Phytopathology® • 2023 • 113:1128-1132 • https://doi.org/10.1094/PHYTO-10-22-0396-A

RESOURCE ANNOUNCEMENT

Complete Circularized Genome Resources of Seven Strains of *Xylella fastidiosa* subsp. *fastidiosa* Using Hybrid Assembly Reveals Unknown Plasmids

María Pilar Velasco-Amo,[1] Luis F. Arias-Giraldo,[1] Miguel Román-Écija,[1] Leonardo De La Fuente,[2] Ester Marco-Noales,[3] Eduardo Moralejo,[4] Juan A. Navas-Cortés,[1] and Blanca B. Landa[1,†]

[1] Instituto de Agricultura Sostenible, Consejo Superior de Investigaciones Científicas (CSIC), Córdoba, Spain
[2] Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, U.S.A.
[3] Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, Spain
[4] Tragsa, Empresa de Transformación Agraria, Delegación de Baleares, 07005 Palma, Spain

Abstract

Xylella fastidiosa is a vascular plant pathogenic bacterium native to the Americas that is causing significant epidemics and economic losses in olive and almonds in Europe, where it is a quarantine pathogen. Since its first detection in 2013 in Italy, mandatory surveys across Europe revealed the presence of the bacterium also in France, Spain, and Portugal. Combining Oxford Nanopore Technologies and Illumina sequencing data, we assembled high-quality complete genomes of seven *X. fastidiosa* subsp. *fastidiosa* strains isolated from different plants in Spain, the United States, and Mexico. Comparative genomic analyses discovered differences in plasmid content among strains, including plasmids that had been overlooked previously when using the Illumina sequencing platform alone. Interestingly, in strain CFBP8073, intercepted in France from plants imported from Mexico, three plasmids were identified, including two (plasmids pXF-P1.CFBP8073 and pXF-P2.CFBP8073) not previously described in *X. fastidiosa* and one (pXF5823.CFBP8073) almost identical to a plasmid described in a *X. fastidiosa* strain from citrus. Plasmids found in the Spanish strains here were similar to those described previously in other strains from the same subspecies and ST1 isolated in the Balearic Islands and the United States. The genome resources from this work will assist in further studies on the role of plasmids in the epidemiology, ecology, and evolution of this plant pathogen.

Resource Announcement

Xylella fastidiosa is a plant-pathogenic bacterium that causes a variety of diseases on agricultural crops, such as almond, citrus, coffee, grapevine, and olive, and it also infects plant species in natural forest landscapes and ornamental plants. This pathogen has been associated with various outbreaks in Europe, the first one detected in Italy, and subsequently, outbreaks have been reported in France, Spain, and Portugal (EFSA et al. 2022).

In Spain, the Balearic Islands represent the largest niche of genetic diversity of this bacterium among the outbreaks that have occurred in Europe. The Balearic Islands are the only place in Europe where the three main subspecies and four different sequence types (STs) (Yuan et al. 2010) have been detected in open fields (Moralejo et al. 2020; Olmo et al. 2021).

[†]Corresponding author: B. B. Landa; blanca.landa@csic.es

The author(s) declare no conflict of interest.

Accepted for publication 22 November 2022.

Funding

Support was provided by the HORIZON-CL6-2021-FARM2FORK-01 project BeXyl (Beyond *Xylella*, Integrated Management Strategies for Mitigating *Xylella fastidiosa* Impact in Europe; grant 101060593); the project ERTA2017-00004-C06-02 (Desarrollo de estrategias de erradicación, contención y control de *X. fastidiosa* en España) from the 'Programa Estatal de I+D+I Orientada a los Retos de la Sociedad' of the Spanish Government (AEI-INIA Spain); the Spanish Olive Oil Interprofessional; and the Thematic Interdisciplinary Platform on *X. fastidiosa*, PTI Sol-Xyl, from Consejo Superior de Investigaciones Científicas. Financial support from the Spanish Ministry of Agriculture and Fisheries to the Institute for Sustainable Agriculture is also acknowledged. L. De La Fuente was a recipient of a research fellowship awarded by the Organization for Economic Cooperation and Development, Cooperative Research Programme, during the writing of this manuscript. M. P. Velasco-Amo was a recipient of a contract financed by the Consejo Superior de Investigaciones Científicas Intramural Project 201840E111.

