Phytopathologia Mediterranea (2018), 57, 1, 8–25
DOI: 10.14601/Phytopathol_Mediterr-22263

RESEARCH PAPERS

Insights on a founder effect: the case of *Xylella fastidiosa* in the Salento area of Apulia, Italy

Matteo RAMAZZOTTI[1,*], Fabio CIMAGLIA[2,3,*], Antonia GALLO[2], Francesco RANALDI[1], Giuseppe SURICO[3], Giovanni MITA[2], Gianluca BLEVE[2] and Guido MARCHI[3]

[1] Dipartimento di Scienze Biomediche, Sperimentali e Cliniche, Università degli Studi di Firenze, viale G.B. Morgagni 50, 50134, Firenze, Italy
[2] Istituto di Scienze delle Produzioni Alimentari, Consiglio Nazionale delle Ricerche, Lecce, Italy
[3] Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente, Sezione di Patologia Vegetale ed Entomologia, Università degli Studi, Piazzale delle Cascine 28, 50144 Firenze, Italy

Summary. *Xylella fastidiosa* causing disease on different plant species has been reported in several European countries, since 2013. Based on multilocus sequence typing (MLST) results, there is evidence of repeated introductions of the pathogen in Spain and France. In contrast, in the Salento area of Apulia (Puglia) in Southern Italy, the existence of a unique Apulian MLST genotype of *X. fastidiosa*, causing the olive quick decline syndrome (OQDS; also referred to as "CoDiRO" or "ST53") was proven, and this was tentatively ascribed to *X. fastidiosa* subsp. *pauca*. In order to acquire information on intra population diversity European Food Safety Authority (EFSA) has strongly called for the characterization of *X. fastidiosa* isolates from Apulia to produce the necessary data to better understand strain diversity and evolution. In this work, for the first time the existence of sub-variants within a set of 14 "ST53" isolates of *X. fastidiosa* collected from different locations was searched using DNA typing methods targeting the whole pathogen genome. Invariably, VNTR, RAPD and rep-PCR (ERIC and BOX motifs) analyses indicated that all tested isolates possessed the same genomic fingerprint, supporting the existence of predominant epidemiological strain in Apulia. To further explore the degree of clonality within this population, two isolates from two different Salento areas (Taviano and Ugento) were completely sequenced using PacBio SMRT technology. The whole genome map and sequence comparisons revealed that both isolates are nearly identical, showing less than 0.001% nucleotide diversity. However, the complete and circularized Salento-1 and Salento-2 genome sequences were different, in genome and plasmid size, from the reference strain 9a5c of *X. fastidiosa* subsp. *pauca* (from citrus), and showed a PCR-proved large genome inversion of about 1.7 Mb. Genome-wide indices ANIm and dDDH indicated that the three isolates of *X. fastidiosa* from Salento (Apulia, Italy), namely Salento-1, Salento-2, and De Donno, whose complete genome sequence has been recently released, share a very recent common ancestor. This highlights the importance of continuous and extensive monitoring of molecular variation of this invasive pathogen to understand evolution of adaptive traits, and the necessity for adoption of all possible measures to reduce the risk of new introductions that may augment pathogen diversity.

Key words: whole genomes, phylogeny, olive quick decline syndrome (OQDS), molecular epidemiology.

Data deposition: The complete genome and plasmid sequences of *Xylella fastidiosa* Salento-1 (NCCPB No. 4595 LMG 29352) and Salento-2, as well as their annotations, are deposited at GenBank under accession numbers CP016608, CP016609 and CP016610, CP016611, respectively. The sequence alignments used are available from the corresponding author upon request.

* These authors equally contributed to this article.
Corresponding author: G. Marchi
E-mail: guido.marchi@unifi.it

ISSN (print): 0031-9465
ISSN (online): 1593-2095
www.fupress.com/pm

Introduction

The bacterium *Xylella fastidiosa* (Wells *et al*., 1987) shows many attributes of the ideal plant pathogen. It is a generalist pathogen that includes in its host range many symptomless species or species to which it is only moderately virulent; it is transmitted by numerous xylem-feeding insects which are often themselves not host specific; its life cycle is within the xylem vessels of host plants or within vectors protected from the environment and chemical, biological, and cultural disease management measures; and it is naturally competent (Chatterjee *et al*., 2008; Kung and Almeida 2011; Nunney *et al*., 2013; Purcell, 2013). Mostly based on the major host plants in which *X. fastidiosa* causes disease, and on DNA-DNA hybridization and molecular typing methods, four subspecies of the bacteriumhave been described. These are subspp. *fastidiosa* (main hosts: *Vitis* spp. and *Prunus amygdalus*), *multiplex* (numerous hosts), *sandyi* (main host: *Nerium oleander*) and *pauca* (main hosts: *Citrus* and *Coffea* spp.) (Schaad *et al*., 2004; Euzéby, 2009; Bull *et al*., 2012; Baker *et al*., 2015).

The evolutionary history of *X. fastidiosa* has been inferred by examining large collections of American strains from different hosts. According to MLSA/MLST based analyses in the United States of America (Spratt, 1999; Gevers *et al*., 2005; Jolley and Maiden, 2014), strains of subspp. *fastidiosa* and *sandyi* are clearly distinguishable from one another, and have been commonly isolated from non-native plants. Because they are genetically monomorphic, an allopatric origin of these two subspecies has been suggested, indicating that they have been introduced recently, though independently, into North America (Nunney *et al*., 2010; Yuan *et al*., 2010). It has been speculated that North American populations of subsp. *fastidiosa* descend from one or a few founders infecting coffee plants that were imported to California from Central America (Nicaragua) at the end of the 19th century, near the time (1882) of the first outbreak of Pierce's Disease of grapevine. The geographical origin of *X. fastidiosa* subsp. *sandyi* is currently unknown, but its origin from a tropical region is suspected (Nunney *et al*., 2010). On the contrary, subsp. *multiplex*, which diverged from the former subspecies, between 15,000 and 30,000 years ago, is considered native to the USA, where it is commonly isolated from many native and non-native plant species (Schuenzel *et al*., 2005; Nunney *et al*., 2013). Some correlations were found between the relatedness of groups of *X. fastidiosa* subsp. *multiplex* sequence types (STs) and their hosts (peach, oak or almond) but not with their geographical origins, highlighting the existence some host specialization. Moreover, homologous recombination has shaped the population structure of subsp. *multiplex* found in the USA, rendering possible its adaptation to new hosts. Multiple events of inter-sub-specific homologous recombination of subsp. multiplex with different strains of subsp. *fastidiosa* have created new genotypes able to invade new native plants: mulberry and blueberry. The subsp. *fastidiosa* strains that may have acted as donors have never been found in the USA, but are known to be present in Central America (Nunney *et al*., 2103; Nunney *et al*., 2014a; 2014b). Based on host range and genetic diversity criteria, the rank of subsp. *morus*(a fifth subspecies) has been proposed (Nunney *et al*., 2014b).

High levels of genetic introgression but limited numbers of point mutations in MLST alleles were detected in the bacterial population of subsp. *pauca* infecting *Citrus* spp. and *Coffea* spp. grown in Brazil (Nunney *et al*., 2012). Both hosts, although non-native, have been cultivated for centuries in Brazil, apparently unaffected by *X. fastidiosa*. The first symptoms of citrus variegated chlorosis were observed around 1987, and, few years later coffee leaf scorch disease was also reported. Thus, the indigenous forms of *X. fastidiosa*, although yet to be found, were supposed to be non-pathogenic to these hosts at the time the hosts were introduced to this new environment. It has been speculated that inter-sub-specific recombination, possibly with subsp. *multiplex* infecting some plum trees imported in 1935 from the USA, has recently generated the conditions for the switch from the native host/s, which are also unknown, to citrus and coffee and the consequent genetic bottleneck (Almeida *et al*., 2008; Nunney *et al*., 2012; Coletta-Filho *et al*., 2017). As exemplified by the results of pathogenicity tests, several recombination events must have occurred. Citrus-associated *X. fastidiosa* strains do colonize coffee but are unable to incite disease, while coffee-associated strains fail in infecting citrus (Almeida *et al*., 2008).

