

## Genome-Wide Analysis Provides Evidence on the Genetic Relatedness of the Emergent *Xylella fastidiosa* Genotype in Italy to Isolates from Central America

Annalisa Giampetruzzi, Maria Saponari, Giuliana Loconsole, Donato Boscia, Vito Nicola Savino, Rodrigo P. P. Almeida, Stefania Zicca, Blanca B. Landa, Carlos Chacón-Díaz, and Pasquale Saldarelli

First, third, and fifth authors: Università degli Studi di Bari Aldo Moro, Dipartimento di Scienze del Suolo della Pianta e degli Alimenti, via Amendola 165/A, Bari, Italy; second, fourth, seventh, and tenth authors: Consiglio Nazionale delle Ricerche, Istituto per la Protezione Sostenibile delle Piante, Bari, via Amendola 122/D, Bari, Italy; sixth author: Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, 94720-3114; eighth author: Institute for Sustainable Agriculture, Consejo Superior de Investigaciones Científicas, 14004 Córdoba, Spain; and ninth author: Centro de Investigación en Enfermedades Tropicales (CIET), Facultad de Microbiología, Universidad de Costa Rica, 2060 San José, Costa Rica.

Accepted for publication 7 April 2017.

### ABSTRACT

*Xylella fastidiosa* is a plant-pathogenic bacterium recently introduced in Europe that is causing decline in olive trees in the South of Italy. Genetic studies have consistently shown that the bacterial genotype recovered from infected olive trees belongs to the sequence type ST53 within subspecies *pauca*. This genotype, ST53, has also been reported to occur in Costa Rica. The ancestry of ST53 was recently clarified, showing it contains alleles that are monophyletic with those of subsp. *pauca* in South America. To more robustly determine the phylogenetic placement of ST53 within *X. fastidiosa*, we performed a comparative analysis based on single nucleotide polymorphisms (SNPs) and the study of the pan-genome of the 27 currently public available whole genome sequences of *X. fastidiosa*. The resulting maximum-parsimony and maximum likelihood trees constructed using the SNPs and the pan-genome analysis are consistent with previously described *X. fastidiosa* taxonomy, distinguishing the subsp. *fastidiosa*,

*multiplex*, *pauca*, *sandyi*, and *morus*. Within the subsp. *pauca*, the Italian and three Costa Rican isolates, all belonging to ST53, formed a compact phylotype in a clade divergent from the South American *pauca* isolates, also distinct from the recently described coffee isolate CFBP8072 imported into Europe from Ecuador. These findings were also supported by the gene characterization of a conjugative plasmid shared by all the four ST53 isolates. Furthermore, isolates of the ST53 clade possess an exclusive locus encoding a putative ATP-binding protein belonging to the family of histidine kinase-like ATPase gene, which is not present in isolates from the subspecies *multiplex*, *sandyi*, and *pauca*, but was detected in ST21 isolates of the subspecies *fastidiosa* from Costa Rica. The clustering and distinctiveness of the ST53 isolates supports the hypothesis of their common origin, and the limited genetic diversity among these isolates suggests this is an emerging clade within subsp. *pauca*.

*Xylella fastidiosa* is a plant-pathogenic bacterium widely distributed in the Americas, where it is associated with a number of plant diseases and relevant losses in commercially important crops (Almeida and Nunney 2015). The bacterium is transmitted by xylem sap-feeding insects (Hemiptera; subfamily Cicadellinae, superfamily Cercopoidea, and potentially the superfamily Cica-doidea) (Almeida and Nunney 2015). *X. fastidiosa* is classified into four recognized subspecies (*fastidiosa*, *multiplex*, *pauca*, and *sandyi*), each of which has a largely nonoverlapping host range and allopatric origins (Nunney et al. 2014a, c). A fifth subspecies has been proposed (subsp. *morus*), which includes isolates infecting mulberry; subsp. *morus* may have originated from intersubspecific homologous recombination (IHR) (Nunney et al. 2014c). IHR and transfer of genetic elements such as plasmids and integrated prophages are major players in *X. fastidiosa* evolution, potentially conferring the bacterium the ability to explore new niches and infecting new hosts (Nunney et al. 2014a, c).

The evolutionary history and geographical distribution of *X. fastidiosa* subspecies in the Americas indicate that the different subspecies have developed in geographical isolation: subsp.

*multiplex* in North America, subsp. *fastidiosa* in Central America, subsp. *pauca* in South America, and subsp. *sandyi* being only detected in southern regions of the United States (Almeida and Nunney 2015). Although *X. fastidiosa* populations appeared to have historically remained largely isolated due to geographical barriers, recent human-mediated invasions are responsible for important plant diseases driven by bacterial introductions in new areas. Examples include the genotype causing disease in grapevines in the United States, which originated from Central America (Nunney et al. 2010); the finding in Brazil of plum trees harboring isolates belonging to the North American subsp. *multiplex* (Nunes et al. 2003); the identification in Taiwan of isolates genetically related to the North American strains of *X. fastidiosa* subsp. *fastidiosa* (Su et al. 2013); and the report of subsp. *pauca* in Central America (Nunney et al. 2014b). Likewise, the recent emergence of *X. fastidiosa* in Europe (Giampetruzzi et al. 2015a; Saponari et al. 2013) illustrates the potential for long-distance dispersal of this bacterium and the vulnerability of the Mediterranean landscape and agriculture to invasion by new organisms.

In Europe, the bacterium was detected in olive trees (*Olea europaea* L.) in southern Italy (Apulia region) in 2013, the first outbreak under field conditions in the European Union, and the first documented event of widespread pathogenic infections in this plant species. Infected olive trees are affected by a severe disease called olive quick decline syndrome (OQDS). Symptoms of OQDS include yellow and brown lesions on leaf tips and margins, extensive branch and twig dieback, and subsequent tree mortality

Corresponding author: P. Saldarelli; E-mail address: pasquale.saldarelli@ipsp.cnr.it

\*The e-Xtra logo stands for “electronic extra” and indicates that one supplementary file and two supplementary tables are published online.

(Saponari et al. 2013). Multilocus sequence typing (MLST) and comparative genomic analyses of the draft genome of an olive-infecting strain (Loconsole et al. 2016) showed its genetic relatedness to *X. fastidiosa* subsp. *pauca* isolates. This novel sequence type was denoted ST53; was concurrently reported in the same year from Costa Rica (Nunney et al. 2014b). Five out of the seven MLST alleles of the ST53 type are identical to those found in Brazil in *X. fastidiosa* subsp. *pauca* isolates, while the two remaining alleles were new. However, a careful sequence analysis of the origin of these two unknown alleles (*cysG#24* and *nuoL#16*) revealed their subsp. *pauca* ancestry (Nunney et al. 2014b). Although exploring only a limited part of the bacterial genome, MLST analysis led to the hypothesis that the ST53 phylotype originated from a single event of introduction of subsp. *pauca* in Central America. The discovery of the ST53, together with recent European Union interceptions of *X. fastidiosa* isolates harboring several new sequence types (Loconsole et al. 2016), illustrates the limited knowledge on the diversity of this bacterium.

Extensive sampling of diseased olive trees and other infected hosts in the epidemic area of Apulia demonstrated the widespread presence of ST53 alone, indicating that a single pathogen introduction led to the current outbreak (Loconsole et al. 2016). A conclusion that was recently confirmed by a panel of experts of EFSA (European Food Safety Authority) that reviewed the existing literature on the *X. fastidiosa* outbreak in Apulia (EFSA PLH Panel 2016). Recent surveillance of plant material shipments from Central and South America to Europe led to the frequent identification of *X. fastidiosa*-infected plants (Baker et al. 2015). Consequently, it is plausible that infected plants introduced from this geographical area were responsible for the introduction of ST53 into Italy. In a recent report by the European Food Safety Authority (Baker et al. 2015), the probability of entry of *X. fastidiosa* from countries where the bacterium is endemic was rated as very likely, because plants for planting are frequently asymptomatic and may constitute an efficient source of inoculum to start new outbreaks. Recent interceptions of *X. fastidiosa*-infected coffee plants in Italy support this conclusion (Loconsole et al. 2016).

Genome-wide studies based on SNPs have been useful in tracking pathogen outbreaks, inferring evolution and spread pathways, in addition to identifying pathogenicity and host specificity determinants (Hendriksen et al. 2011; Rusconi et al. 2016). A previous SNP-based study was limited to only four complete *X. fastidiosa* genomes (Doddapaneni et al. 2006). Recently, the availability of whole genome sequence data for the four ST53 isolates (Nunney et al. 2014b) from Costa Rica allowed a study that identified SNPs characterizing the four ST53 isolates (Marcelletti and Scortichini 2016). Additional average nucleotide identity analysis concluded that these isolates belong to a clonal complex of the subsp. *pauca* (Marcelletti and Scortichini 2016).

We performed a comparative genome-wide analysis on a group of 27 finished or draft genome sequences of *X. fastidiosa* available in GenBank database to explore the genetic relatedness of the four ST53 isolates from Italy and Costa Rica with all genomic data available for this species, particularly subspecies *pauca*. Phylogenetic relationships in the *X. fastidiosa* subspecies were assessed and analysis of the pan-genome of this bacterial species performed. The study was extended to a conjugative plasmid and a coding sequence (CDS) encoding a putative ATPase binding protein found in all four ST53 and two additional subsp. *fastidiosa* isolates from Costa Rica.