Keywords

bacteriology, hybrid assembly, Illumina, Oxford Nanopore, plasmids, population biology, *Xylella fastidiosa*

Table 1. General genome features and assembly statistics for strains of *Xylella fastidiosa* subsp. *fastidiosa* used in the study

Strain	ST[a]	Host	Geographical origin	Number of contigs	Sequence type	Size	Contamination (%)[b]	Completeness (%)[b]	GC content (%)	Number of CDSs[c]	Number of tRNA	Number of rRNA	Number of ncRNA	Accession number
R2XF4358/18	1	*Rhamnus alaternus*	Mallorca, Spain	2	Chromosome	2,556,791	0.3	99.6	51.8	2,246	49	6	4	CP109898
					Plasmid	38,297				41	–	–	–	CP109899
XYL461	1	*Rhamnus alaternus*	Mallorca, Spain	2	Chromosome	2,540,594	0	99.6	51.7	2,224	49	6	4	CP109897
					Plasmid	38,297				41	–	–	–	CP109896
CFBP8351	1	*Vitis vinifera*	California (CA), U.S.A.	2	Chromosome	2,534,856	0	99.6	51.7	2,224	49	6	4	CP109894
					Plasmid	38,294				41	–	–	–	CP109895
IVIA5770	1	*Vitis vinifera*	Mallorca, Spain	2	Chromosome	2,542,564	0	99.6	51.8	2,218	51	9	4	CP109892
					Plasmid	38,297				42	–	–	–	CP109893
CFBP8083	2	*Vitis vinifera*	North Carolina (NC), U.S.A.	1	Chromosome	2,515,619	0	99.6	51.7	2,185	49	6	4	CP109891
WM1-1	2	*Vitis vinifera*	Georgia (GA), U.S.A.	1	Chromosome	2,514,513	0	99.6	51.7	2,193	49	6	4	CP109890
CFBP8073	75	*Coffea canephora*	Mexico	4	Chromosome	2,669,408	0.1	99.6	51.9	2,379	49	6	4	CP109886
					Plasmid	5,821				7	–	–	–	CP109887
					Plasmid	38,041				49	–	–	–	CP109888
					Plasmid	28,508				38	–	–	–	CP109889

[a] Sequence type (ST) was known and/or confirmed by genome query at the *Xylella fastidiosa* pubMLST database (Jolley et al. 2018).
[b] Genome contamination (redundancy) and completeness were checked with checkM (Parks et al. 2015).
[c] Coding sequences (CDSs) were annotated with NCBI Prokaryotic Genome Annotation Pipeline.

X. fastidiosa subsp. *fastidiosa* ST1 and *X. fastidiosa* subsp. *multiplex* ST7 have been detected only in Mallorca Island. *X. fastidiosa* subsp. *pauca* ST80 occurs only in Ibiza Island, and this particular ST has not been described elsewhere. Finally, *X. fastidiosa* subsp. *multiplex* ST81 has been described in Mallorca and Menorca Islands and is genetically different from other *X. fastidiosa* subsp. *multiplex* isolates detected in Europe. Further studies on the presence and role of plasmids among genetically diverse *X. fastidiosa* strains can provide important information in studies of epidemiology, ecology, and evolution of this plant pathogen.

Additionally, high-quality complete genomes are required to perform genome comparative analysis and to fully describe the genetic diversity of the different strains isolated in new outbreaks. For that, the combination of both long-read and short-read methods of DNA sequencing, mainly Oxford Nanopore Technologies (ONT) and Illumina, respectively, have enabled the complete high-quality assembly of *X. fastidiosa* genomes, revealing novel features such as plasmids, phages, and genome reorganizations that could not be detected otherwise (Arias-Giraldo et al. 2020; O'Leary et al. 2022).

In this resource announcement, we describe the complete circularized genomes of seven strains of *X. fastidiosa* (Table 1) obtained by combining ONT and Illumina sequencing data. Strains were isolated from *Rhamnus alaternus* (R2XF4358/18 and XYL461), *Vitis vinifera* (IVIA5770, CFBP8351, CFBP8083, and WM1-1), and *Coffea canephora* (CFBP8073). All these strains belong to subspecies *fastidiosa* but to different STs (i.e., ST1, ST2, and ST75) (Table 1).