Based on this evidence, the entry and adaptation of *X. fastidiosa* into new areas outside North and South America has been long-predicted (Purcell, 1997). In 2013, the presence of *X. fastidiosa* was reported in a large area (approx. 8,000 ha) of the Salento peninsula, in Apulia, Italy (Cariddi *et al*., 2014). The host species first examined were oleander and olive. Soon after, the role of the pathogen in a new form of decline named

"olive quick decline syndrome" (OQDS) was proven (Saponari *et al.*, 2016). To date, the OQDS epiphytotic has spread to the whole province of Lecce and to the provinces of Brindisi and Taranto (April 2017, http://webapps.sit.puglia.it/freewebapps/DatiFasceXF/index.html). MLSA/MLST studies have shown that all isolates from olive and oleander, and from five other plant hosts (*Prunus amygdalus*, *Catharanthus roseus*, *Prunus avium*, *Polygala myrtifolia* and *Westringia fruticosa*) from Apulia share the same sequence type (ST53), which has been reported from Costa Rica, France and the Netherlands in imported coffee plants (Nunney *et al.*, 2014c; Bergsma-Vlami *et al.*, 2017; Denancé *et al.*, 2017). All phylogenetic analyses have unequivocally shown that ST53 is grouped with other STs from the reference strains of *X. fastidiosa* subsp. *pauca* (Cariddi *et al.*, 2014; Nunney *et al.*, 2014c; Bleve *et al.*, 2016; Loconsole *et al.*, 2016). Accordingly, it has been proposed that a genotype of *X. fastidiosa* subsp. *pauca*, referred to as CoDiRO (Complesso del Disseccamento Rapido dell'Olivo), the Italian name first given to the OQDS, has reached the Salento area from Costa Rica with imported plant material, finding favourable conditions for infection and spread (Martelli *et al.*, 2016).

The MLSA/MLST approach is well-suited for global epidemiology and detection of pathogen recombination. Nonetheless, genome sequencing or molecular methods with increased discriminatory power may assist the search for variation among isolates of *X. fastidiosa* recovered within short periods from narrow geographic ranges (Spratt 1999; Almeida *et al.*, 2008; Yuan *et al.*, 2010; Bragard *et al.*, 2016). In the present study, 14 *X. fastidiosa* isolates from the Apulian region were typed using different DNA fingerprinting approaches, and the complete genomes of two isolates, namely Salento-1 (NCPPB 4595; LMG 29352) previously described by Bleve *et al.* (2016), and the recently isolated Salento-2, were sequenced, assembled by the PacBio SMRT and analyzed. Comparative analyses between these isolates and available genomes of *X. fastidiosa* were performed in order to evaluate if and how much genetic variation possibly exists within the population causing the OQDS epidemic in Apulia.

Figure 1. Geographical origin of 14 strains of *Xylella fastidiosa* isolated from different hosts (four *Olea europaea* cultivars and *Nerium oleander*) naturally infected and displaying symptoms of OQDS in the outbreak area of Salento (data reported in Table 1). Site 1, Ugento; site 2, Taviano; site 3, Gallipoli; site 4, Seclì-Galatone; site 5, Galatone. The map is adapted from "http://webapps.sit.puglia.it/freewebapps/DatiFasceXF/index.html" June 2017. The dark blue line designates Provinces of Lecce, Taranto and Brindisi; the red line the containment area, the yellow line, the last 20 km where infected plants have to be removed if present). The light blue line indicates buffer zones (infected plants, if found, and all host plants within a 100 m radius to be removed). Green lines indicate main roads and railways.

Materials and methods

Plant sampling and isolation of *Xylella fastidiosa*

Samples of leaves and twigs of olive (*Olea europaea* L.) or oleander (*Nerium oleander* L.) were collected from plants displaying symptoms of OQDS during spring and autumn of 2015, from five different geographical sites in the OQDS outbreak area of Salento (Figure 1). Symptom severity varied between olive plants, ranging from leaf scorch and desiccation of a few twigs to dieback of entire branches, while oleander plants had leaf scorch and chlorosis. Isolation and presumptive identification of *X. fastidiosa* was accomplished with the procedure of Bleve *et al.* (2016), using the PCR primer set FXYgyr499/RXYgyr907 and conditions described by Rodrigues *et al.* (2003). PCR-positive bacterial colonies were stored at -80°C in sterile PBS buffer (Sigma) containing 50% glycerol (Sigma).

DNA Extraction

Xylella fastidiosa DNA was routinely obtained from 15-20 d-old buffered charcoal yeast extract agar (BCYE: LaBM) cultures using the GenElute Genomic

Minipreps Kit (Sigma) following manufacturer's instructions. Each purified DNA sample suitable for PacBio sequencing (see below) was produced by a treatment with phenol-chloroform-isoamyl alcohol (25:24:1). The resulting supernatant was extracted twice with chloroform-isoamyl alcohol (24:1). The quality, quantity and integrity of DNA were determined through spectrophotometry using a NanoDrop One UV-Vis Spectrophotometer (Thermo Fisher Scientific), fluorospectrometry using the Qubit fluorometer platform (Thermo Fisher Scientific), and agarose gel electrophoresis. High-molecular weight DNA with OD (260/280) > 1.9 and OD 260/230 > 1.9 and yield of at least 10 μg was sent for sequencing to the GATC Biotech genomics sequencing facility at Lake Constance, Germany.

Sanger sequencing characterization of *Xylella fastidiosa* isolates

A fragment of each of the seven MLST loci (*cysG*, *holC*, *leuA*, *gltT*, *nuoL*, *petC* and *malF*) and of the non-MLST *pilU* gene were amplified following the procedure of Yuan *et al.* (2010), using GoTaq DNA Polymerase (Promega) and an annealing temperature of 68°C for the *malF* gene primers.

Since Salento-1/Salento-2 genome comparison obtained with Gepard v.1.40 (Krumsiek *et al.*, 2007) against *X. fastidosa pauca* 9a5c strain revealed the presence of an inverted region of about 1.67 Mbp (see results), two sets of primers P1/P2 and P3/P4 were constructed corresponding to hypothetical inversion points on the genomic DNA sequence of Salento-1/Salento-2. The amplification of a 694 bp fragment was carried out using the primer set P1/P2, and of a 747 bp fragment using the primer set P3/P4 (Supplementary Table S1). Each reaction mixture for amplification contained 1 × GoTaq reaction buffer (Promega), 0.2 mM each dNTP, 0.3 μM each primer, 1.25 U GoTaq DNA Polymerase (Promega), approx. 5 ng of template DNA, and water to a final volume of 25 mL. The conditions used for the amplification were: an initial denaturation at 95°C for 3 min; 35 cycles each at 95°C for 30 s, 65°C for 30 s, and 72°C for 1 min; and a final extension at 72°C for 10 min. Since Sanger sequencing is an appropriate orthogonal technology for determining variant calls of NGS, primer sets PAF/PAR, PBF/PBR, PCF/PCR and PDF/PDR (Supplementary Table S1) were designed and used in the same amplification conditions described above, to verify four of the detected mismatches between the genome sequences of isolates Salento-1 and Salento-2.

All PCR products were visualized after electrophoresis in 1 or 2% agarose gels in 1 × Tris–acetate–EDTA (TAE) buffer and stained with SYBR Safe DNA gel stain (0.1 μL mL^{-1}). They were purified using ExoSAP-IT (USB-Affymetrix), and both strands were sequenced on an ABI prism 3130 Genetic Analyzer system (Applied Biosystems). Sense and antisense electro-pherograms were visualized using CHROMAS LITE 2.01 (Technelysium) and checked for single nucleotide polymorphisms. After removal of primer oligonucleotides, identity searches were performed on the INSDC database (http://www.insdc.org). For MLST data, searches were carried out on the *Xylella fastidiosa* MLST databases website (http://pubmlst.org/xfastidiosa), to determine the allele numbers and the resulting ST.