## MATERIALS AND METHODS

**SNP analysis.** A total of 27 *X. fastidiosa* genomes were available in the NCBI Genomes database (<https://www.ncbi.nlm.nih.gov/genome/genomes/173>) at the date of data download (Table 1). Complete genome or draft genome sequences were downloaded in FASTA format along with their annotations and used as input for the

kSNP3.02 software (Gardner et al. 2015). The software identifies SNPs in whole genome data, which can consist of finished, draft (i.e., list of contigs), or raw unassembled reads. The strategy of SNPs finding relies on the identification of stretches of homologous nucleotides in all genomes having SNPs in their median position. The size of the nucleotide regions flanking the SNPs is defined by the *k-mer* and is set by the user after running of a build-in program Kchooser. Briefly, the software generates several files, which list the identified SNPs loci and their annotation in the proteins inferred by the downloaded GenBank information. In addition, an SNP matrix (SNPs\_all\_matrix.fasta) was created from which a tree was produced by maximum-parsimony (MP) as a consensus of up to 100 equally parsimonious trees. The use of all SNPs instead of a subset would reduce the effect of sequencing errors, which cannot be distinguished from real SNPs by the software. The same SNP matrix was used as input in MEGA 7.0.18 (Kumar et al. 2016) to be analyzed by the maximum likelihood (ML) method. The analysis was based on the general time-reversible model (Nei and Kumar 2000) and a discrete Gamma distribution. Robustness of the ML tree was assessed by 100 bootstrap experiments.

**ST53 genome-wide comparisons.** Whole genome comparison of the four ST53 draft genomes OLS0478, OLS0479, and COF0407 from Costa Rica, and CoDiRO from Italy, was made by sequence alignment using progressive Mauve tool and the construction of a circular genomic map for each genome by the BLAST Ring Image Generator (BRIG, version 0.95; Alikhan et al. 2011). Each circular genomic map was drawn using as reference genome the strain CoDiRO of *X. fastidiosa* subsp. *pauca* (henceforth referred to as “alignment reference genome”) based on a BLAST+ analysis performed on a local computer, with default options and parameter values of 50% lower –70% upper cut-off for identity and E-value threshold of  $10^{-5}$ .

The online software PHASTER (faster version of the PHAST phage search tool) (Arndt et al. 2016) was used to identify DNA prophage regions in the chromosome drafts of the four available ST53 isolates.

Comparative genomic of 27 *X. fastidiosa* genomes was performed using GET\_HOMOLOGS package (Contreras-Moreira and Vinuesa 2013) to identify the core/pan-genome according to the recommendations of Tettelin et al. (2005). Prior to conducting the analyses, all genomes were reannotated using RAST (Aziz et al. 2008; Overbeek et al. 2014) to reduce the bias introduced by the different pipeline (or software) used for the annotation of the 27 genomes. This step produced the input files in GenBank format allowing the clustering of nucleotide and protein sequences. The software includes a step of all-against-all BlastP followed by clustering based on OrthoMCL v1.4 (Li et al. 2003) and bidirectional best hit, COGsoft algorithm (Kristensen et al. 2010) to yield homologous gene clusters (Tettelin et al. 2005). This result was filtered using compare\_cluster.pl (a script in the GET\_HOMOLOGS package) with option “-t n,” where n is the number of genomes. An average nucleotide identity matrix was calculated among CDSs of the four (i.e., “cloud”, “shell”, “soft core”, and “core”) pan-genome clusters using GET\_HOMOLOGS. The produced identity matrix was used to calculate a heatmap and a dendrogram showing pairwise dissimilarities (Gower’s distance) among CDSs of the 27 pan-genomes.

**DNA amplification.** DNAs purified from bacterial cultures of isolates from different *X. fastidiosa* subspecies (Table 2) were used for amplification. Oligonucleotide primers ATPasefor1 (5'-ACCTGCATAGGTGTGAG-3') and ATPaserev2 (5'-GCTGCCGATAGCTCGAGTTCC-3') targeting the CoDiRO 2876 nucleotide sequence, which contains the putative histidine kinase-like ATPase gene (locus RA12\_07795), were located on contig000005 (GenBank JUJW01000005.1) between nucleotides 26,055 and 28,859. Primers were designed on the alignment of the genomes of the four Costa Rican isolates. PCR reactions were set following standard amplification conditions using an annealing temperature of 57°C.

TABLE 1. List and general features of *Xylella fastidiosa* strains used in the study<sup>a</sup>

Strain	Host plant	Origin	Collection date	Size (Mb)/CDS	GC%	Accession number (chromosomes/WGS)	Plasmids accession number	Citation or release date
9a5c	<i>Citrus sinensis</i>	São Paulo, Brazil	1992	2.73175/2,982	52.64	NC002488.3	pXF1.3: NC002489.3; pXF51: NC002490.1	Simpson et al. 2000
Dixon	<i>Prunus dulcis</i>	California, USA	na	2.62233/2,914	52	AAAL02	–	7/10/2002
Temecula1	<i>Vitis vinifera</i>	California, USA	1998	2.52115/2,757	51.80	NC004556.1	pXFPD1.3: NC004554.1	Van Sluys et al. 2003
M12	<i>P. dulcis</i>	California, USA	2003	2.47513/2,706	51.9	NC010513.1	–	Chen et al. 2010
M23	<i>P. dulcis</i>	California, USA	2003	2.57399/2,803	51.76	NC010577.1	pXFAS01: NC010579.1	Chen et al. 2010
GB514	<i>Vitis vinifera</i>	Texas, USA	na	2.51738/2,835	51.77	NC017562.1	Unnamed: NC017561.1	9/23/2010
EB92.1	<i>Sanbuccus nigra</i>	USA	1992	2.47543/2,750	51.5	AFDJ01	–	Zhang et al. 2011
ATCC 35871	<i>P. salicina</i>	Georgia, USA	na	2.41626/2,676	51.7	AUAJ01	–	7/15/2013
Griffin-1	<i>Quercus rubra</i>	Georgia, USA	2006	2.38731/2,724	51.7	AVGA01	–	Chen et al. 2013
32c	<i>Coffea</i> sp.	São Paulo, Brazil	na	2.60755/2,915	52.5	AWYH01	–	Alencar et al. 2014
6c	<i>Coffea</i> sp.	São Paulo, Brazil	na	2.606/2,916	52.35	AXB01	pXF6c: NZAXBS01000046.1	Alencar et al. 2014
Mul-MD	<i>Morus alba</i>	Maryland, USA	2011	2.52055/2,838	51.6	AXDP01	–	Guan et al. 2014b
Ann-1	<i>Nerium oleander</i>	USA	na	2.78091/3,162	52.07	NZCP006696.1	Unnamed1: NZCP006697.1	6/6/2014
MUL0034	<i>M. alba</i>	USA	na	2.66658/2,974	51.97	NZCP006740.1	Unnamed2: NZCP006739.1	6/6/2014
Sycamore	<i>Platanus occidentalis</i>	Virginia, USA	2002	2.47588/2,758	51.6	JMHP01	–	Guan et al. 2014a
Sy-VA	<i>Vitis vinifera</i>	Florida, USA	1987	2.52233/2758	51.8	JQAP01	–	10/21/2014
ATCC 35879	<i>Olea europaea</i>	Italy	2013	2.54293/2,816	51.96	JUJW01	Unnamed: NZCM003178.1	Giampetruzzi et al. 2015a
CoDiRO	( <i>Vinca minor</i> )							
CO33	<i>Coffea arabica</i>	Costa Rica	2015	2.68193	51.7	LJZW01	–	Giampetruzzi et al. 2015b
CFBP8072	<i>Coffea arabica</i>	Ecuador	2012	2.49666/2,763	51.9	LKDK01	–	Jacques et al. 2016
CFBP8073	<i>Coffea canephora</i>	Mexico	2012	2.58215/2,840	51.6	LKES01	–	Jacques et al. 2016
OLS0479	<i>Nerium oleander</i>	Costa Rica, San Jose Province	2011	2.53996/2,827	51.93	LRVH01	pXF-P1.COF0407: NZCM003743.1; pXF-P4.COF0407: NZCM003744.1; pXF-PS.COF0407: NZCM003745.1; pXF-RC.COF0407: NZCM003746.1	2/5/2016
CVC0256	<i>Citrus sinensis</i>	São Paulo, Brazil	1999	2.70214/3,045	52.51	LRVF01	pXF-BHR.CVC0256: NZCM003748.1; pXF-P1.CVC0256: NZCM003749.1; pXF-P4.CVC0256: NZCM003750.1; pXF-PS.CVC0256: NZCM003751.1	2/5/2016
OLS0478	<i>Nerium oleander</i>	Costa Rica, San Jose Province	2011	2.55541/2,861	51.91	LRVI01	pXF-P1.OLS0478: NZCM003752.1; pXF-P4.OLS0478: NZCM003753.1	2/5/2016
CVC0251	<i>Citrus sinensis</i>	São Paulo, Brazil	1999	2.74025/3,103	52.51	LRVE01	pXF-BHR.CVC0251: NZCM003754.1; pXF-P1.CVC0251: NZCM003755.1; pXF-P4.CVC0251: NZCM003756.1; pXF-PS.CVC0251: NZCM003757.1	2/5/2016
COF0324	<i>Coffea</i> sp.	Minas Gerais, Brazil	2006	2.77256/3,143	52.43	LRVG01	pXF-BHR-COF0324: NZCM003758.1; pXF-P1.COF0324: NZCM003759.1; pXF-PCCFO0324: NZCM003760.1; pXF-RC.COF0324: NZCM003761.1	2/5/2016
COF0407	<i>Coffea</i> sp.	Costa Rica, San Jose Province	2009	2.53847/2,837	51.84	LRVJ01	pXF-P1.OLS0479: NZCM003762.1; pXF-P4.OLS0479: NZCM003763.1; pXF-PS.OLS0479: NZCM003764.1; pXF-RC.OLS0479: NZCM003765.1	2/5/2016
Stag's Leap	<i>Vitis vinifera</i>	California, USA	2005	2.5108/2,755	51.7	LSMJ01	–	Chen et al. 2016

<sup>a</sup> The coding sequence (CDS) number was obtained by RAST. Priority was given to citation, when existing, alternatively the release date is reported. na, not available.

The use of *X. fastidiosa* DNA material for genetic diversity studies was revised and approved by the “Comité Institucional de Biodiversidad” from University of Costa Rica, and in agreement with the corresponding law “Ley de Biodiversidad,” CR (Law 7788, Law on Biodiversity).

