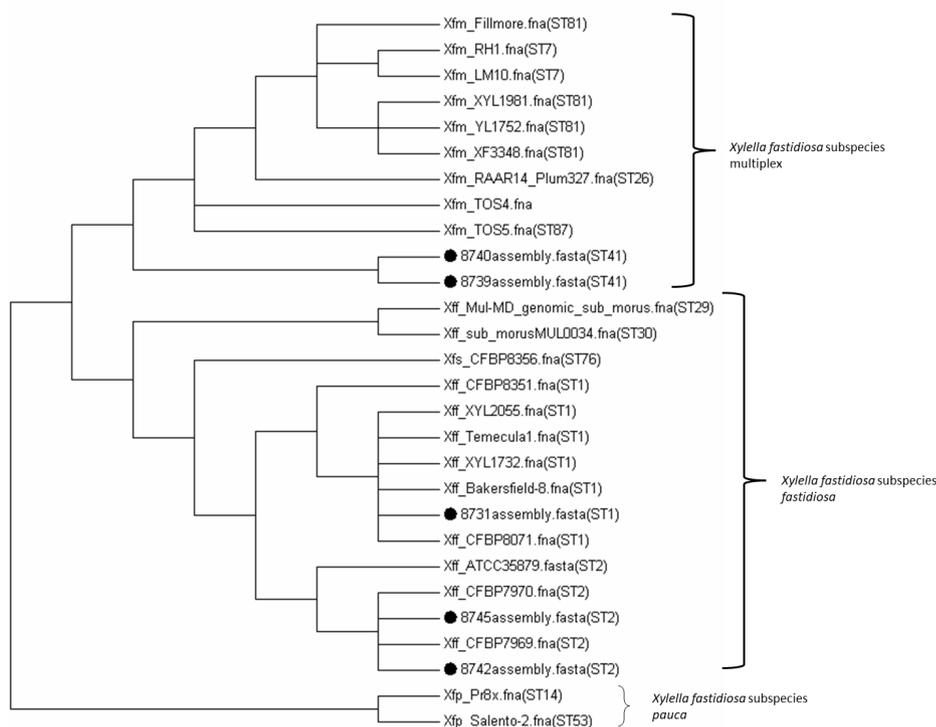


FIG 1 Phylogenetic tree displaying relatedness of the of 29 *Xylella fastidiosa* genomes using UBCG analysis. The tree was visualized in MEGA V7 (<https://www.megasoftware.net/>). Filled dots represent strains sequenced in this study. Sequence type (ST) is provided in parentheses.

bacterial genes by maximum-likelihood approach (7). In addition, sequence types were determined using Linux-based single script (<https://github.com/tseemann/mlst>). All five strains were accurately placed to their respective clades in the dendrogram (Fig. 1). For analysis, softwares were used with default parameter, unless specified otherwise.

## ACKNOWLEDGMENTS

This study was funded through Horticulture Innovation and Agriculture Victoria Research for project MT17006: Improving preparedness of the Australian horticultural sector to the threat potentially posed by *Xylella fastidiosa* (a severe biosecurity risk).

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Energy, Environment and Climate Action, Agriculture Victoria Research, Bundoora, Victoria, Australia

<sup>2</sup>Plant Health and Environment Laboratory, Ministry for Primary Industries, Auckland, New Zealand

<sup>3</sup>Department of Primary Industries, Plant Biosecurity Research and Diagnostics, New South Wales, Australia

<sup>4</sup>Department of Primary Industries and Regional Development, Biosecurity and Sustainability, Western Australia, Australia

<sup>5</sup>Department of Agriculture and Fisheries, Queensland, Australia

## PRESENT ADDRESS

Rachel Mann, Plant Health Australia, Canberra, Victoria, Australia

## AUTHOR ORCID*s*

Pragya Kant  <http://orcid.org/0000-0001-8056-8103>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">Hort Innovation (HIAL)</a>	MT17006	Fiona Constable
<a href="#">Agriculture Victoria</a>	MT17006	Brendan Rodoni

## AUTHOR CONTRIBUTIONS

Pragya Kant, Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review and editing | Natasha Brohier, Methodology | Rachel Mann, Conceptualization, Project administration, Supervision | Luciano Rigano, Methodology, Resources, Writing – review and editing | Robert Taylor, Funding acquisition, Writing – review and editing | Johanna Wong-Bajracharya, Visualization | Toni A. Chapman, Conceptualization | Rebecca Roach, Visualization | Brendan Rodoni, Funding acquisition | Fiona Constable, Funding acquisition, Supervision.

## DATA AVAILABILITY

Genome assemblies are accessible in GenBank as Bioproject- [PRJNA930964](#) consisting five accessions [CP117821.1](#), [JAQQT000000000](#), [CP117820.1](#), [CP117823.1](#), [JAQQT000000000](#). The raw reads are deposited in Sequence Read Archive, accession numbers [SRR24958443](#) to [SRR24958447](#) (Illumina) and [SRR24955971](#) to [SRR24955975](#) (Nanopore).

## REFERENCES

1. Almeida RPP, Nunney L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? *Plant Dis* 99:1457–1467. <https://doi.org/10.1094/PDIS-02-15-0159-FE>
2. Denancé N, Briand M, Gaborieau R, Gaillard S, Jacques M-A. 2019. Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*. *BMC Genomics* 20:239. <https://doi.org/10.1186/s12864-019-5565-9>
3. Yuan X, Morano L, Bromley R, Spring-Pearson S, Stouthamer R, Nunney L. 2010. Multilocus sequence typing of *Xylella fastidiosa* causing pierce's disease and oleander leaf scorch in the United States. *Phytopathology* 100:601–611. <https://doi.org/10.1094/PHTO-100-6-0601>
4. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
5. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>
6. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
7. Na S-I, Kim YO, Yoon S-H, Ha S-M, Baek I, Chun J. 2018. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol* 56:280–285. <https://doi.org/10.1007/s12275-018-8014-6>