DNA fingerprinting techniques

For all PCR experiments, DNA was quantified with NanoDrop One (Thermo Fisher Scientific) and GoTaq flexi DNA polymerase (Promega). The PCR reaction products were separated by electrophoresis in 1.5% agarose TAE gels and stained with ethidium bromide. The size of each DNA fragment was estimated by comparison with a 1-Kb plus DNA ladder marker (Life Technologies). For variable number tandem repeat (VNTR) analyses, sets of primers (SSR21, SSR36 and SSR40) for some 7-, 8-, and 9-nucleotide repeats, were used using the amplification conditions indicated by Coletta-Filho *et al.* (2001). For RAPD analyses, primers OPG10, OPG17, OPG19, OPH03, OPH07, OPH12, OPN04, OPQ05 and OPW07 were used, using the amplification conditions indicated by Coletta-Filho *et al.* (2001). The tRNA gene primer T3A (Welsh and McClelland, 1991) was used following amplification conditions indicated by de La Puente-Redondo *et al.* (2000). For repetitive elements identification, rep-PCR was performed using BOX and ERIC primers (Louws *et al.*, 1994) following amplification conditions indicated by de La Puente-Redondo *et al.* (2000).

Genome sequencing, assembly and annotation

Two sequencing libraries were constructed and subjected to the PacBio SMRT sequencing, according

to manufacturer protocols and the quality standard of the GATC Biotech sequencing service. After quality check and read filtering, sequencing of Salento-1 and Salento-2 isolates yielded a total , respectively, of 96,971 and 94,066 reads with aN50 (weighted median of the read length considering the total sequencing effort in base pairs), 19,437 and 18,001 kb, for a total sequencing of 1,270,032,650 bp for Salento-1 and 1,112,749,307 bp for Salento-2, (approx. 500× coverage considering the combined size of genome and plasmid of the type strain *X. fastidiosa* 9a5c). The *de novo* assembly was carried out using the Hierarchical Genome Assembly Process (HGAP), encompassing an Overlap-Layout-Consensus assembly algorithm (specifically suited for PacBio reads and bacterial genomes), and including final consensus polishing with Quiver (Chin *et al.*, 2013). All assembly steps were performed according to tools and protocols available in the SMRT Portal (PacBio). Sequence circularization was carried out with Circlator 1.2.1 (Hunt *et al.*, 2015). Gene prediction and functional annotation were performed using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAP; see the book NBK174280 at the NCBI Bookshelf for details). This pipeline encompasses gene prediction tools such as cmsearch from Infernal (Nawrocki and Eddy, 2013) for rDNA genes, tRNAscan-SE (Lowe and Eddy, 1997) for tRNA genes, GeneMarkS+ (Besemer *et al.*, 2001) for genes and CDS, and BLAST for annotation. The Circos v. 0.69 program (Krzywinski *et al*,. 2009) was used for representation.

Genome-genome distance, and Average Nucleotide Identity (ANI)

Forty genome sequences of *X. fastidiosa* strains were downloaded from the NCBI refseq database as listed in the NCBI assembly database with the query "Xylella fastidiosa". Contig-, draft- and complete-level genomes were considered. For estimating genome-wide distances among the known strains and novel genome sequences of isolates Salento-1 and Salento-2, three widely accepted metrics were used. The Average Nucleotide Identity (ANI) and the correlation indices of the tetranucleotide signatures (TETRA) were calculated with respect to Salento-1 genome sequence using the Jspecies 1.2.1 program (Richter and Rosselló-Móra 2009). In ANI calculations we tested both the BLAST- and the MUMmer-based method, finding no significant discordance. Digital DNA/DNA hybridization (dDDH) values were calculated using the Genome-to-Genome Distance Calculator (GGDC 2.1, Meier-Kolthoff *et al.*, 2013) available at the DMSZ portal (http://ggdc.dsmz.de/distcalc2.php), using BLAST+ to detect pairwise HSPs and retaining dDDH values from formula 2 (the less biased by genome incompleteness, according to GGDC guidelines). This was because most of the available genomes of *X. fastidiosa* strains are resolved at the contig level only. In all comparisons, plasmid sequences, although available for some strains, were not used.

Cluster analyses of strains

For all the genomes included in this study, dDDH calculations (formula 2, see above) were performed, and all vs. all distance matrices were extracted. These matrices were imported into the R statistical package (R Core Team 2015) and subjected to complete hierarchical clustering (hclust, default parameters). The final plots of the clusters was refined and coloured with the R package dendextend (Galili 2015) according to the geographic region (M49 standard; https://unstats.un.org/unsd/methodology/m49/) of the host plant from which the strain was isolated.

Triple-wise comparison of Apulia *Xylella fastidiosa* genome sequences

The genome sequences of *X. fastidiosa* Salento-1, Salento-2 and De Donno (Giampetruzzi *et al.*, 2017) were initially subjected to accurate manual multiple sequence alignment using the sequence editor SeaView (Gouy *et al.*, 2010). Once the genome-wide alignment was completed, an in-house perl script was used to extract relevant positions discriminating the three strains. Briefly, all the positions of the alignment showing nucleotide differences (both indel and substitutions) in at least one of the three genomes were located, collected as differential regions (i.e. single or contiguous positions showing variations) and written as separate records with a 5′ and 3′ expansions of 5 bp. This was to improve description of the genomic context of the variation. Such regions were further annotated by parsing the GenBank Feature Table, in particular by locating their positions in the "gene" feature key (which tracks begin and end positions of each gene) and recovering the relevant annotations from the qualifiers "/locus_tag=" and "/product=" from the "CDS" feature key.

Results

Characterization of isolates by MLST and DNA fingerprinting

The analyses of seven house-keeping gene loci, following the MLST scheme proposed by Yuan and coauthors (2010), have clearly shown that all the isolates in our analyses, including 13 from olive and one from oleander, are indistinguishable from one another. The isolates shared the same sequence type, ST53 (Table 1), with 18 previously described Apulian isolates of *X. fastidiosa* from *Olea europaea*, *Prunus amygdalus*, *Nerium oleander*, *Catharanthus roseus*, *Prunus avium*, *Polygala myrtifolia*, or *Westringia fruticosa* (Giampetruzzi *et al.*, 2015; Bleve *et al.*, 2016; Bragard *et al.*, 2016; Loconsole *et al.*, 2016). Similarly, no differences were found in the sequenced fragment of the *pilU* gene (Accession Number GenBank: KU214457).

To determine if the Apulian *X. fastidiosa* isolates were epidemiologically related, different PCR-based typing methods were used in conjunction, allowing creation and examination of fragments of the pathogen genome. These approaches unambiguously revealed that all the isolates tested possessed the same fingerprint profiles of isolate Salento-1 (Figure 2).

Whole genome sequencing of two novel *Xylella fastidiosa* isolates

To better understand the origins and dynamics of the OQDS, two *X. fastidiosa* isolates were selected and their genomes were completely sequenced, assembled and circularized using the PacBio SMRT technology. Salento-1 (NCPPB 4595; LMG 29352), was isolated in Taviano (Lecce, Italy) from the olive cultivar Ogliarola, as previously described by Bleve *et al.* (2016), and Salento-2, was isolated in Ugento (Lecce, Italy), 10 km away from Taviano from olive cultivar Cellina di Nardò (Table 1). The genomes of Salento-1 and Salento-2 shared GC content of 51.98%, and had almost identical lengths: 2,508,097 bp for Salento-1 and 2,508,296 bp for Salento-2. In each strain a plasmid of 35,270 bp was also found, and these were completely sequenced, assembled and circularized. Two genome and plasmid sequences were deposited on the GeneBank with the following accession numbers: Salento-1

Table 1. Geographic and host origins, and disease symptoms, for *Xylella fastidiosa* isolates used in this study. Isolate MLST-derived sequence type (ST) is included. All sites are localities in Apulia region of Italy.