## RESULTS

**SNP analysis of 27 *X. fastidiosa* genomes.** A total of 119,685 SNPs were identified in the 27 *X. fastidiosa* genomes, of which 36,264 (30% of the total SNPs) were homoplastic. This number corresponds to SNPs shared by groups of genomes that are not clustered in the consensus MP tree. The majority (119,538) of the total SNPs was found in the annotated genomes and 45,378 of them (37.9%) determined nonsynonymous amino acid substitutions. The average SNP frequency, considering coding and noncoding regions, was  $4.78 \times 10^{-2}$ , which is almost five times larger than that estimated by Doddapaneni et al. (2006) using four genomes. Comparison with a recent study (Marcelletti and Scortichini 2016) was not possible since these data, regarding the four ST53 genomes, was not reported.

The consensus MP tree reconstructed from the corresponding SNP matrix clustered *X. fastidiosa* isolates consistently with the current phylogenetic results obtained with MLST, identifying main clusters for the subsp. *fastidiosa*, *multiplex*, *pauca*, and *sandyi* (Fig. 1A). The phylogenetic placement of the isolate CO33, which is basal to the *fastidiosa* cluster and close to the subsp. *sandyi* (i.e., *sandyi* Ann 1 in Figure 1A), confirms the already reported phylogeny (Loconsole et al. 2016), whereas isolate CFBP8073 groups with the subsp. *fastidiosa* (Jacques et al. 2016). In addition, the SNP-based study supported the separate clustering of the subsp. *morus* from subsp. *multiplex* and subsp. *fastidiosa*, in agreement

with the suggested intersubspecific homologous recombination origin of this group (Nunney et al. 2014c).

Isolates of the subsp. *pauca* shared 12,940 SNP alleles (i.e., meaning that there are 12,940 SNP loci for which all subsp. *pauca* isolates share the same allele). Isolate CFBP8072, intercepted in France in coffee (Jacques et al. 2016), was placed on a separate basal branch from the remaining subsp. *pauca* isolates, exclusively bearing 5,411 SNPs. The remaining two clades, whose members share 2,452 SNPs, are composed by the Brazilian isolates from coffee and citrus, and the Costa Rican isolates belonging to ST53. Significantly, the ST53 isolates from Costa Rica (OLS0478, OLS0479 from oleander, and COF0407 from coffee) formed a monophyletic cluster containing the Italian CoDiRO isolate. This group is genetically supported by 4,076 synapomorphic SNPs and branches in two subgroups, one of which has CoDiRO and COF0407 and the other containing the two isolates from oleander, OLS0478 and OLS0479. According to Nunney et al. (2014b), these latter two isolates were obtained by repeated sampling of isolates from oleander in a short period of time. The overall genetic diversity of the ST53 isolates, inferred by the short branch lengths of this cluster, is very limited as shown by the low number of SNPs (103 and 151) that separates the two subgroups. In comparison with a recent study (Marcelletti and Scortichini 2016), a higher number of SNPs distinguishing CoDiRO from COF0407 and the other two ST53 isolates was found; likely because of the use of different bioinformatics approaches.

Phylogenetic relationships among the 27 genomes were also reconstructed by maximum likelihood (ML) using all SNPs detected by kSNP3.0 as input file for the analysis with MEGA 7.0.18. The inferred ML tree (Fig. 1B) was strongly supported by high bootstrap values at each node. Topologies of the ML and MP trees were almost congruent and subspecies are clearly distinguished, with the exception of isolates CO33 and *sandyi* Ann 1.

TABLE 2. List of *Xylella fastidiosa* isolates tested with primers ATPase for 1/ATPase rev2 with relevant information regarding their origin, host plant, and phylotype<sup>a</sup>

Isolate	Subspecies	Phylotype	Host	Country	PCR (bp) approximate
J4	<i>pauca</i>	Unknown	Coffee	Brazil	700 weak
J41	<i>pauca</i>	Unknown	Citrus	Brazil	700
J48	<i>pauca</i>	Unknown	Coffee	Brazil	700
J68	<i>pauca</i>	Unknown	Coffee	Brazil	700 weak
J12	<i>pauca</i>	Unknown	Citrus	Brazil	700
J62	<i>pauca</i>	Unknown	Citrus	Brazil	700
J55	<i>pauca</i>	Unknown	Citrus	Brazil	700
J65	<i>pauca</i>	Unknown	Citrus	Brazil	700
J1	<i>pauca</i>	Unknown	Citrus	Brazil	700
J44	<i>pauca</i>	Unknown	Citrus	Brazil	700
J2	<i>pauca</i>	Unknown	Citrus	Brazil	700
6570	<i>pauca</i>	Unknown	Citrus	Brazil	700
J33	<i>pauca</i>	Unknown	Coffee	Brazil	–
Dixon	<i>multiplex</i>	ST6	Almond	USA	–
BB08-1	<i>multiplex</i>	Unknown	Blueberry	USA	–
Temecula	<i>fastidiosa</i>	ST1	Grape	USA	–
AlmaEm3	<i>multiplex</i>	Unknown	Blueberry	USA	–
Sandyi	<i>sandyi</i>	Unknown	Oleander	USA	–
M23	<i>fastidiosa</i>	ST1	Almond	USA	–
9a5c	<i>pauca</i>	ST13	Citrus	Brazil	–
CoDiRO	<i>pauca</i>	ST53	Olive	Italy	2,800
No1	<i>pauca</i>	ST53	Oleander	Costa Rica	2,800
No2	<i>pauca</i>	ST53	Oleander	Costa Rica	2,800
No3	<i>pauca</i>	ST53	Oleander	Costa Rica	2,800
No4	<i>pauca</i>	ST53	Oleander	Costa Rica	2,800
No5	<i>pauca</i>	ST53	Oleander	Costa Rica	2,800
No7	<i>pauca</i>	ST53	Oleander	Costa Rica	2,800
Vv1	<i>fastidiosa</i>	ST18	Grapevine	Costa Rica	–
Vv2	<i>fastidiosa</i>	ST18	Grapevine	Costa Rica	–
Ca4	<i>fastidiosa</i>	ST21	Coffee	Costa Rica	2,800
Ca7	<i>fastidiosa</i>	ST21	Coffee	Costa Rica	2,800
Ca5	<i>fastidiosa</i>	ST33	Coffee	Costa Rica	–
CO33	<i>undetermined</i>	ST72	Coffee	Costa Rica	–

<sup>a</sup> The size of the amplified product is indicated. –, no amplification.

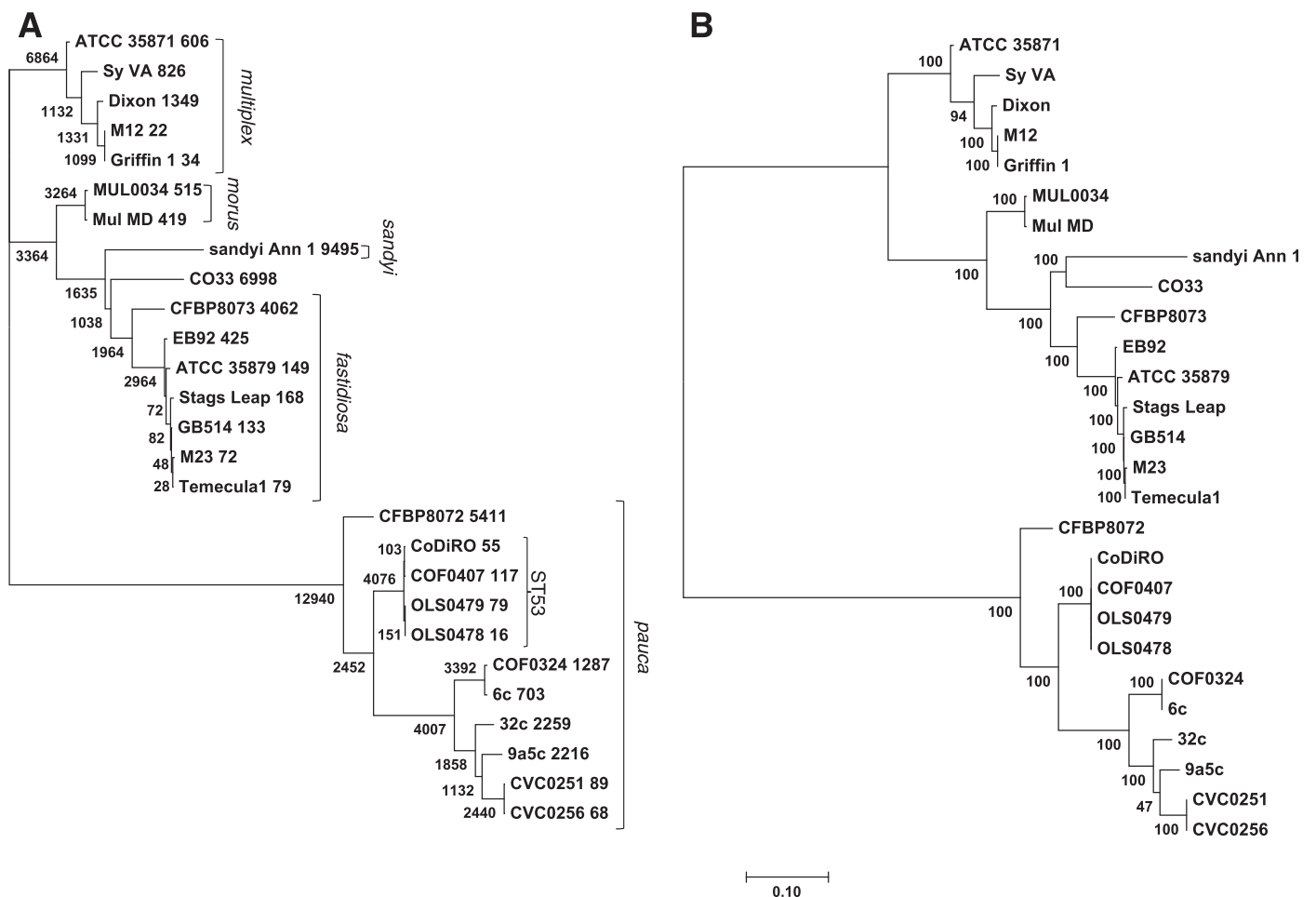
Indeed, the ML tree shows that these two isolates cluster together in a strongly supported subsp. *sandyi* clade. However, comparison of the MP and ML trees cannot allow to conclusively assign isolate CO33 to a subspecific clade, as recombination studies were not addressed in the present work.