Genomic DNA was extracted using the Quick DNA Fungal/Bacteria Miniprep kit (Zymo Research Group, Tustin, CA, U.S.A.) from *X. fastidiosa* pure cultures grown on PD2 agar medium (Davis 1980) at 28°C for 7 to 10 days. The integrity of DNA was evaluated by using 2% agarose gel electrophoresis, and DNA was quantified by spectrofluorometric methods (Qubit; Thermo Fisher Scientific, Waltham, MA, U.S.A.). ONT sequencing libraries were prepared using the LSK-109 ligation sequencing kit (ONT, Cambridge, U.K.). Libraries were sequenced in multiplexed runs on a MinION device using R9.4.1 flow cells (ONT). Sequencing reads were base called with Guppy for MK1C v6.0.7 and trimmed with Porechop v0.2.4 (Wick et al. 2017). Illumina sequencing libraries for R2XF4358/18, XYL461, IVIA5770, and CFBP8083, were prepared and paired-end sequenced with a Hiseq 4000 platform with other strains in the same run. Illumina sequencing yielded a total of 5,008,800 2 $\times$ 150-bp paired-end reads. For the remaining strains, Illumina data were retrieved from the Sequence Read Archive database: CFBP8073 (SRR8501431), CFBP8351 (SRR8454583), and WM1-1 (SRR6796150). We used the fastp tool v0.23.2 for adapter and quality trimming of short reads (Chen et al. 2018). Prior to assembly, we used the taxonomic classification tool Kraken2 v2.1.2 (Wood et al. 2019) with the PlusPFP v.9-8-2022 database (https://benlangmead.github.io/aws-indexes/k2) to remove contaminating sequences, with a total of 99 to 100% of the sequenced reads for each strain dataset being assigned to *X. fastidiosa*.

Long-read de novo genome assemblies were performed using Canu v2.2 with default parameters. The resulting draft genomes were polished with high-quality short reads (Q $>$ 25) to correct errors that long-read-only assemblies are prone to have (Watson and Warr 2019)