Isolate	ST	Site	Host plant (cultivar)	Symptoms
CG3 A1	53	Galatone	*Olea europaea* (Ogliarola)	Leaf scorch, dieback of branches
CM4A1	53	Galatone	*Olea europaea* (Cima di Melfi)	Leaf scorch, desiccation of twigs
CM4A2	53	Galatone	*Olea europaea* (Cima di Melfi)	Leaf scorch, desiccation of twigs
Cast1A	53	Gallipoli	*Olea europaea* (Ogliarola)	Leaf scorch, dieback of branches
Cast2AT	53	Gallipoli	*Olea europaea* (Ogliarola)	Leaf scorch, dieback of branches
L2 st A1	53	Seclì-Galatone	*Olea europaea* (Leccino)	Leaf scorch, desiccation of twigs
L2 st A2	53	Seclì-Galatone	*Olea europaea* (Leccino)	Leaf scorch, desiccation of twigs
L6A1	53	Seclì-Galatone	*Olea europaea* (Leccino)	Leaf scorch, desiccation of twigs
L6A2	53	Seclì-Galatone	*Olea europaea* (Leccino)	Leaf scorch, desiccation of twigs
L3 st A2	53	Seclì-Galatone	*Olea europaea* (Leccino)	Leaf scorch, desiccation of twigs
Salento-1[a]	53	Taviano	*Olea europaea* (Ogliarola)	Leaf scorch, dieback of branches
O2A1	53	Taviano	*Nerium oleander*	Leaf scorch and chlorosis
S2A1	53	Ugento	*Olea europaea* (Cellina di Nardò)	Leaf scorch, desiccation of twigs
Salento-2	53	Ugento	*Olea europaea* (Cellina di Nardò)	Leaf scorch, desiccation of twigs

[a] Described in Bleve *et al.*, 2016

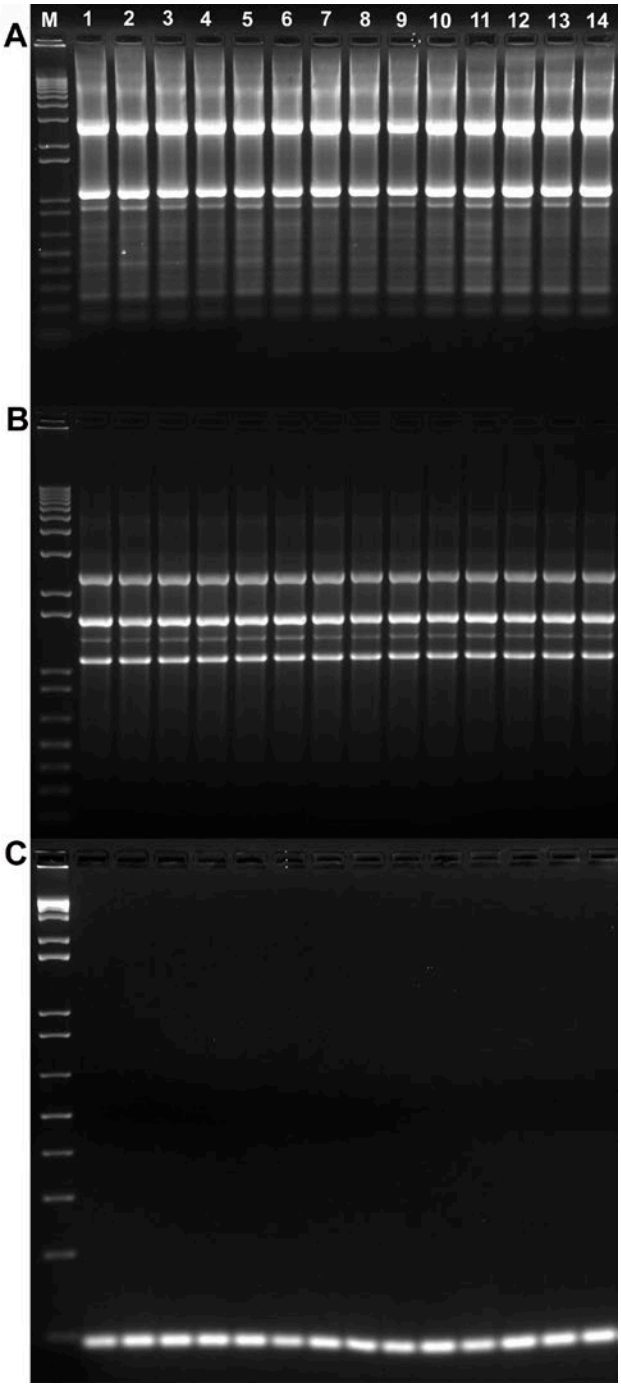

Figure 2. Gel electrophoresis of amplifications obtained by ERIC1R/ERIC2 primers (Panel A) RAPD primer OPH07 (Panel B) and VNTR primer SSR21 (Panel C) for fourteen *Xylella fastidiosa* isolates used in this study. Lane M: 1 Kb-plus ladder marker (Life Technologies); Lanes 1-14: L2stA1, L2stA2, L6A1, L6A2, L3stA2, Salento-1, Salento-2, CG3A1, S2A1, CM4A1, CM4A2, Cast1A, Cast2AT, O2A1.

chromosome CP016608, Salento-1 plasmid CP016609 (pSal1), Salento-2 chromosome CP016610, Salento-2 plasmid CP016611 (pSal2). Annotation results of the Salento-1 and Salento-2 genomes produced by the NCBI system are listed in Table 2. Salento-1 (Figure 3A) and Salento-2 chromosomes were predicted to contain, respectively, totals of 2,276 and 2,288 genes, among which 1,993 and 2,049 were complete protein-coding genes, 224 and 180 were pseudogenes, and six were rRNA genes, 49 were tRNA loci and four were noncoding RNAs. The plasmid annotation identified 35 complete protein-coding genes and two pseudogenes (Figure 3B and Table 2).

Isolates Salento-1 and Salento-2 both showed values of 100 in a genome-wide comparison based on digital DNA/DNA Hybridization (dDDH), indicating that they are the same genotype (Figure 4A).

Whole genome comparisons of *Xylella fastidiosa* isolates Salento-1 and Salento-2 with other strains from Apulia and a reference strain for subsp. *pauca*

Currently, two *X. fastidiosa* genome sequences are available from the Apulian olive isolates of the bacterium, namely CoDiRO (NCBI Accession Number CM003178) and De Donno (CP020870). Although the CoDiRO genome is not fully assembled, the recently deposited De Donno genome is the first complete genome, sequenced by the PacBio and Illumina HiSeq technologies, of *X. fastidiosa* from the Salento area. This was very recently indicated as the reference strain for *X. fastidiosa* causing OQDS in Apulia.

From comparative genome-wide analyses, a dDDH value of 100 has been found between isolates Salento-1 and Salento-2, and of 99.7 with the previously described CoDiRO isolate (Table 3). This difference may be partly because the CoDiRO genome is not assembled, showing a relatively high number of scaffolds. Instead the comparison of Salento-1 and Salento-2 genome sequences with the De Donno genome revealed a dDDH value of 99.9.

The alignment of these three genomes showed the presence of 324 indels and 48 substitutions. Among these, 312 indels corresponded to deletions/insertions of one nucleotide in stretches of poly-A, -T, -G or -C polynucleotides, 40 resembled SNPs in coding and noncoding regions, and 20 were due to presence of regions extended for more than one nucleotide often within repetitive elements (Supplementary File S1). Re-analysis of four identified mismatches between Salento-1 and Salento-2 genome sequences carried out by Sanger se-