Comparisons of the whole genome sequences of CoDiRO with the other ST53 isolates confirmed its close relatedness with isolate COF407 (Fig. 2). CoDiRO and the Costa Rican isolate COF407 have the highest sequence identity (Fig. 2) in paired alignment, whereas the other two isolates from oleander, OLS478 and OLS479, showed a higher number of sequence regions having lower identity. The circular representation (Fig. 2) reports the six regions identified as prophage or phage remnants in the CoDiRO genome identified by Phaster.

**Genomic elements characterizing CoDiRO and Costa Rican isolates.** In silico analysis revealed that all ST53 isolates harbor a common plasmid. Specifically, CoDiRO has an approximate 35-kb plasmid (Fig. 3A) sharing 99% nucleotide sequence identity and respectively, 99%, 92% and 87% coverage with the corresponding plasmids described in the isolates COF407 (pXF-P4.OLS0479), OLS0478 (pXF-P4.OLS0479), and OLS0479 (pXF-P4.COF407) (Table 1). The common 35-kb plasmid was found to be similar to the conjugative plasmid pXF-RIV5 (NC020121.1) found in subsp. *multiplex* and almost identical to pXFAS01 (NC010579.1) described in the M23 strain from subsp. *fastidiosa*

(Rogers and Stenger 2012). The highest (98%) identity of the CoDiRO plasmid with pXF-RIV5/pXFAS01 refers to protein components of the type IV secretion system (T4SS), responsible for DNA and protein translocation to the extracellular environment, to eukaryotic cells during bacterial infections, or to other bacterial cells (Alvarez-Martinez and Christie 2009). As such, this system is mainly involved in DNA transfer among bacteria by conjugative plasmids or genome-integrated DNA that excises and circularizes for translocation. T4SS genes necessary for conjugative transfer (*tra*) and mating pair formation (*trb*) are present in CoDiRO (Fig. 3A) as well as genes responsible for replication (*repA*, *kleE*, and *ssBP*) and plasmid partitioning (*parA* and *parB*), similarly present in pXF-RIV5. This homology is interrupted in the sequence region located between the Tra/Trb protein modules (Fig. 3A), which is considered a site of insertion of gene accessory modules occurring in pXF-RIV5, between transposon and resolvase genes (Rogers and Stenger 2012). Indeed, an in-depth investigation of the CoDiRO plasmid sequence reveals the presence of a partial resolvase CDS, at nucleotides 17,740 to 18,216, that lacks a methionin start codon and a complete CDS encoding a TnpA transposase between nucleotides 21,384 and 21,869 (Supplementary File S1). A 100% nucleotide identical resolvase and TnpA transposase are present in the pXF-P4.OLS0479 plasmid from strain COF407.

RAST annotation of this region in the CoDiRO plasmid containing gene accessory modules shows that mapped CDSs encode proteins



**Fig. 1.** Genetic relationships among ST53 isolates and *Xylella fastidiosa* subspecies. **A**, A maximum-parsimony tree was reconstructed using 119,685 single nucleotide polymorphisms (SNPs) from 27 draft or finished whole-genome sequences. The tree is unrooted. Main *X. fastidiosa* subspecies are indicated. Numbers at each node indicate SNPs common to descendants. Numbers after the isolate name (isolate name number) indicate SNPs exclusive of the isolate. **B**, A maximum likelihood tree was generated based on the general time-reversible model of nucleotide evolution and discrete gamma distribution. The percentage of trees in which the associated taxa clustered together from a total of 100 bootstrap experiments is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 27 nucleotide sequences. There was a total of 33,016 positions in the final dataset. The roots have been placed arbitrarily.

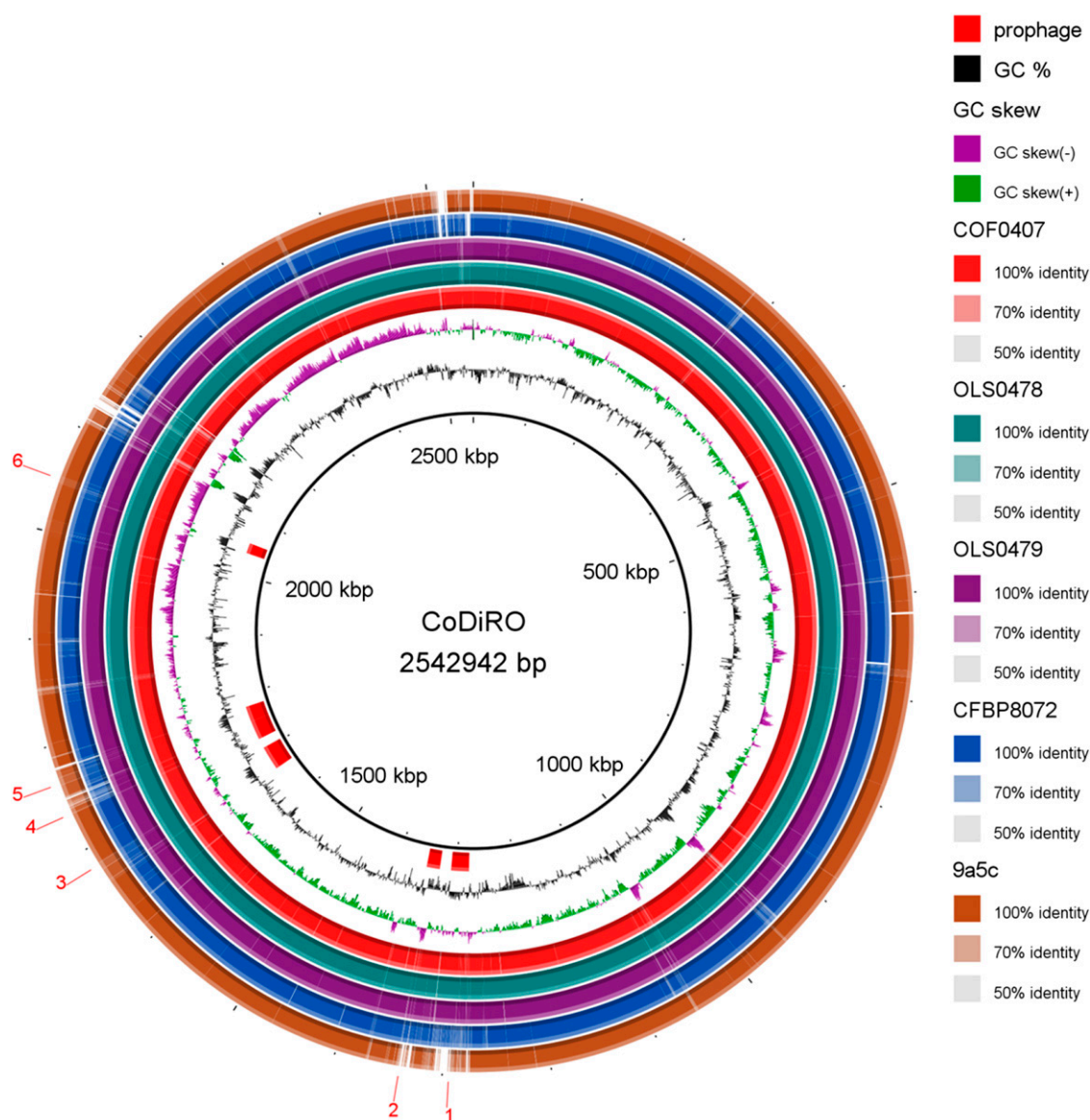
homologs to the *parD/parE* toxin/antitoxin system (TA) (Roberts et al. 1994). This system induces postsegregational killing of cells that have not received such plasmid after cell division, therefore ensuring plasmid transmission to the descent. Moreover, a second module encoding a YdcE-YdcD TA system (Pellegrini et al. 2005) is present in this region of the CoDiRO plasmid, whose role is to induce cells to enter in a dormant persistent state of low metabolic activity when subjected to diverse stresses (Merfa et al. 2016; Muranaka et al. 2012; Muthuramalingam et al. 2016). An addition module annotated by RAST as a transcriptional regulator of the ArsR family is present between the two components of the YdcE-YdcD TA system. As shown in Figure 3B (light gray/pink CDSs), a sequence containing the same gene accessory modules is present in pXF-P4.OLS0479 plasmid from strain COF407.

**Pan-genome analysis.** The *X. fastidiosa* pan-genome (Tettelin et al. 2005) was investigated using GET\_HOMOLOGS to obtain the genes repertoire of the species. The software identifies orthologous genes in a set of bacterial genomes and groups them in clusters (i.e., genes or protein clusters). Intrinsic in the pan-genome concept

is the distinction of a “core” set of genes shared by all isolates and a “accessory” set of genes, which are present in one/few members. This accessory genome may provide information on adaptations to different environments or virulence factors. Considering the *X. fastidiosa*–olive association as a new pathosystem, one of the aims of the study was to identify ST53 and/or CoDiRO exclusive genes in addition to support the clustering of the ST53 isolates.

GET\_HOMOLOGS intersects data from a “tree-based” (COG) algorithm and a “pair-linking” (OrthoMCL) method to identify clusters of orthologous genes and proteins starting from input files in the GenBank format (Kristensen et al. 2010). A preliminary step of reannotation of all 27 genomes with RAST increased the number of orthologous genes and proteins compared with the original annotated genomes (data not shown). The intersection of the proteins obtained with the COG and OMCL algorithms of the 27 genomes produced a pan-genome containing 6,869 protein clusters (Fig. 4A).