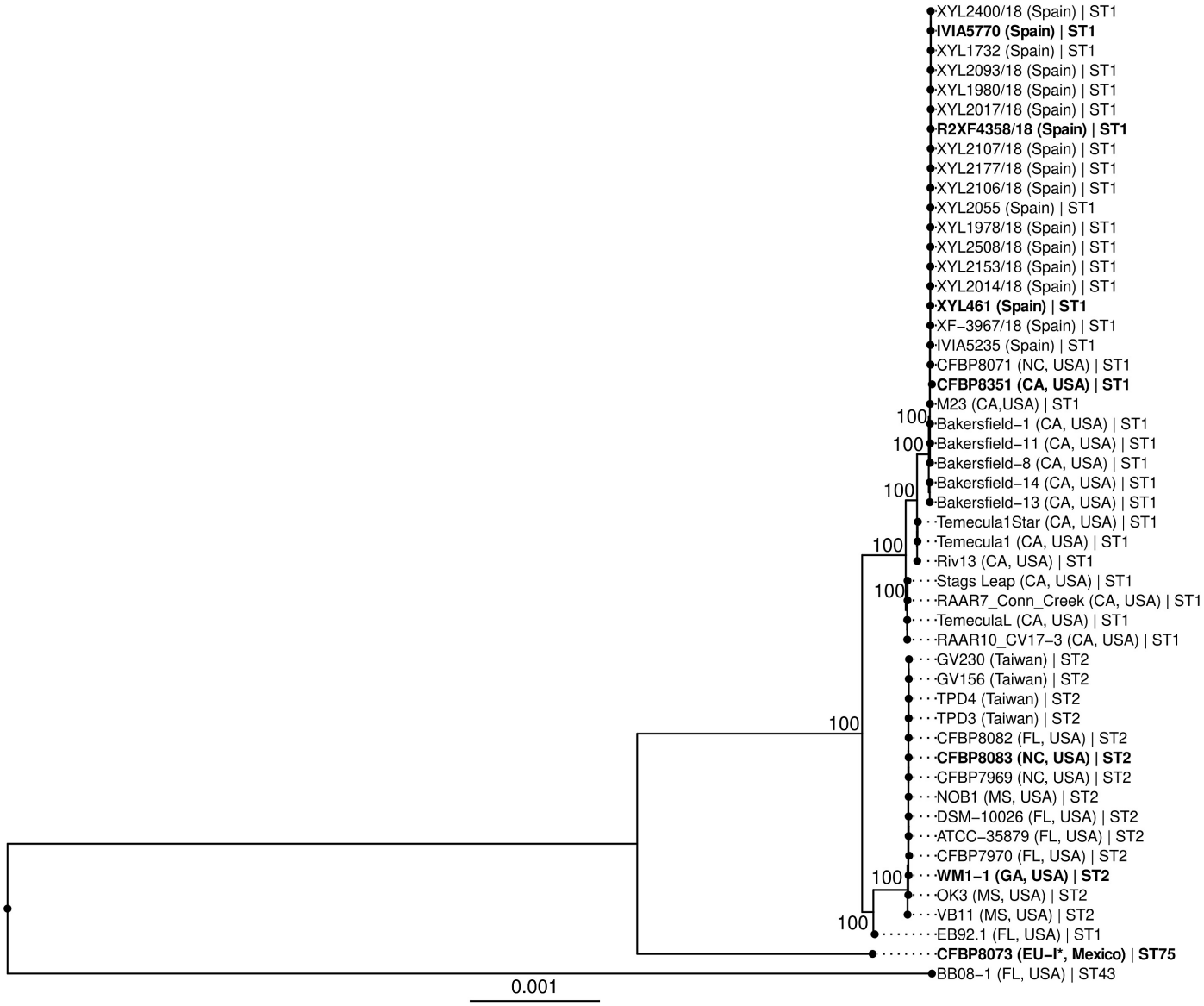

Fig. 1. Maximum likelihood phylogenetic reconstruction of the core genome for 50 *Xylella fastidiosa* subsp. *fastidiosa* strains. The strain BB08-1 (*X. fastidiosa* subsp. *multiplex*) was used as an outgroup. Strains sequenced in this study are shown in bold. The location of isolation is included in parenthesis after the strain names. *EU-I: Strain intercepted in the European Union. Numbers indicate bootstrap value.

using first Polypolish v0.5.0 (Wick and Holt 2022) and, in subsequent steps, POLCA v4.0.9 (Zimin and Salzberg 2020) until no further changes were necessary. The hybrid approach allowed for the reconstruction and circularization of the complete genome of the seven strains and revealed the presence of plasmids in all strains except for the two strains belonging to *X. fastidiosa* subsp. *fastidiosa* ST2 (i.e., CFBP8083 and WM1-1) (Table 1).

To understand the phylogenetic context of the sequenced strains within *X. fastidiosa* subsp. *fastidiosa*, a phylogenetic reconstruction of the core genome was performed using maximum likelihood analysis including all genomes from this subspecies available at GenBank (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/173/). The core genome was inferred with the CoreCruncher pipeline (Harris et al. 2021) and then processed with ClipKIT v1.3.0 (Steenwyk et al. 2020). The maximum likelihood phylogenetic tree was obtained using IQ-TREE v2.2.0.3 (Nguyen et al. 2015) applying the GTR + I + G4 model of nucleotide substitution and 1,000 bootstrap replicates. All strains were placed in different clusters according to their ST, with all Spanish strains and strain CFBP8351 being included in the same subcluster within the ST1 cluster. Strain CFBP8073 constituted an independent clade at high phylogenetic distance from the rest of *X. fastidiosa* subsp. *fastidiosa* strains (Fig. 1).

We identified in the four Spanish strains the same plasmid (pXFAS_5235) that was previously detected in strain IVIA5235, which was the first Spanish *X. fastidiosa* subsp. *fastidiosa* strain whose genome was completely circularized (Arias-Giraldo et al. 2020; Landa et al. 2018). This plasmid also showed high identity (100%) over 100% query cover with the plasmid identified in this study in strain CFBP8351, also belonging to ST1, and originated from California. Plasmid pXFAS_5235 also shows an identity ranging from 99.98 to 100% with plasmids pXF-Riv5, pXFAS01, and pXFAS01-Riv13 identified in several strains of *X. fastidiosa* belonging to two distinct subspecies (i.e., *multiplex* and *fastidiosa*) from California, suggesting the existence of recent horizontal transfer (Rogers and Stenger 2012). These results confirm the previous hypothesis of a single introduction of *X. fastidiosa* subsp. *fastidiosa* in the Balearic Islands (Velasco-Amo et al. 2022).