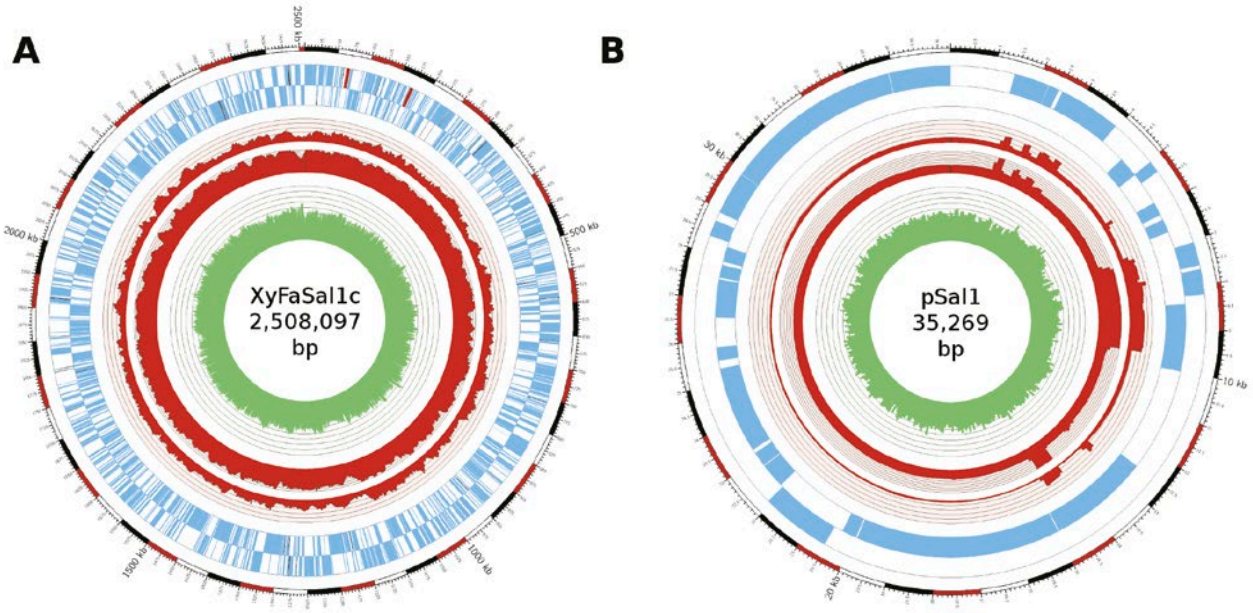

Figure 3. Circular representation of *Xylella fastidiosa* Salento-1 (NCPPB 4595; LMG 29352) complete genome (A) and plasmid (B) obtained with Circos v. 0.69. Circle colours (from outside to inside) describe: 1) location of coding sequences in the two strands (light blue), tRNAs (black) and rRNA (red) genes; 2) coverage in terms of consensus reads (range 0–50) (red); 3) coverage in terms of subreads (range 0–500) (red); and 4) the GC% (range 0–100%) (green).

quencing, indicated that the two genomes are identical in these regions (Supplementary Figure S1).

Salento-1 and Salento-2 genome sequences were further compared by dot matrix analysis with the genome of *X. fastidiosa* 9a5c, the best-known and studied genotype of subsp. *pauca*. This comparison revealed the presence of an inverted region of 1,646,881 bp (Figure 4A).

In order to exclude that this finding could be derived from an assembly error in Salento-1 and Salento-2 genomes, two primer sets were designed using the genome sequences of Salento-1 and Salento-2 over the two points of inversion. Amplification and sequencing of two regions over the two points of inversion confirmed the accurateness of our assembly (Figure 4B). Further, the complete genome of De Donno showed an orientation of the 1.65 Kbp region in accordance with the Salento-1 and Salento-2 genome sequences.

Whole genome comparison can be suitable for taxonomy studies

In the first attempt to classify the two genomes with respect to all available *X. fastidiosa* genomes (in scaffold, contig or complete form), the variation in numbers of complete *rrn* operons was considered across 40 genomes (Table 4). All complete genomes of isolates Salento-1, Salento-2, 9a5c, De Donno, CoDiRO, M12, Temecula1, GB514, M23, Mul0034, Fb7, J1a12, U24d, Pr8x, Hib4 and 3124, contig-level genomes of isolates 11399, CVC0251, CVC0256, OLS0478, Stag's leap, Ann-1, and CFBP8416, and the scaffold genomes of isolates ATCC35879 and DSM10026 were found to contain two copies of a complete *rrn* operon. At this time, the presence of a single copy or the absence of a complete *rrn* operon in the remaining incomplete genomes (contigs or scaffolds) should not be considered a difference of taxonomic relevance, since this may be due to poor quality of genome sequencing, assembly and/or the corresponding annotation (Bhattacharyya *et al.*, 2002). Isolates Salento-1 and Salento-2 were identical in their 16S, 23S and 5S components, and their ribosomal operons were further investigated with respect to 9a5c. The 16S and 5S rRNA sequences were found to be identical between Salento-1, Salento-2 and 9a5c. Five point mutations in 23S rRNA, namely C270T, A289G, G313A, A533G and C872T, were instead found as peculiar of Salento-1

Table 2. Annotation results of the genomes and pSal1 (CP016609) and pSal2 (CP016611) plasmids of *Xylella fastidiosa* isolates Salento-1 (CP016608) and Salento-2 (CP016610), as indicated by the NCBI Prokaryotic Genome Annotation Pipeline v 3.3.

Feature	CP016608	CP016610	CP016609	CP016611
Type	Salento-1	Salento-2	pSal1	pSal2
Genes (total)	2239	2251	37	37
CDS (total)	2180	2192	37	37
Genes (coding)	1993	2049	37	37
CDS (coding)	1993	2049	37	37
Genes (RNA)	59	59	0	0
rRNA operons	2	2	0	0
rRNAs	6	6	0	0
5S	2	2	0	0
16S	2	2	0	0
23S	2	2	0	0
tRNAs	49	49	0	0
ncRNAs	4	4	0	0
Pseudo Genes (total)	221	178	3	2
Pseudo Genes (ambiguous residues)	0	0	0	0
Pseudo Genes (frameshifted)	180	136	1	0
Pseudo Genes (incomplete)	42	44	2	2
Pseudo Genes (internal stop)	28	28	0	0
Pseudo Genes (multiple problems)	28	29	0	0

and Salento-2 with respect to 9a5c (numbering from the 9a5c genome), corresponding to an edit distance of 0.1735%.

The 40 genomes of *X. fastidiosa* strains were also tested to determine their genome-wide distances using ANI, TETRA and dDDH methods (Table 3). Regardless from the geographic region of origin of the isolates, ANIb and ANIm indices where always > 95. In contrast, when the Salento-1 genome sequence of was compared by dDDH analysis to all *X. fastidiosa* genomes considered in this study, within-species variation was clearly detectable (Table 3; Figure 5, panel A). Salento-1 dDDH index ranged from 81.6 to 86.6 in the pairwise comparisons with *X. fastidiosa* subsp. *pauca* genomes from South America (Brazil, Argentina and Ecuador), and was nearly 100% similar with those from Central America (Costa Rica). In contrast, the dDDH index with genomes from Western Europe (subsp. *multiplex*) was ≤ 67.4. Hierarchical cluster analysis of strains based on the distance matrix obtained from the dDDH (formula 2) algorithm (Figure 5, panel B), confirmed the extreme closeness of Apulia isolates with isolates OLS0478, OLS0479 and COF0407 from Central America, and highlighted the existence of two major clades within *X. fastidiosa*. The first clade ranks all the members of *X. fastidiosa* subsp. *pauca* that were included in the analysis regardless of their geographic origins (South America, Central America or Southern Europe). The second clade groups all members of *X. fastidiosa* subspecies *multiplex*, *sandyi*, *morus* and *fastidiosa* from North America, *multiplex* from Western Europe, and isolates CFB8073 and CO33 which were recently isolated from coffee plants imported into Europe from, respectively, Mexico and Costa Rica. Although cluster analyses reveal some segregation according to latitudinal lev-