The partition of the pan-genome showed that the largest number of protein clusters (2,680) was in the “shell” (i.e., exclusive of a

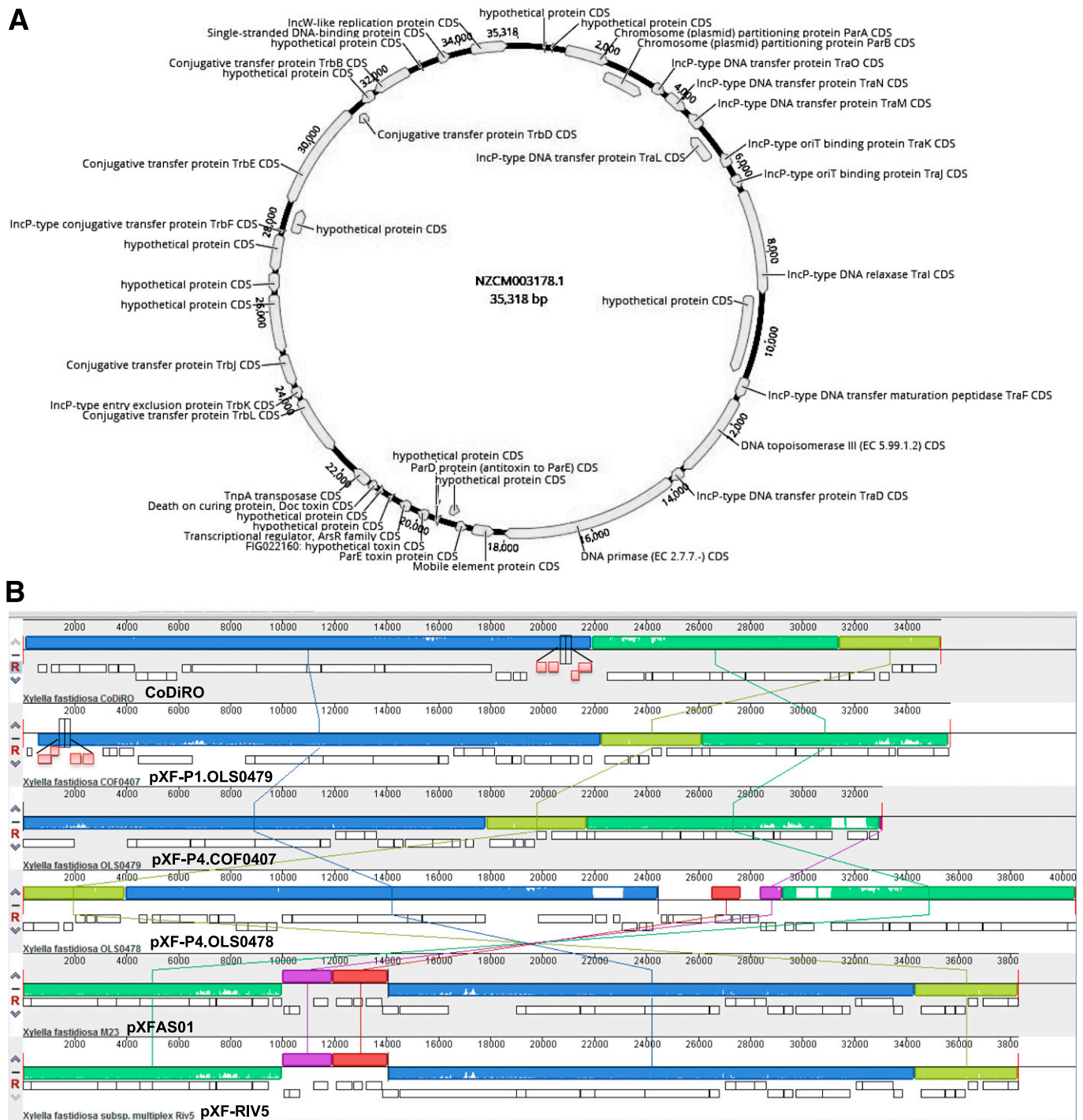


**Fig. 2.** Whole genome comparison of the Costa Rican and the CoDiRO ST53 phylotypes. Genome sequence assemblies were aligned by BLASTN. The innermost ring indicates the draft genome sequence of the isolate CoDiRO (accession number JIJW01), used as reference. Moving outward, the next rings are as follows: a plot of the G + C content; a plot of the GC skew  $(G - C)/(G + C)$  averaged over a moving window of 10,000 bp, with excess G and excess C shown in varying shades of gray. The remaining four rings correspond to the three ST53 Costa Rican draft assemblies and the complete *pauca* 9a5c genome. Low levels of sequence identity are indicated by blurred shades. Whitespace gaps represent genomic regions not covered by the BLASTN alignments. The software arbitrarily assembles the scaffolds of the draft genome sequences of the ST53 isolates and the complete *pauca* 9a5c genome, which contain the plasmids sequences.

limited number of genomes) category of proteins scattered in 3 to 24 genomes followed by a “cloud” category of 2,668 protein clusters (i.e., present in only few genomes), which are present in 2 or less of the 27 genomes (Fig. 4B). The “soft core” (i.e., identified in 95% of isolates) category contains 1,521 protein clusters, which are present in 25 or 26 genomes, whereas protein clusters common to all 27 genomes and composing the “core” class are 1,269 (Fig. 4B). A heatmap and a dendrogram generated using an average nucleotide identity matrix based on the CDSs of the four pan-genome clusters (Fig. 4C) showed that a topology consistent with the *X. fastidiosa* phylogeny observed in MP and ML tree is obtained by the analysis

of the 27 bacterial pan-genomes with few differences. Isolates of the subspecies *multiplex* cluster distinctively from those of subsp. *fastidiosa*, *morus*, and *sandyi* and, as observed in the ML tree, isolate CO33 groups in the subsp. *sandyi*. In addition, two separate clusters are identified in the subsp. *pauca*, which contain Brazilian isolates 9a5c, COF0324, 6c, 32c, CVC0251, and CVC0256, and the isolates OLS0478, OLS0479, COF0407, and CoDiRO from Costa Rica, and CFBP8072 from Ecuador (Fig. 4C).

**Genes unique to the ST53 clade.** The interrogation of these 6,869 protein clusters produced a list of 51 proteins encoded by the four draft genomes of the ST53 clade but not in the remaining

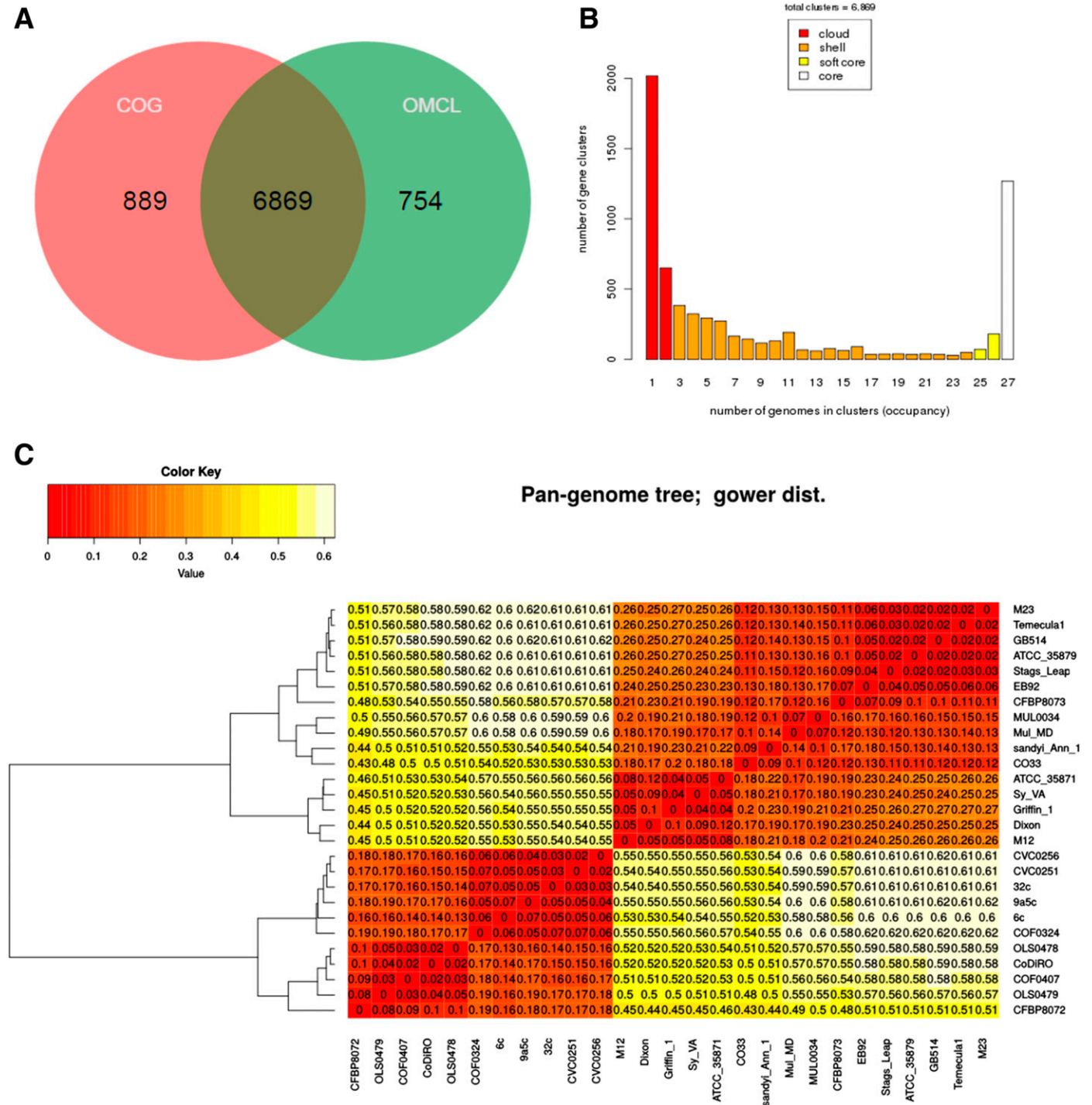


**Fig. 3. A,** Annotation map of CoDiRO unnamed plasmid (NZCM003178.1). Map shows the RAST annotations stored in GenBank file and toxin/antitoxin sequence region between nucleotides 17,740 and 21,869. **B,** Progressive Mauve multiple alignment of plasmids of the four ST53 isolates, pXFAS01, and pXF-RIV5. The ParD/ParE containing sequence, additionally present in CoDiRO and pXF.P1.OLS0479 is indicated by gray boxes.