Interestingly, previously, no plasmids were reported in the genome assembly of strains CFBP8351 and CFBP8073 deposited in GenBank (GCA_004016405.1 and GCA_001469395.1, respectively) that were obtained using Illumina sequencing only (Denancé et al. 2019; Jacques et al. 2016). Strain CFBP8073 was isolated from coffee plants imported from Mexico that were intercepted in France and was described as an atypical *X. fastidiosa* subsp. *fastidiosa* strain (Denancé et al. 2019; Jacques et al. 2016). However, in this study, with the combination of long and short reads, we were able to reconstruct three complete plasmids for this strain, named pXF5823.CFBP8073 (5,821 bp), pXF-P1.CFBP8073 (28,508 bp), and pXF-P2.CFBP8073 (38,041 bp). Results of BLASTN searches against sequences deposited in GenBank revealed that the plasmid pXF5823.CFBP8073 has 94.74% identity over 99% query cover with a plasmid from a citrus strain of *X. fastidiosa* of unknown subspecies (Qin and Hartung 2001). On the other hand, the two largest plasmids were different from any previously described *X. fastidiosa* plasmid. Plasmid pXF-P1.CFBP8073 had an identity of 97.8%, but only 75% of query cover, with a plasmid from *X. fastidiosa* subsp. *sandyi* strain Ann-1, and plasmid pXF-P2.CFBP8073 showed a 96.42% identity but a 79% of query cover with plasmid pXF51ud from *X. fastidiosa* subsp. *pauca* strain U24d.

Within the different *X. fastidiosa* strains sequenced so far, for which genomes are available at GenBank, the presence of a large number of plasmids have been described mostly on those belonging to subspecies *pauca* (e.g., OLS0479, CVC0256, CVC0251, COF0324, and COF0407), with a maximum of four plasmids identified within a single strain (Giampetruzzi et al. 2017; Pierry et al. 2020). On the other hand, only strain Riv13 belonging to *X. fastidiosa* subsp. *fastidiosa* ST1 and isolated from the western redbud tree (*Cercis occidentalis*) in California also harbors three plasmids; however, they differ from those of strain CFBP8073.

Studies of inter- and intrasubspecific homologous recombination among *X. fastidiosa* subsp. *pauca* strains showed that the strains harboring a high number of plasmids have a high number of recombination events (Coletta-Filho et al. 2017; Kahn and Almeida 2022). However, the role of plasmids in the recombination of these strains has not been addressed yet. The genome resources from this work will assist in further studies on the role of plasmids in the epidemiology, ecology, and evolution of this plant pathogen.

Data Availability

The complete genome sequences of *X. fastidiosa* subsp. f*astidiosa* were deposited in NCBI under BioSample accession numbers SAMN31318971 (R2XF4358/18), SAMN31318972 (XYL461), SAMN31318973 (CFBP8351), SAMN31318974 (IVIA5770), SAMN31318975 (CFBP8083), SAMN31318976 (WM1-1), and SAMN31318977 (CFBP8073) (Table 1). All the accession numbers are associated with BioProject PRJNA891297. The strains CFBP8351, CFBP8083, and CFBP8073 are available at the CIRM-CFBP (International Center for Microbial Resources, Collection for Plant-Associated Bacteria, IRHS UMR, Beaucouzé Cedex, France), and the remaining are deposited at the *X. fastidiosa* strain collection at the Institute for Sustainable Agriculture, Córdoba, Spain.

Literature Cited

Arias-Giraldo, L. F., Giampetruzzi, A., Metsis, M., Marco-Noales, E., Imperial, J., Velasco-Amo, M. P., Román-Écija, M., and Landa, B. B. 2020. Complete circularized genome data of two Spanish strains of *Xylella fastidiosa* (IVIA5235 and IVIA5901) using hybrid assembly approaches. Phytopathology 110:969-972.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.

Coletta-Filho, H. D., Francisco, C. S., Lopes, J. R. S., Muller, C., and Almeida, R. P. P. 2017. Homologous recombination and *Xylella fastidiosa* host–pathogen associations in South America. Phytopathology 107:305-312.

Davis, M. J. 1980. Isolation media for the Pierce's disease bacterium. Phytopathology 70:425.

Denancé, N., Briand, M., Gaborieau, R., Gillard, S., and Jacques, M.-A. 2019. Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*. BMC Genomics 20:239.

EFSA (European Food Safety Authority), Delbianco, A., Gibin, D., Pasinato, L., and Morelli, M. 2022. Scientific report on the update of the *Xylella* spp. host plant database – systematic literature search up to 30 June 2021. EFSA J. 20:67.

Giampetruzzi, A., Saponari, M., Loconsole, G., Boscia, D., Savino, V. N., Almeida, R. P. P., Zicca, S., Landa, B. B., Chacón-Díaz, C., and Saldarelli, P. 2017. Genome-wide analysis provides evidence on the genetic relatedness of the emergent *Xylella fastidiosa* genotype in Italy to isolates from Central America. Phytopathology 107: 816-827.