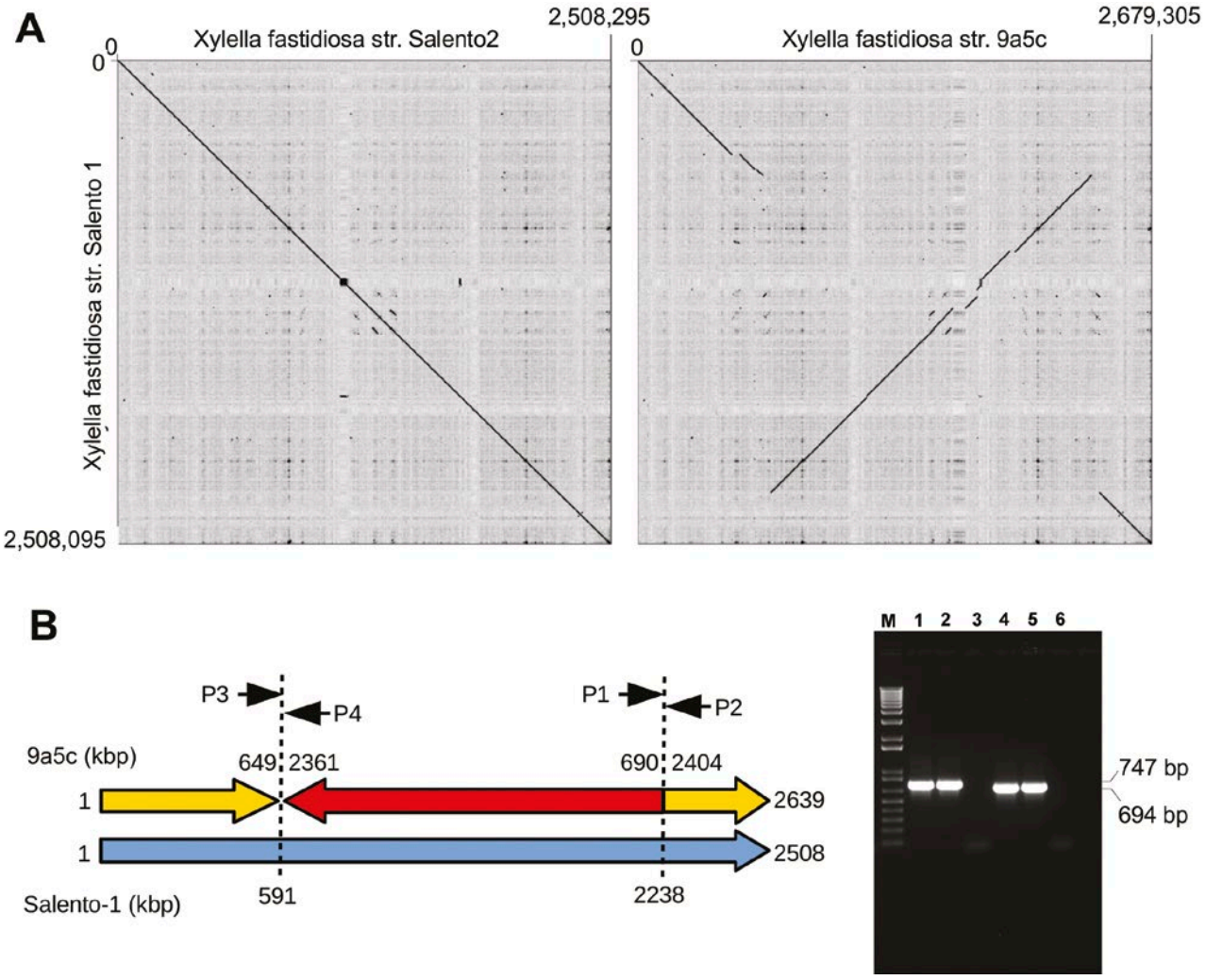

Figure 4. A) Dot matrix representation of the pairwise alignment between complete genomes of Salento-1/Salento-2 and Salento-1/9a5c isolates of *Xylella fastidiosa* created with the suffix-tree method by Gepard v.1.40, with default parameters. B) Schematic linear representation of chromosomal DNA of *X. fastidiosa* 9a5c and of Salento-1 (NCPPB 4595; LMG 29352) reporting points of inverted regions, the localization of primer sets (considering the different starting point number assigned to the bases of the two genomes in GeneBank) and gel electrophoresis of amplification of regions including inversion point performed on *X. fastidiosa* Salento-1 and Salento-2 DNA. 5′ to 3′ direction is determined considering orientation of Salento-1 and Salento-2 chromosome DNA. Lane M, 1 Kb plus ladder marker (Life Technologies); Lane 1, Salento-1 5′ inversion; Lane 2, Salento-2 5′ inversion; Lane 3, no template control; Lane 4, Salento-1 3′ inversion; Lane 5, Salento-2 3′ inversion; Lane 6, no template control.

els and subspecies designation, they also show that strains from Central America split in the two major clusters forming loose associations with other members. Whole genome sequencing of more individual *X. fastidiosa* isolates must be carried out to determine if Central America represents a region of biodiversity for this pathogen.

Discussion

The first aim of this study was to produce multilocus genotype data on 13 *X. fastidiosa* isolates (Figure 1 and Table1) collected from different sites of the OQDS outbreak in Salento, Italy), to acquire more information on *X. fastidiosa* phylogeny and potential pathogen

Table 3. Genome-wide indices for *Xylella fastidiosa* isolates.ANIb = Average Nucleotide Identity obtained using the BLAST aligner. ANIm = Average Nucleotide Identity obtained using the MUM-mer aligner. Tetra = correlation between tretramer frequency. ANIb, ANIm and Tetra were calculated with jspecies. dDDH = Digital DNA/DNA hybridization values obtained with GGDC 2.1 (DDH-2 value only is reported because of its greater accuracy with incomplete genomes). Calculations were carried out with respect to the *X. fastidiosa* Salento-1 isolate. SE = Southern Europe (039), WE = Western Europe (155), CA = Central America (013), SA = South America (005), NA = North America (021).

Strain	Geographic region of the host plant[a]	ANIb	ANIm	Tetra	dDDH	p(dDDH) >= 70%
11399	SA	98.29	97.9	0.99678	82.6	92.40
3124	SA	98.13	97.86	0.99723	81.6	91.73
32	SA	98.88	97.98	0.99773	83.1	92.69
6c	SA	98.15	98.08	0.99742	83.4	92.85
9a5c	SA	98.19	98.04	0.99723	82.8	92.48
Ann-1	NA	96.02	95.82	0.99756	65.1	67.65
ATCC35871	NA	96.54	96.22	0.99794	68.2	74.83
ATCC35879*	NA	NaN	95.71	0.9982	65.3	68.19
BB01	NA	96.27	96.05	0.99845	67.2	72.61
CFBP8072	SA[b]	NaN	98.4	0.99907	86.6	94.5
CFBP8073	CA[b]	96.24	95.82	0.99792	66.2	70.43
CFBP8416	WE	96.04	95.9	0.99835	66.0	69.76
CFBP8417	WE	96.83	96.09	0.99862	67.4	73.16
CFBP8418	WE	96.11	96.07	0.99867	67.4	73.06
CO33	CA[b]	95.61	95.78	0.99811	65.6	68.78
CoDiRO	SE	99.95	99.79	0.99964	99.7	98.22
COF0324	SA	97.8	97.91	0.99693	82.5	92.34
COF0407	CA	99.94	99.75	0.99916	99.6	98.18
CVC0251	SA	98.33	97.92	0.9968	82.9	92.58
CVC0256	SA	98.28	97.94	0.99702	82.8	92.54
De Donno	SE	99.98	99.85	0.9999	99.9	98.25
Dixon	NA	NaN	95.95	0.99855	66.7	71.53
DSM10026*	NA	96.12	95.91	0.99799	66.2	67.88
EB92-1	NA	96.02	95.81	0.9982	65.8	69.43
Fb7	SA	98.17	98.05	0.99691	81.6	91.77
GB514	NA	95.97	95.83	0.99848	65.4	68.30
Griffin-1	NA	96.72	96.13	0.99798	67.5	73.37
Hib4	SA	98.26	98.15	0.99568	83.0	92.61
J1a12	SA	98.24	98.04	0.99581	82.1	92.07
M12	NA	96.14	96.16	0.99872	67.0	72.26

(Continued)

Table 3. (Continued).

Strain	Geographic region of the host plant[a]	ANIb	ANIm	Tetra	dDDH	p(dDDH) >= 70%
M23	NA	95.94	95.82	0.99831	65.1	67.57
Mul-MD	NA	96.39	95.75	0.99822	65.5	68.55
Mul0034	NA	95.98	95.79	0.99806	65.1	67.51
OLS0478	CA	99.96	99.75	0.99939	99.4	98.15
OLS0479	CA	99.88	99.72	0.99926	99.5	98.16
Pr8x	SA	98.18	98.09	0.99736	82.5	92.32
Salento-2	SE	99.99	99.84	1.00000	100	98.28
Stag's leap	NA	95.97	95.73	0.99856	65.4	68.35
sycamore Sy-VA	NA	96.65	96.07	0.99826	67.6	73.43
Temecula1	NA	95.96	95.83	0.99843	65.2	67.91
U24d	SA	98.19	98.03	0.99727	82.7	92.47

* Co-identical strains according to www.dsmz.de.
[a] https://unstats.un.org/unsd/methodology/m49/
[b] Strains CFBP8072, CFBP8073 and CO33 were isolated in Europe from coffee plants that had been imported from Ecuador, Mexico and Costa Rica, respectively (Jacques *et al*., 2016; Loconsole *et al*., 2016).

introduction route(s) (Bragard *et al*., 2016). Although the very first observations of the disease (Elbeaino *et al*., 2014) showed possible existence of variability within the *X. fastidiosa* population in the Salento area of Apulia, the MLST analyses of seven housekeeping gene loci revealed that the 14 Apulian isolates that we have typed in this study belong to the same sequence type, ST53, together with 18 previously characterized isolates from Apulia (Giampetruzzi *et al*., 2015; Bragard *et al*., 2016; Loconsole *et al*., 2016).