23 genomes (Supplementary Table S1). However, a detailed BLASTP analysis disclosed that only 1 out of the 51 proteins was exclusively present in the ST53 clade, whereas 50 were found in some of the remaining 23 genomes. This unique 663 amino acid-long protein was homologous to several bacterial ATP-binding proteins belonging to the family of histidine kinase-like ATPase. In bacteria, these transmembrane proteins have the role of signal transduction in response to environmental stresses (Capra and Laub 2012).

The nucleotide sequence encoding the ATP-binding protein is flanked in the genomes of CoDiRO, OLS0478 and partially in that

of COF0407 by prophage and other conserved CDSs (Fig. 5), while this information is incomplete in the OLS0479 genome. A primer set (ATPase for1/ATPase rev2) was designed on the alignment of the available ST53 genomes (Fig. 5) to verify the exclusive presence of the ATP-binding protein in ST53 isolates and to exclude its presence in other *X. fastidiosa* subspecies (Table 2). PCR analysis showed that a DNA fragment of the expected size (~2,876 bp) was correctly amplified from the CoDiRO strain from olive (Fig. 6A, lane 1), six Costa Rican isolates from oleander (Fig. 6B, No3 lane 1 and 4, No1 lane 5, No5 lane 6, No2 lane 10, No4 lane 11, and No7 lane 12), and



**Fig. 4.** Pan-genome analysis of the 27 *Xylella fastidiosa* genomes. **A**, Venn diagram showing the overlap between clusters of orthologous sequences computed by COG and OMCL algorithms; **B**, distribution of the pan-genome matrix into core, soft core, cloud, and shell partitions; and **C**, heatmap and dendrogram obtained using an average nucleotide identity matrix based on the coding sequences composing the pan-genome. Numbers refer to dissimilarity values (Gower's distance) among the 27 pan-genomes calculated on the average nucleotide identity matrix.



two Costa Rican isolates of subsp. *fastidiosa* isolated from coffee and having the ST21 phylotype (Fig. 6B, Ca4 lane 3, Ca7 lane 8). However, two additional ST18 subsp. *fastidiosa* isolates from grapevine (Fig. 6B, Vv2 lane 2, Vv1 lane 7) and a ST33 subsp. *fastidiosa* from coffee (Fig. 6B, Ca5 lane 9) of the same country were negative. Nucleotide sequencing of the obtained PCR product from the CoDiRO strain confirmed it corresponds to the expected sequence. This DNA fragment was not amplified from total DNA of representative isolates of the subspecies *multiplex* (Fig. 6A, Dixon lane 4, AlmaEm3 lane 6, BB08-1 lane 7), *fastidiosa* (Fig. 6A, Temecula lane 3), *sandyi* (Fig. 6A, Sandyi lane 5), and an undetermined subspecies (Fig. 6A, Co33 lane 9), whereas an approximate 700-bp DNA band was observed in 12 out of 21 isolates of the subsp. *pauca* (Fig. 6A, J4 lane 8; Fig. 6C, J1 lane 2, and Table 2). BLASTN search showed that the 692 bp-long sequences of subsp. *pauca* isolate J2 (Fig. 6A, lane 2; accession number KY271786) and J4 (Fig. 6A, lane 8; accession number KY271787) have 99% nucleotide identity and 100% coverage with a genome region of the subsp. *pauca* isolate CVC0256. This sequence belongs to contig 16 (GenBank LRVF01000016.1) and spans from a hypothetical CDS annotated as Figure 012199071 (CDS A in Fig. 5) to a phage-related protein (CDS F in Fig. 5), both present in the CoDiRO genome. CDSs F, which contains primers ATPase for1, is missing in the 9a5c genome therefore explaining the failure in amplification. Besides this approximate 700-bp sequence that in 9a5c is replaced by four CDSs (CDSs p, q, r, and s in Figure 5), genome organization and CDSs of this region of 9a5c genome are highly identical to that of the CoDiRO genome (Fig. 5 and Supplementary Table S2).

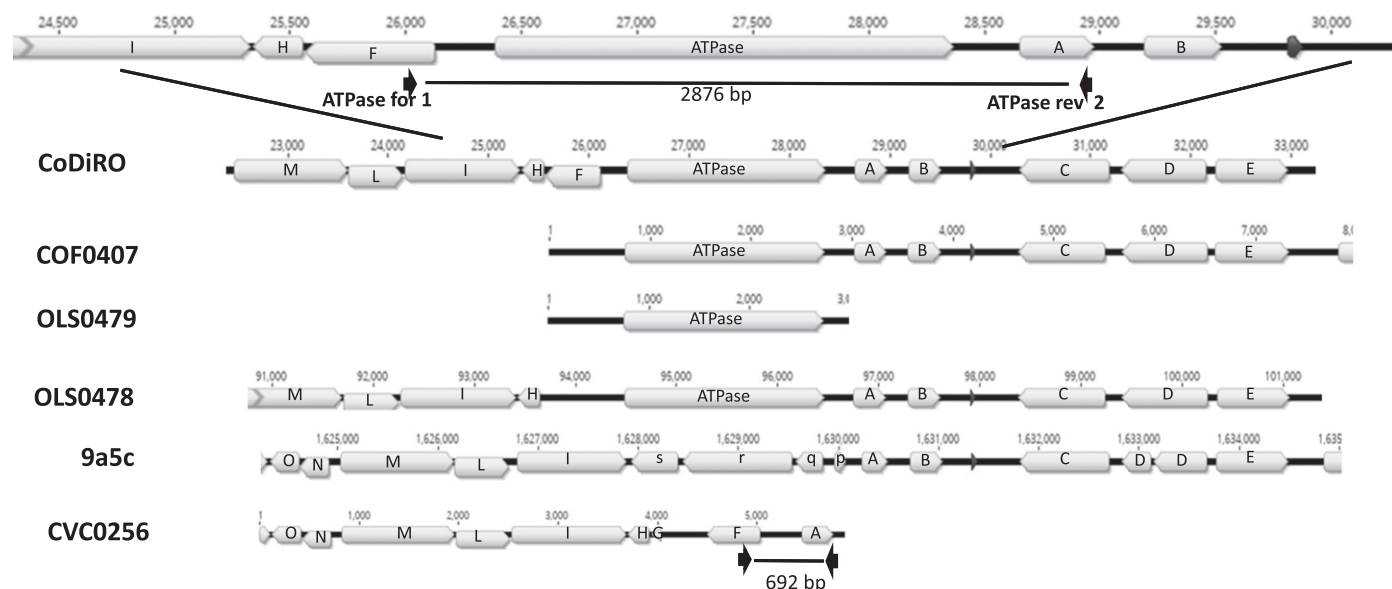
## DISCUSSION

A comprehensive whole-genome study of 27 *X. fastidiosa* genomes was performed by SNP identification and analysis of the gene repertoire (pan-genome) of this bacterial species. The whole-genome SNP approach agrees with the known *X. fastidiosa* phylogeny, although depicting a more complex scenario with the inclusion of the recently acquired genomes described in Costa Rica or intercepted in Europe. Obtained MP and ML phylogenetic trees using data from SNP analysis basically overlaps except for the

placement of the new isolate CO33 (Giampetruzzi et al. 2015b), which clusters with isolate sandyi Ann 1 in the subsp. *sandyi* in ML tree and confirms the observed presence of blurred boundaries between the subsp. *fastidiosa* and *sandyi* (Loconsole et al. 2016). Both MP and ML trees distinguish, in the subsp. *fastidiosa*, a compact cluster including U.S. isolates EB92, ATCC 35879, Stags Leap, GB514, M23, and Temecula 1 from the recently described Mexican isolate CFBP8073 (Jacques et al. 2016). Whereas in line with its initial description (Jacques et al. 2016), the recombinant isolate CFBP8072 is grouped in the subsp. *pauca* in the phylogenetic trees obtained by SNPs and pan-genome analysis. The 27 analyzed genomes do not have comparable qualities, as many of them are assemblies generated by diverse technologies of sequencing and assembly. Nevertheless, even considering potential errors caused by this lack of homogeneity, the present data proved that whole-genome phylogenetic SNP analysis is effective in delineating the genetic relatedness of *X. fastidiosa* isolates.

Phylogenetic studies and the identification of *X. fastidiosa*-infected plant material in several consignments originated from Costa Rica and Honduras support the hypothesis that the *X. fastidiosa* ST53 genotype in southern Italy may have originated from Central America (Loconsole et al. 2016). Genetic evidence in support of this assumption, previously limited to MLST datasets, was recently corroborated by average nucleotide identity analysis (Marcelletti and Scortichini 2016). According to Nunney et al. (2014b), ST53 was not native to Central America but originated from the introduction of a single *pauca* genotype from South America. Our SNP and pan-genome analyses support that conclusion, as all the ST53 isolates studied here formed a separate clade associated with subsp. *pauca*. This clade is separate from other subsp. *pauca* isolates including South American isolates, as well as from isolate CFB8072 from Ecuador. The limited genetic diversity among isolates of this new clade suggests a recent origin from a common ancestor as there has not been enough time for genetic diversification. Alternatively, it is possible that ST53 is particularly virulent and therefore is overrepresented in the obtained isolates from Costa Rica since it has been obtained from several hosts; or, finally, the number of available sequences from Central America is still very limited, potentially leading to biased results.

