Title
Genome-wide analysis provides evidence on the genetic relatedness of the emergent xylella fastidiosa genotype in Italy to isolates from Central America

Permalink
https://escholarship.org/uc/item/3x88q8bj

Journal
Phytopathology, 107(7)

ISSN
0031-949X

Authors
Giampetruzzi, A
Saponari, M
Loconsole, G
et al.

Publication Date
2017-07-01

DOI
10.1094/PHYTO-12-16-0420-R

Peer reviewed
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 Pianta, 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 subs. 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 subs. fastidiosa, multiplex, pauca, sandyi, and morus. Within the subs. pauca, the Italian and three Costa Rican isolates, all belonging to ST53, formed a compact 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 subs. 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 Cicaloidea) (Almeida and Nunney 2015). The bacterium is transmitted by a large number of plants, including mulberry; subsp. pauca, sandyi, and morus. Within the subs. pauca, the Italian and three Costa Rican isolates, all belonging to ST53, formed a compact 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 subs. pauca.

In Europe, the bacterium was detected in olive trees (Olea europaea L.) in southern Italy (Apulia region) in 2013, 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.

*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; it 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; Alkan 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 (fastest 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/pa-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 v1.2.2 (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 ATPaselr1 (5′-ACCTGCATAGTTGTGAG-3′) and ATPaserev2 (5′-GCTGCGATAGTCGAGTTCC-3′) targeting the CoDiRO 2876 nucleotide sequence, which contains the putative histidine kinase-like ATPase gene (locus RA12_07795), were located on contig0000005 (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>
<thead>
<tr>
<th>Strain</th>
<th>Host plant</th>
<th>Origin</th>
<th>Collection date</th>
<th>Size (Mb)/CDS</th>
<th>GC%</th>
<th>Accession number (chromosomes/WGS)</th>
<th>Plasmids accession number</th>
<th>Citation or release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>9a5c</td>
<td><em>Citrus sinensis</em></td>
<td>São Paulo, Brazil</td>
<td>1992</td>
<td>2.7317/2.982</td>
<td>52.64</td>
<td>NC002488.3</td>
<td>pXF1.3: NC002489.3; pXF51: NC002490.1</td>
<td>Simpson et al. 2000</td>
</tr>
<tr>
<td>Dixon</td>
<td><em>Prunus dulcis</em></td>
<td>California, USA</td>
<td>na</td>
<td>2.6223/2.914</td>
<td>52</td>
<td>AAAL02</td>
<td></td>
<td>7/10/2002</td>
</tr>
<tr>
<td>Temecula</td>
<td><em>Vitis vinifera</em></td>
<td>California, USA</td>
<td>1998</td>
<td>2.5211/2.757</td>
<td>51.80</td>
<td>NC000456.1</td>
<td>pXPD1.3: NC000455.1</td>
<td>Van Sluys et al. 2003</td>
</tr>
<tr>
<td>M12</td>
<td><em>P. dulcis</em></td>
<td>California, USA</td>
<td>2003</td>
<td>2.4751/3.706</td>
<td>51.9</td>
<td>NC010513.1</td>
<td></td>
<td>Chen et al. 2010</td>
</tr>
<tr>
<td>M23</td>
<td><em>P. dulcis</em></td>
<td>California, USA</td>
<td>2003</td>
<td>2.5739/2.803</td>
<td>51.76</td>
<td>NC010577.1</td>
<td>pXFS01: NC010579.1</td>
<td>Chen et al. 2010</td>
</tr>
<tr>
<td>GB514</td>
<td><em>Vitis vinifera</em></td>
<td>Texas, USA</td>
<td>na</td>
<td>2.5173/2.835</td>
<td>51.77</td>
<td>NC017562.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EB92.1</td>
<td><em>Sambucus nigra</em></td>
<td>USA</td>
<td>1992</td>
<td>2.4754/2.750</td>
<td>51.5</td>
<td>AFDJ01</td>
<td></td>
<td>Zhang et al. 2011</td>
</tr>
<tr>
<td>ATCC 35871</td>
<td><em>P. salicina</em></td>
<td>Georgia, USA</td>
<td>na</td>
<td>2.4162/2.676</td>
<td>51.7</td>
<td>AUAJ01</td>
<td></td>
<td>7/15/2013</td>
</tr>
<tr>
<td>Griffin-1</td>
<td><em>Quercus rubra</em></td>
<td>Georgia, USA</td>
<td>2006</td>
<td>2.3871/2.724</td>
<td>51.7</td>
<td>AVGA01</td>
<td></td>
<td>Chen et al. 2013</td>
</tr>
<tr>
<td>32c</td>
<td><em>Coffeea sp.</em></td>
<td>São Paulo, Brazil</td>