To the best of our knowledge, ST53 was first detected in *X. fastidiosa* isolates from oleander and coffee plants in Costa Rica (Nunney *et al*., 2014c). Since three loci (*leuA*, *petC* and *holC*) are identical, and two (*malF* and *gltT*) differ by a single base pair from the corresponding alleles of Brazilian strains of *X. fastidiosa* subsp. *pauca*, Apulian as well as Costa Rican isolates have been tentatively ascribed to this subspecies (Nunney *et al*., 2014c; Martelli *et al*., 2016). However, as pointed out previously (Nunney *et al*., 2014c; Coletta-Filho *et al*., 2017), the number of nucleotide polymorphisms found in *cysG* and *nuoL* loci are consistent with a different phylogeny of ST53 from the *X. fastidiosa* subsp. *pauca* populations that are currently known to exist in Brazil and Argentina, infecting citrus, coffee, olive and hibiscus.

The MLST approach is very suitable for grouping bacteria that share recent common ancestors into distinct clades or clonal complexes (Maiden *et al*., 2013). Nevertheless, being based on analyses of selected housekeeping gene fragments, the method is intrinsically inappropriate for resolving differences among isolates that belong to one lineage. Thus, to understand if the OQDS epiphytotic was or was not caused by an outbreak strain (*sensu* Maiden *et al*., 2013) of *X. fastidiosa*, typing approaches that index more variable loci need to be used, as suggested by Yuan *et al*. (2010) and Almeida *et al*. (2008). Apart from MLST data, other lines of evidence indicate that the current Apulian disease outbreak is the result of a founder event by one lineage of *X. fastidiosa*. Prior to 2013 there was no formal evidence of *X. fastidiosa* diseases in the countries of the Mediterranean basin (https://gd.eppo.int/taxon/XYLEFA/distribution), Furthermore, when OQDS was initially detected in 2013 the disease was restricted to an area of about 8,000 ha of the Lecce province (Cariddi *et al*., 2014). In the four following years, the disease has spread unhalted beyond the Province of Lecce, moving North through the Brindisi province and, more recently, the Province of Taranto (Bollettino Ufficiale della Regione Puglia - n. 64 del 01-6-2016). In this scenario, finding that none of the conventional

Table 4. *Xylella fastidiosa* genomes used in this study.

Strain name	Biosample (NCBI)	Assembly status	Genome scaffolds	No. of complete rrn operons	Sequence Type
11399	SAMN02786837	Contig	36	2	11
3124	SAMN03166214	Complete	1	2	16 [a]
32	SAMN02471372	Contig	56	1	16
6c	SAMN02471371	Contig	19	1	14
9a5c	SAMN02603773	Complete	1	2	13
Ann-1	SAMN03081476	Contig	1	2	5
ATCC35871	SAMN02441559	Contig	62	1	41
ATCC35879*	SAMN02997312	Scaffold	16	2	2
BB01	SAMN05982167	Scaffold	83	1	42[a]
CFBP8072	SAMN04075656	Scaffold	278	0	74
CFBP8073	SAMN04075659	Scaffold	328	0	75
CFBP8416	SAMN04546448	Contig	128	2	7
CFBP8417	SAMN04546482	Contig	256	1	6
CFBP8418	SAMN04546487	Contig	271	1	6
CO33	SAMN04100274	Contig	96	0	72
CoDiRO	SAMN03247589	Contig	12	2	53
COF0324	SAMN03862122	Contig	143	1	14
COF0407	SAMN03862125	Contig	172	1	53
CVC0251	SAMN03862120	Contig	130	2	11
CVC0256	SAMN03862121	Contig	128	2	11
De Donno	SAMN06765826	Complete	1	2	53
Dixon	SAMN02441075	Scaffold	32	1	6
DSM10026*	SAMN05660380	Scaffold	63	2	2[a]
EB92-1	SAMN02471770	Contig	168	0	1
Fb7	SAMN03154693	Complete	1	2	69[a]
GB514	SAMN02603754	Complete	1	2	1
Griffin-1	SAMN02472062	Contig	84	1	7
Hib4	SAMN03173341	Complete	1	2	70[a]
J1a12	SAMN03166204	Complete	1	2	11[a]
M12	SAMN02598402	Complete	1	2	7
M23	SAMN02598408	Complete	1	2	1
Mul-MD	SAMN02630185	Contig	101	1	29
Mul0034	SAMN03081485	Complete	1	2	30
OLS0478	SAMN03862124	Contig	48	2	53

(Continued)

Table 4. (Continued).

Strain name	Biosample (NCBI)	Assembly status	Genome scaffolds	No. of complete rrn operons	Sequence Type
OLS0479	SAMN03862123	Contig	183	1	53
Pr8x	SAMN03166213	Complete	1	2	14[a]
Salento-1	SAMN05429913	Complete	1	2	53
Salento-2	SAMN05429914	Complete	1	2	53
Stag's leap	SAMN04485942	Contig	15	2	1
sycamore Sy-VA	SAMN02709772	Contig	128	1	8
Temecula1	SAMN02603844	Complete	1	2	1
U24d	SAMN03154423	Complete	1	2	13[a]

* Co-identical strains according to www.dsmz.de.
[a] According to https://pubmlst.org/xfastidiosa/

whole genome fingerprinting techniques that we used reveals the existence of genetic variation suggests that the Apulian population was founded by a single genotype or very similar genotypes, for example by different members of the clonal complex that was described recently (Marcelletti and Scortichini 2016). The time scale of the Apulian outbreak is extremely short compared with the estimated *X. fastidiosa* evolutionary time scale (Schuenzel *et al.*, 2005). Therefore, it could be difficult to track microevolution events (Feil, 2004; Tang *et al.*, 2017), particularly for a bacterium such as *X. fastidiosa* that is prone to recombination (Scally *et al.*, 2005), by using molecular methods such as RAPDs or rep-PCR. This is because these methods gain their signals from small fractions of the genome.

As demonstrated by recent studies (Rhoads and Au, 2015), the PacBio sequencing, adopted for sequencing of *X. fastidiosa* Salento-1 and Salento-2 strains, provides long reads (> 10 kbp) and high coverage for small genomes, overcoming many of the obstacles faced by short-read producing NGS instruments (e.g. Illumina). Long overlapping reads are particularly suited for sequencing regions with low complexity or amplification (e.g. duplication, such as ribosomal genes), that frequently impair correct genome assembly. This allowed bacterial chromosomes to be fully resolved and circularized starting from reads obtained by a single PacBio run. At the present time, the three Apulian *X. fastidiosa* isolates sequenced with PacBio technology (De Donno, Salento-1 and Salento-2) should be considered "genomically indistinguishable" at the whole genome level, since the observed differences can result from technical artifacts associated to sequencing technology (Miyamoto *et al.*, 2014). When we used Sanger sequencing, none of the mismatches between the Salento-1 and Salento-2 genome sequences that we arbitrarily chose, was endorsed. Refinements by Sanger sequencing may be necessary to also verify if the sequences of genes annotated as pseudo-genes are due to miss start or stop codons, internal frameshifts or other coding abnormalities. Several studies have reported that the accumulation of single nucleotide polymorphism can indicatively range between two and ten per year in various bacteria (Salipante *et al.*, 2015). Re-sequencing of all identified SNPs and regions extended for more than one single nucleotide, is therefore necessary to acquire valuable information about the epidemiology of *X. fastidiosa* in Apulia, and new insights on possible effects of micro-evolutionary forces acting on this bacterium in Salento. Nevertheless, whole genome comparisons, confirming the evidence obtained by the MLST analyses, demonstrated that Apulia strains are very closely related to strains of *X. fastidiosa* from Central America. Taken together, DDH and ANI calculations confirmed that previously known *X. fastidiosa* subsp. *pauca* strains from Central America and Apulia (Southern Europe) are very closely related to each other, suggesting a common geographical origin. They also show a loose relationship with *X. fastidiosa* subsp. *pauca* strains from South America. These differences are more accentuated when North America and Western Europe and counterparts are considered, raising