Harris, C. D., Torrance, E. L., Raymann, K., and Bobay, L.-M. 2021. CoreCruncher: Fast and robust construction of core genomes in large prokaryotic data sets. Mol. Biol. Evol. 38:727-734.

Jacques, M.-A., Denancé, N., Legendre, B., Morel, E., Briand, M., Mississipi, S., Durand, K., Olivier, V., Portier, P., Poliakoff, F., and Crouzillat, D. 2016. New coffee plant-infecting *Xylella fastidiosa* variants derived via homologous recombination. Appl. Environ. Microbiol. 82:1556-1568.

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 3:124.

Kahn, A. K., and Almeida, R. P. P. 2022. Phylogenetics of historical host switches in a bacterial plant pathogen. Appl. Environ. Microbiol. 88.

Landa, B. B., Velasco-Amo, M. P., Marco-Noales, E., Olmo, D., López, M. M., Navarro, I. Monterde, A., Barbé, S., Montes-Borrego, M., Román-Écija, M., Saponari, M., and Giampetruzzi, A. 2018. Draft genome sequence of *Xylella fastidiosa* subsp. *fastidiosa* strain IVIA5235, isolated from *Prunus avium* in Mallorca Island, Spain. Microbiol. Resour. Announc. 7:e01222-18.

Moralejo, E., Gomila, M., Montesinos, M., Borràs, D., Pascual, A., Nieto, A., Adrover, F., Gost, P. A., Seguí, G., Busquets, A., Jurado-Rivera, J. A., Quetglas, B., García, J. D., Beidas, O., Juan, A., Velasco-Amo, M. P., Landa, B. B., and Olmo, D. 2020. Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe. Commun. Biol. 3: 560.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268-274.

O'Leary, M. L., Arias-Giraldo, L. F., Burbank, L. P., De La Fuente, L., and Landa, B. B. 2022. Complete genome resources for *Xylella fastidiosa* strains AlmaEM3 and BB08-1 reveal prophage-associated structural variation among blueberry-infecting strains. Phytopathology 112:732-736.

Olmo, D., Nieto, A., Borràs, D., Montesinos, M., Adrover, F., Pascual, A., Gost, P. A., Quetglas, B., Urbano, A., García, J. D., Velasco-Amo, M. P., Olivares-García, C., Beidas, O., Juan, A., Marco-Noales, E., Gomila, M., Rita, J., Moralejo, E., and Landa, B. B. 2021. Landscape epidemiology of *Xylella fastidiosa* in the Balearic Islands. Agronomy 11:473.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Gen. Res. 25:1043-1055.

Pierry, P. M., Uceda-Campos, G., Feitosa-Junior, O. R., Martins-Junior, J., de Santana, W. O., Coletta-Filho, H. D., Zaini, P. A., and Silva, A. M. 2020. Genetic diversity of *Xylella fastidiosa* plasmids assessed by comparative genomics. Trop. Plant Pathol. 45:342-360.

Qin, X., and Hartung, J. S. 2001. Construction of a shuttle vector and transformation of *Xylella fastidiosa* with plasmid DNA. Curr. Microbiol. 43:158-162.

Rogers, E. E., and Stenger, D. C. 2012. A conjugative 38 kB plasmid is present in multiple subspecies of *Xylella fastidiosa*. PLoS One 7:e52131.

Steenwyk, J. L., Iii, T. J. B., Li, Y., Shen, X.-X., and Rokas, A. 2020. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 18:e3001007.

Velasco-Amo, M. P., Arias-Giraldo, L. F., Olivares-García, C., Denancé, N., Jacques, M.-A., and Landa, B. B. 2022. Use of *tra*C gene to type the incidence and distribution of pXFAS_5235 plasmid-bearing strains of *Xylella fastidiosa* subsp. *fastidiosa* ST1 in Spain. Plants 11:1562.

Watson, M., and Warr, A. 2019. Errors in long-read assemblies can critically affect protein prediction. Nat. Biotechnol. 37:124-126.

Wick, R. R., and Holt, K. E. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput. Biol. 18:e1009802.

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13:e1005595.

Wood, D. E., Lu, J., and Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20:257.

Yuan, X., Morano, L., Bromley, R., Spring-Pearson, S., Stouthamer, R., and 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.

Zimin, A. V., and Salzberg, S. L. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLOS Comput. Biol. 16: e1007981.