<td>na</td>
<td>2.6075/2.915</td>
<td>52.5</td>
<td>AWYH01</td>
<td></td>
<td>Alencar et al. 2014</td>
</tr>
<tr>
<td>6c</td>
<td><em>Coffeea sp.</em></td>
<td>São Paulo, Brazil</td>
<td>na</td>
<td>2.6002/2.916</td>
<td>52.35</td>
<td>AXBS01</td>
<td></td>
<td>Alencar et al. 2014</td>
</tr>
<tr>
<td>Mul-MD</td>
<td><em>Morus alba</em></td>
<td>Maryland, USA</td>
<td>2011</td>
<td>2.5205/2.838</td>
<td>51.6</td>
<td>AXDP01</td>
<td></td>
<td>6/6/2014</td>
</tr>
<tr>
<td>MUL0034</td>
<td><em>M. alba</em></td>
<td>USA</td>
<td>na</td>
<td>2.6668/2.974</td>
<td>51.97</td>
<td>NZCP0066740.1</td>
<td>Unnamed1: NZCP0066739.1</td>
<td>Guan et al. 2014b</td>
</tr>
<tr>
<td>Sycamore</td>
<td><em>Platanus occidentalis</em></td>
<td>Virginia, USA</td>
<td>2002</td>
<td>2.4758/2.758</td>
<td>51.6</td>
<td>JMH01</td>
<td></td>
<td>Guan et al. 2014a</td>
</tr>
<tr>
<td>Sy-V</td>
<td><em>Vitis vinifera</em></td>
<td>Florida, USA</td>
<td>1987</td>
<td>2.5223/3.875</td>
<td>51.8</td>
<td>JQAP01</td>
<td></td>
<td>10/21/2014</td>
</tr>
<tr>
<td>ATCC 35879</td>
<td><em>Olea europaea</em> (Vinca minor)</td>
<td>Italy</td>
<td>2015</td>
<td>2.68193</td>
<td>51.7</td>
<td>LJZW01</td>
<td></td>
<td>Giampetruzzi et al. 2015a</td>
</tr>
<tr>
<td>CoDiRO</td>
<td><em>Coffeea arabica</em></td>
<td>Costa Rica</td>
<td>2015</td>
<td>2.68193</td>
<td>51.7</td>
<td>LJZW01</td>
<td></td>
<td>Giampetruzzi et al. 2015b</td>
</tr>
<tr>
<td>CFBP8072</td>
<td><em>Coffeea arabica</em></td>
<td>Ecuador</td>
<td>2012</td>
<td>2.4966/2.763</td>
<td>51.9</td>
<td>LKDK01</td>
<td></td>
<td>Jacques et al. 2016</td>
</tr>
<tr>
<td>CFBP8073</td>
<td><em>Coffeea canephora</em></td>
<td>Mexico</td>
<td>2012</td>
<td>2.58215/2.840</td>
<td>51.6</td>
<td>LKES01</td>
<td></td>
<td>Jacques et al. 2016</td>
</tr>
<tr>
<td>OLS0479</td>
<td><em>Nerium oleander</em></td>
<td>Costa Rica, San Jose Province</td>
<td>2011</td>
<td>2.53996/2.827</td>
<td>51.93</td>
<td>LRVH01</td>
<td></td>
<td>2/5/2016</td>
</tr>
<tr>
<td>CVC0256</td>
<td><em>Citrus sinensis</em></td>
<td>São Paulo, Brazil</td>
<td>1999</td>
<td>2.70214/3.045</td>
<td>52.51</td>
<td>LRVF01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS0478</td>
<td><em>Nerium oleander</em></td>
<td>Costa Rica, San Jose Province</td>
<td>2011</td>
<td>2.55541/2.861</td>
<td>51.91</td>
<td>LRVI01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVC0251</td>
<td><em>Citrus sinensis</em></td>
<td>São Paulo, Brazil</td>
<td>1999</td>
<td>2.74025/3.103</td>
<td>52.51</td>
<td>LRVE01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COF0324</td>
<td><em>Coffeea sp.</em></td>
<td>Minas Gerais, Brazil</td>
<td>2006</td>
<td>2.77256/3.143</td>
<td>52.43</td>
<td>LRVG01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COF0407</td>
<td><em>Coffeea sp.</em></td>
<td>Costa Rica, San Jose Province</td>
<td>2009</td>
<td>2.53847/2.837</td>
<td>51.84</td>
<td>LRVJ01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 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 \textit{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 \textit{X. fastidiosa} genomes. A total of 119,685 SNPs were identified in the 27 \textit{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 \textit{X. fastidiosa} isolates consistently with the current phylogenetic results obtained with MLST, identifying main clusters for the subsp. \textit{fastidiosa}, \textit{multiplex}, \textit{pauca}, and \textit{sandyi} (Fig. 1A). The phylogenetic placement of the isolate CO33, which is basal to the \textit{fastidiosa} cluster and close to the subsp. \textit{sandyi} (i.e., sandyi Ann I in Figure 1A), confirms the already reported phylogeny (Loconsole et al. 2016), whereas isolate CFBP8073 groups with the subsp. \textit{fastidiosa} (Jacques et al. 2016). In addition, the SNP-based study supported the separate clustering of the subsp. \textit{morus} from subsp. \textit{multiplex} and subsp. \textit{fastidiosa}, in agreement with the suggested intersubspecific homologous recombination origin of this group (Nunney et al. 2014c).