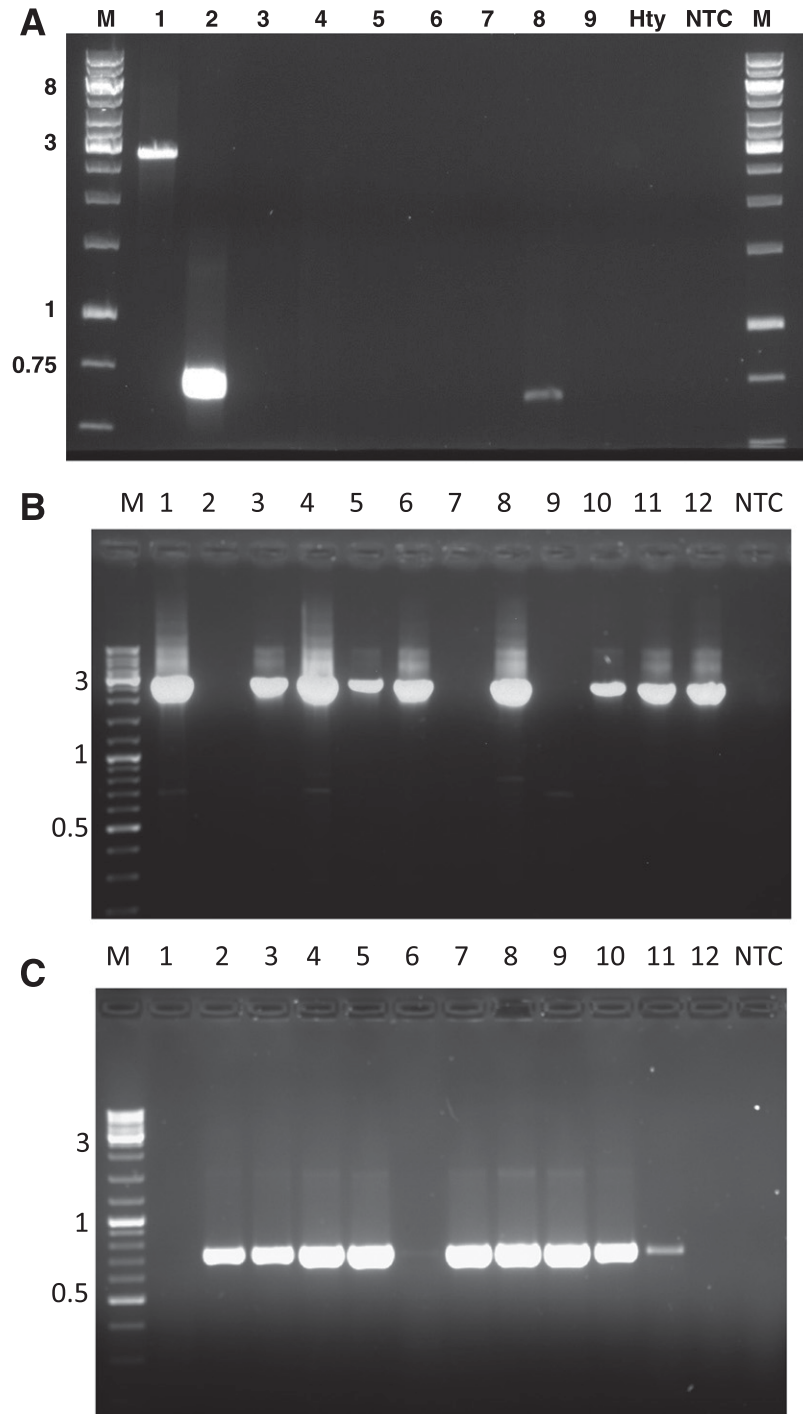
The presence of an identical plasmid in all four isolates supports their common ancestry as distinct from the rest of the *pauca*



**Fig. 5.** Graphical representation of the coding sequences (CDSs) flanking the putative histidine kinase-like ATPase present in all four ST53 isolates. CDS organization with the corresponding regions in 9a5c and CVC0256 genomes is reported for comparison except for the presence of the putative histidine kinase-like ATPase. Similar CDSs are indicated with identical letters. The close-up of the nucleotide region surrounding the putative histidine kinase-like ATPase in CoDiRO is reported to show the location of primers for PCR analysis and the amplified DNA. The position in the CV0256 genome of the 692-bp DNA fragment amplified from subsp. *pauca* isolates J2 and J4 is also shown.

isolates. Moreover, the highest nucleotide identity between the CoDiRO and the COF0407 plasmids is in agreement with the topology obtained by the SNP-based phylogeny. The high nucleotide identity of the ST53 common plasmids with the pXF-RIV5 conjugative plasmid described by Rogers and Stenger (2012) raises the following question regarding their origin. It is known that a condition for conjugation to occur is the close proximity of the donor and receiving bacterial cells, implying that

a strain bearing a pXF-RIV-5-like plasmid would have shared the same ecological niche with an ancestor of the ST53 phylotype. Considering that ST53 isolates have been discovered in Italy and Costa Rica so far, whereas pXF-RIV5 has been described in a North American isolate of the subspecies *multiplex*, the origin of the ST53 plasmids remains unexplained. We propose that pXF-RIV5 occurs in other endemic populations of *X. fastidiosa* in Costa Rica not sampled so far.



**Fig. 6.** Gel electrophoresis showing the products of PCR amplifications of different *Xylella fastidiosa* subspecies with primers ATPase for1/ATPase rev2. Amplifications were from bacterial DNA of isolates: **A**, CoDiRO (lane 1), J2 (lane 2), Temecula (lane 3), Dixon (lane 4), Sandyi (lane 5), AlmaEm3 (lane 6), BB08-1 (lane 7), J4 (lane 8); CO33 (lane 9), total DNA from xylem of healthy olive tree (lane Hty), and no template control (lane NTC); Generuler 1-kb DNA ladder, Thermo Scientific (lane M). **B**, No3 (lane 1), Vv2 (lane 2), Ca4 (lane 3), No3 (lane 43), No1 (lane 5), No5 (lane 6), Vv1 (lane 7), Ca7 (lane 8), Ca5 (lane 9), No2 (lane 10), No4 (lane 11), No7 (lane 12), and no template control (lane NTC); O'Gene-ruler DNA ladder mix, Fermentas, St. Leon-Rot, Germany (lane M). **C**, J33 (lane 1), J1 (lane 2), J41 (lane 3), J44 (lane 4), J48 (lane 5), J68 (lane 6), 6570 (lane 7), J12 (lane 8), J62 (lane 9), J55 (lane 10), J65 (lane 11), 9a5c (lane 12), and no template control (lane NTC). Relevant bands of the Generuler 1-kb DNA ladder (lane M; Thermo Scientific) are reported.

The pan-genome analysis showed that the gene repertoire of the species contains 6,869 gene clusters, a large pool if compared with 2,869 genes, which is the average content of the species (Table 1). The ability of *X. fastidiosa* in acquiring new genes is proved by the high number of “accessory” genes (i.e., 2,668 genes are encoded in only two genomes). Pan-genome analysis of the *X. fastidiosa* genomes led to a clustering of *X. fastidiosa* subspecies, which is consistent with the known phylogeny and the obtained MP and ML trees.

Histidine kinase-like ATPase are widespread in the bacterial kingdom representing a system to perceive and respond to environmental stimuli (Capra and Laub 2012). When stimulated this signal transduction system induces corresponding modification in gene expression allowing bacterial cells to better adapt to changes in external or environmental conditions. The discovery of a putative histidine kinase-like ATPase protein exclusive of the ST53 clade but also detected in ST21 isolates from Costa Rica further confirms the existence of a common evolution history which occurred in this country. The presence of conserved flanking regions in several genomes of the subsp. *pauca* suggests that this CDS may have originated from an event of DNA insertion. This scenario is confirmed by the presence of the same CDS in the ST21 phylotype, which also allows us to exclude the alternative hypothesis that the ST53 genome may represent the ancestor of the subspecies *pauca* that lost this CDS.

Besides demonstrating the common evolution of the ST53 isolates at genome level, which suggests their classification in a separate clade, this study conclusively supports the Central America origin of the *X. fastidiosa* subsp. *pauca* strain CoDiRO, which is causing disease in olive trees in the South of Italy and provides clues in the study of their biological features.

## ACKNOWLEDGMENTS

The work was supported by funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement 635646: POnTE (Pest Organisms Threatening Europe). The equipment used for the next-generation sequencing analysis were from the “Rete di Laboratori Pubblici SELGE-Regione Puglia (cod. 14).” We thank L. De La Fuente, Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, and J. R. S. Lopes, Departamento de Entomologia e Acarologia, ESALQ, Universidade de São Paulo, Piracicaba, São Paulo 13418-970, Brazil for providing isolates of *Xylella fastidiosa*.