Isolates of the subsp. \textit{pauca} shared 12,940 SNP alleles (i.e., meaning that there are 12,940 SNP loci for which all subsp. \textit{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. \textit{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.

\begin{table}[h]
\centering
\caption{List of \textit{Xylella fastidiosa} isolates tested with primers ATPase for 1/ATPase rev2 with relevant information regarding their origin, host plant, and phylotype\textsuperscript{a}}
\begin{tabular}{lllll}
\hline
Isolate & Subspecies & Phylotype & Host & Country & PCR (bp) approximate \\
\hline
J4 & pauca & Unknown & Coffee & Brazil & 700 weak \\
J41 & pauca & Unknown & Citrus & Brazil & 700 \\
J48 & pauca & Unknown & Coffee & Brazil & 700 \\
J68 & pauca & Unknown & Coffee & Brazil & 700 weak \\
J12 & pauca & Unknown & Citrus & Brazil & 700 \\
J62 & pauca & Unknown & Citrus & Brazil & 700 \\
J55 & pauca & Unknown & Citrus & Brazil & 700 \\
J65 & pauca & Unknown & Citrus & Brazil & 700 \\
J1 & pauca & Unknown & Citrus & Brazil & 700 \\
J44 & pauca & Unknown & Citrus & Brazil & 700 \\
J2 & pauca & Unknown & Citrus & Brazil & 700 \\
6570 & pauca & Unknown & Citrus & Brazil & 700 \\
J33 & pauca & Unknown & Coffee & Brazil & – \\
Dixon & multiplex & ST6 & Almond & USA & – \\
BB08-1 & multiplex & Unknown & Blueberry & USA & – \\
Temecula & fastidiosa & ST1 & Grape & USA & – \\
AlmaEm3 & multiplex & Unknown & Blueberry & USA & – \\
Sandyi & sandyi & Unknown & Oleander & USA & – \\
M23 & fastidiosa & ST1 & Almond & USA & – \\
9a5c & pauca & ST13 & Citrus & Brazil & – \\
CoDiRO & pauca & ST53 & Olive & Italy & 2,800 \\
No1 & pauca & ST53 & Oleander & Costa Rica & 2,800 \\
No2 & pauca & ST53 & Oleander & Costa Rica & 2,800 \\
No3 & pauca & ST53 & Oleander & Costa Rica & 2,800 \\
No4 & pauca & ST53 & Oleander & Costa Rica & 2,800 \\
No5 & pauca & ST53 & Oleander & Costa Rica & 2,800 \\
No7 & pauca & ST53 & Oleander & Costa Rica & 2,800 \\
Vv1 & fastidiosa & ST18 & Grapevine & Costa Rica & – \\
Vv2 & fastidiosa & ST18 & Grapevine & Costa Rica & – \\
Ca4 & fastidiosa & ST21 & Coffee & Costa Rica & 2,800 \\
Ca7 & fastidiosa & ST21 & Coffee & Costa Rica & 2,800 \\
Ca5 & fastidiosa & ST33 & Coffee & Costa Rica & – \\
CO33 & undetermined & ST72 & Coffee & Costa Rica & – \\
\hline
\end{tabular}
\footnotesize{\textsuperscript{a} The size of the amplified product is indicated. –, no amplification.}
\end{table}
Indeed, the ML tree shows that these two isolates cluster together in a strongly supported subs. *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 COF0407 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 COF0407 (pXF-P4.OLS0479), OLS0478 (pXF-P4.OLS0479), and OLS0479 (pXF-P4.COF0407) (Table 1). The common 35-kb plasmid was found to be similar to the conjugative plasmid pXF-RIV5 (NC020121.1) found in subs. *multiplex* and almost identical to pXFA01 (NC010579.1) described in the M23 strain from subs. *fastidiosa* (Rogers and Stenger 2012). The highest (98%) identity of the CoDiRO plasmid with pXF-RIV5/pXFA01 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 COF0407.

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.

---

820 PHYTOPATHOLOGY
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 addiction 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 JUJW01), 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.
A 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, CVCO251, and CVCO256, 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...
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 assays 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 for 1, 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*
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-RIV5-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 in external or environmental conditions. The discovery of a putative histidine kinase-like ATPase protein exclusive of the ST53 clade but in external or environmental conditions. The discovery of a putative histidine kinase-like ATPase protein exclusive of the ST53 clade but

ACKNOWLEDGMENTS

The work was supported by funding from the European Union’s Horizon 2020 RE-PEST (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


Amaro, S. H. F., Furtado, D., and Almeida, R. P. P. 2016. Intercepted isolates of Xylella 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.

EU territory, with the identification and evaluation of risk reduction options. EFSA J. 13:3989.


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.