## LITERATURE CITED

- Alencar, V. C., Barbosa, D., Santos, D. S., Oliveira, A. C. F., de Oliveira, R. C., and Nunes, L. R. 2014. Genomic sequencing of two coffee-infecting strains of *Xylella fastidiosa* isolated from Brazil. *Genome Announc.* 2:e01190-13.
- Alikhan, N.-F., Petty, N. K., Zakour, N. L. B., and Beatson, S. A. 2011. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* 12:402.
- Almeida, R. P. P., and Nunney, L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? *Plant Dis.* 99:1457-1467.
- Alvarez-Martinez, C. E., and Christie, R. J. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* 73:775-808.
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D. S. 2016. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44:W16-21.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. 2008. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- Baker, R., Bragard, C., Caffier, D., Candresse, T., Gilioli, G., Grégoire, J. C., Holb, I., Jeger, M. J., Evtimova Karadjova, O., Magnusson, C., Makowski, D., Manceau, C., Navajas, M., Rafoss, T., Rossi, V., Schans, J., Schrader, G., Urek, G., Vloutoglou, I., Winter, S., and van der Werf, W. 2015. Scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. *EFSA J.* 13:3989.
- Capra, E. J., and Laub, M. T. L. 2012. The evolution of two-component signal transduction systems. *Annu. Rev. Microbiol.* 66:325-347.
- Chen, J., Huang, H., Chang, C. J., and Stenger, D. C. 2013. Draft genome sequence of *Xylella fastidiosa* subsp. *multiplex* strain Griffin-1 from *Quercus rubra* in Georgia. *Genome Announc.* 1(5):e00756-13.
- Chen, J., Wu, F., Zheng, Z., Deng, X., Burbank, L. P., and Stenger, D. C. 2016. Draft genome sequence of *Xylella fastidiosa* subsp. *fastidiosa* strain Stag’s Leap. *Genome Announc.* 4:e00240-16.
- Chen, J., Xie, G., Han, S., and Civerolo, E. L. 2010. Two whole genome sequences of *Xylella fastidiosa* (strains M12 and M23) causing almond leaf scorch disease in California. *J. Bacteriol.* 192:4534.
- Contreras-Moreira, B., and Vinuesa, P. 2013. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79:7696-7701.
- Doddapaneni, H., Yao, J., Lin, H., Walker, M. A., and Civerolo, E. L. 2006. Analysis of the genome-wide variations among multiple strains of the plant pathogenic bacterium *Xylella fastidiosa*. *BMC Genomics* 7:225.
- EFSA PLH Panel. 2016. EFSA Panel on Plant Health. Statement on diversity of *Xylella fastidiosa* subsp. *pauca* in Apulia. *EFSA J.* 14:e04542.
- Gardner, S. N., Slezak, T., and Hall, B. G. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31:2877-2878.
- Giampetruzzi, A., Chiumenti, M., Saponari, M., Donvito, G., Italiano, A., Loconsole, G., Boscia, D., Cariddi, C., Martelli, G. P., and Saldarelli, P. 2015a. Draft genome sequence of the *Xylella fastidiosa* CoDiRO strain. *Genome Announc.* 3:e01538-14.
- Giampetruzzi, A., Loconsole, G., Boscia, D., Calzolari, A., Chiumenti, M., Martelli, G. P., Saldarelli, P., Almeida, R. P., and Saponari, M. 2015b. Draft genome sequence of CO33, a coffee-infecting isolate of *Xylella fastidiosa*. *Genome Announc.* 3:e01472-15.
- Guan, W., Shao, J., Davis, R. E., Zhao, T., and Huang, Q. 2014a. Genome sequence of a *Xylella fastidiosa* strain causing sycamore leaf scorch disease in Virginia. *Genome Announc.* 2:e00773-14.
- Guan, W., Shao, J., Zhao, T., and Huang, Q. 2014b. Genome sequence of a *Xylella fastidiosa* strain causing mulberry leaf scorch disease in Maryland. *Genome Announc.* 2:e00916-13.
- Hendriksen, R. S., Price, L. B., Schupp, J. M., Gillece, J. D., Kaas, R. S., Engelthaler, D. M., Bortolaia, V., Pearson, T., Waters, A. E., Upadhyay, B. P., Shrestha, S. D., Adhikari, S., Shakya, G., Keim, P. S., and Aarestrup, F. M. 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the origin of the Haitian outbreak. *MBio* 2:e00157-11.
- Jacques, M. A., Denancé, N., Legendre, B., Morel, E., Briand, M., Mississippi, S., Durand, K., Olivier, V., Portier, P., Poliakov, F., and Crouzillat, D. 2016. New coffee plant-infecting *Xylella fastidiosa* variants derived via homologous recombination. *Appl. Environ. Microbiol.* 82:1556-1568.
- Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V., and Mushegian, A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* 26:1481-1487.
- Kumar, S., Stecher, G., and Tamura, K. 2016. Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870-1874.
- Li, L., Stoeckert, C. J., and Roos, D. S. 2003. OrthoMC: Identification of ortholog groups for eukaryotic genomes. *Genet. Res.* 81:2178-2189.
- Loconsole, G., Saponari, M., Boscia, D., D’Attoma, G., Morelli, M., Martelli, G., and Almeida, R. P. P. 2016. Intercepted isolates of *Xylella fastidiosa* in Europe reveal novel genetic diversity. *Eur. J. Plant Pathol.* 146:85-94.
- Marcelletti, S., and Scortichini, M. 2016. The *Xylella fastidiosa* CoDiRO strain associated with the olive quick decline syndrome in southern Italy belongs to a clonal complex of the subspecies *pauca* that evolved in Central America. *Microbiology* 162:2087-2098.
- Merfa, M. V., Niza, B., Takita, M. A., and De Souza, A. A. 2016. The MqsRA toxin-antitoxin system from *Xylella fastidiosa* plays a key role in bacterial fitness, pathogenicity, and persister cell formation. *Front. Microbiol.* 7:904.
- Muranaka, L. S., Takita, M. A., Olivato, J. C., Kishi, L. T., and DeSouza, A. A. 2012. Global expression profile of biofilm resistance to antimicrobial compounds in the plant-pathogenic bacterium *Xylella fastidiosa* reveals evidence of persister cells. *J. Bacteriol.* 194:4561-4569.
- Muthuramalingam, M., White, J. C., and Bourne, C. R. 2016. Toxin-antitoxin modules are pliable switches activated by multiple protease pathways. *Toxins (Basel)* 8:214.
- Nei, M., and Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Nunes, L. R., Rosato, Y. B., Muto, N. H., Yanai, G. M., da Silva, V. S., Leite, D. B., Gonçalves, E. R., de Souza, A. A., Coletta-Filho, H. D., Machado, M. A., Lopes, S. A., and de Oliveira, R. C. 2003. Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements. *Genome Res.* 13:570-578.

- Nunney, L., Hopkins, D. L., Morano, L. D., Russell, S. E., and Stouthamer, R. 2014a. Intersubspecific recombination in *Xylella fastidiosa* strains native to the United States: Infection of novel hosts associated with an unsuccessful invasion. *Appl. Environ. Microbiol.* 80:1159-1169.
- Nunney, L., Ortiz, B., Russell, S. A., Sánchez, R. R., and Stouthamer, R. 2014b. The complex biogeography of the plant pathogen *Xylella fastidiosa*: Genetic evidence of introductions and subspecific introgression in Central America. *PLoS One* 9:e112463.
- Nunney, L., Schuenzel, E. L., Scally, M., Bromley, R. E., and Stouthamer, R. 2014c. Large-scale intersubspecific recombination in the plant-pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. *Appl. Environ. Microbiol.* 80:3025-3033.
- Nunney, L., Yuan, X., Bromley, R., Hartung, J., Montero-Astúa, M., Moreira, L., Ortiz, B., and Stouthamer, R. 2010. Population genomic analysis of a bacterial plant pathogen: Novel insight into the origin of Pierce's disease of grapevine in the U.S. *PLoS One* 5:e15488.
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42:D206-D214.
- Pellegrini, O., Mathy, N., Gogos, A., Shapiro, L., and Condon, C. 2005. The *Bacillus subtilis* ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. *Mol. Microbiol.* 56:1139-1148.
- Roberts, R. C., Ström, A. R., and Helinski, D. R. 1994. The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *J. Mol. Biol.* 237:35-51.
- 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.
- Rusconi, B., Sanjar, F., Koenig, S. S., Mammel, M. K., Tarr, P. I., and Eppinger, M. 2016. Whole genome sequencing for genomics-guided investigation of *Escherichia coli* of O157:H7 outbreaks. *Front. Microbiol.* 7:985.
- Saponari, M., Boscia, D., Nigro, F., and Martelli, G. P. 2013. Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy). *J. Plant Pathol.* 95:668.
- Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bové, J. M., Briones, M. R., Bueno, M. R., Camargo, A. A., Camargo, L. E., Carraro, D. M., Carrer, H., Colauto, N. B., Colombo, C., Costa, F. F., Costa, M. C., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A. P., Ferreira, A. J., Ferreira, V. C., Ferro, J. A., Fraga, J. S., França, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuramae, E. E., Laigret, F., Lambais, M. R., Leite, L. C., Lemos, E. G., Lemos, M. V., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M., Madeira, H. M., Marino, C. L., Marques, M. V., Martins, E. A., Martins, E. M., Matsukuma, A. Y., Menck, C. F., Miracca, E. C., Miyaki, C. Y., Monteiro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L., Netto, L. E., Nhani, A., Jr., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A., Pereira, H. A., Jr., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M Rosa, A. J., de Rosa, V. E., Jr., de Sá, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C., da Silva, A. M., da Silva, F. R., da Silva, W. A., Jr., da Silveira, J. F., Silvestri, M. L., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tshako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J., and Setubal, J. C. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature*. 406:151-159.
- Su, C. C., Chang, C. J., Chang, C. M., Shih, H. T., Tzeng, K. C., Jan, F. J., Kao, C. W., and Deng, W. L. 2013. Pierce's disease of grapevines in Taiwan: Isolation, cultivation and pathogenicity of *Xylella fastidiosa*. *J. Phytopathol.* 161:389-396.
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., and Ward, N. L. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci.* 102:13950-13955.
- Van Sluys, M. A., de Oliveira, M. C., Monteiro-Vitorello, C. B., Miyaki, C. Y., Furlan, L. R., Camargo, L. E., da Silva, A. C., Moon, D. H., Takita, M. A., Lemos, E. G., Machado, M. A., Ferro, M. I., da Silva, F. R., Goldman, M. H., Goldman, G. H., Lemos, M. V., El-Dorry, H., Tsai, S. M., Carrer, H., Carraro, D. M., de Oliveira, R. C., Nunes, L. R., Siqueira, W. J., Coutinho, L. L., Kimura, E. T., Ferro, E. S., Harakava, R., Kuramae, E. E., Marino, C. L., Giglioti, E., Abreu, I. L., Alves, L. M., do Amaral, A. M., Baia, G. S., Blanco, S. R., Brito, M. S., Cannavan, F. S., Celestino, A. V., da Cunha, A. F., Fenille, R. C., Ferro, J. A., Formighieri, E. F., Kishi, L. T., Leoni, S. G., Oliveira, A. R., Rosa, V. E., Jr., Sasaki, F. T., Sena, J. A., de Souza, A. A., Truffi, D., Tsukumo, F., Yanai, G. M., Zaros, L. G., Civerolo, E. L., Simpson, A. J., Almeida, N. F., Jr., Setubal, J. C., and Kitajima, J. P. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J. Bacteriol.* 185:1018-26.
- Zhang, S., Flores-Cruz, Z., Kumar, D., Chakrabarty, P., Hopkins, D. L., and Gabriel, D. W. 2011. The *Xylella fastidiosa* biocontrol strain EB92-1 genome is very similar and syntenic to Pierce's disease strains. *J. Bacteriol.* 193:5576-5577.