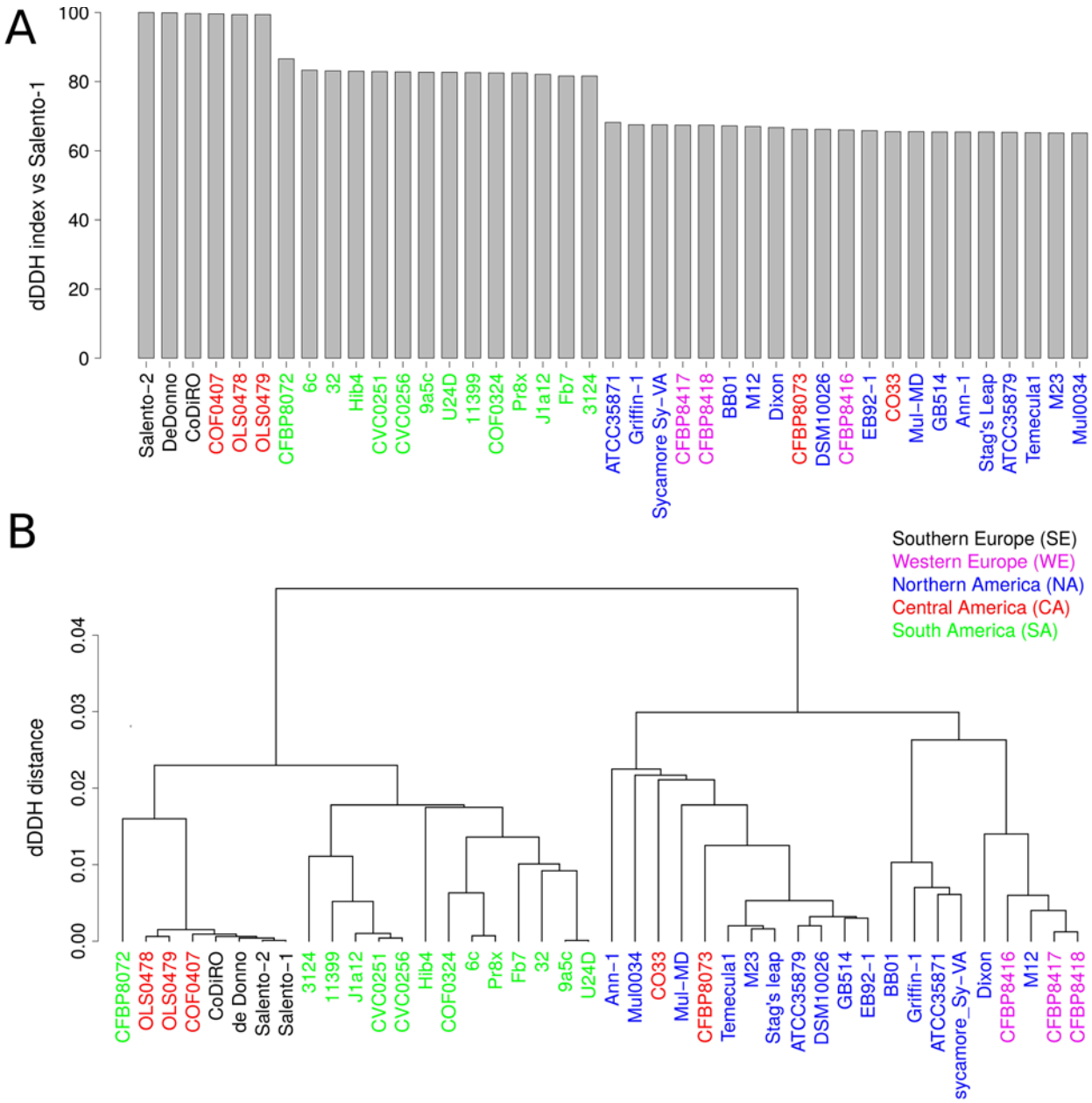

Figure 5. A, barplot showing, in decreasing order, the genomic (at all assembly status) dDDH indices of *Xylella fastidiosa* Salento-1 (NCPPB 4595; LMG 29352) versus all other strains included in this study. The Salento-1 genome sequence was used as the reference genome. B, complete hierarchical clustering based on dDDH-derived distance among strains. Isolates names are coloured according to the geographic region of origin of host plants from which they were isolated (see legend in the middle of the figure).

questions regarding the positioning of all *X. fastidiosa* strains as members of the same species, as already observed by Barbosa *et al.* (2015). Indeed, the dDDH values were below the 70% threshold useful for bacterial species boundaries (Goris *et al.*, 2007; Richter and Rosselló-Móra, 2009) for all remaining genomes of *X. fastidios*a strains belonging to subspecies different from the subsp. *pauca*. Genome sequencing of isolates from

the disease progress areas will be very important for monitoring the possible development of new strains deriving from recombination events, with new genotypes that can possibly enter Apulia. This approach will also provide understanding of genome dynamics and evolution of the bacterial pathogen in a new environment. In this regard, it was recently announced that the genomes of 40 *X. fastidiosa* isolates from Apulian olives have been sequenced (Sicard *et al.*, 2017). Such new genome sequencing efforts will take advantage of complete, fully resolved and circularized genomes, as those characterized in the present study.

Conclusions

Data presented here highlight the existence of a widely spread genotype of *X. fastidiosa* in the OQDS affected orchards of the Salento area of Italy. However, we acquired only partial evidence that the detected mismatches between three complete genomes (of isolates Salento-1, Salento-2 and De Donno) are due only to technical artifacts. Therefore we cannot exclude that micro-evolutionary forces have already started to produce some variation in the pathogen population in this area. Since 2013, infected plant material has been intercepted in different European countries (the Czech Republic, France, Germany, Italy, Spain, Switzerland, the Netherlands), and several disease foci caused by *X. fastidiosa* subsp. *multiplex* and subsp. *fastidiosa* affecting various ornamental plants have been confirmed in France, or fruit crops in Spain (https://gd.eppo.int/taxon/XYLEFA/distribution; Bergsma-Vlami *et al.*, 2017). These findings indicate that a long time after the first spread of *X. fastidiosa* in the American continents, repeated introduction events of different STs from different subspecies have occurred in several European countries (Bergsma-Vlami *et al.*, 2017; Denancé *et al.*, 2017; Olmo *et al.*, 2017) and Iran (Amanifar *et al.*, 2014). Furthermore, different genotypes of this highly adaptable bacterial species have suddenly found the right conditions to survive in new environments, as is the case of Spain (Balearic Islands) and France (Corse and mainland) and, eventually, to become established as is the case of ST53 in Apulia on olive plants (https://gd.eppo.int/taxon/XYLEFA/distribution). In such a scenario, the fortuitous encounter of diverse genotypes leading to events of intra- and inter-sub-specific homologous recombination, which possibly generate more virulent forms or new pathotypes, is a pending risk (Feil and Spratt 2001; Nunney *et al.*, 2014a; 2014c). In the near future massive genome sequencing of isolates from the Apulian region will make rapid monitoring possible of how the evolutionary forces will or did concur, to shape the population structures of this key plant pathogen after its arrival in a new environment.

Acknowledgements

Dr Fabio Cimaglia was supported by a grant (Pratica N. 2014.0727) from Fondazione Ente Cassa di Risparmio di Firenze (Italy). This study was supported in part by the "TAPASS" project [Regione Puglia FSC 2007-2013 "Cluster Tecnologici Regionali 2014" (BURP n. 138/2014)] and by the Regione Toscana. The authors wish to thank Dr. Vittorio Falco for his high specific technical assistance for the DNA sequencing.

Literature cited

Almeida R.P.P., F.E. Nascimento, J. Chau, S.S. Prado, C.W. Tsai, … and J.R. Lopes, 2008.. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Applied and Environmental Microbiology* 74, 3690–3701.

Amanifar N., M. Taghavi, K. Izadpanah and G. Babaei, 2014. Isolation and pathogenicity of *Xylella fastidiosa* from grapevine and almond in Iran. *Phytopathologia Mediterranea* 53, 318–327

Baker R., C. Bragard, D. Caffier, T. Candresse, G. Gilioli, ... and D. Makowski, 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 Journal* 13, 3989.

Barbosa D., V.C. Alencar, D.S. Santos, A.C. de Freitas Oliveira, A.A. de Souza, ... and L.R. Nunes, 2015. Comparative genomic analysis of coffee-infecting *Xylella fastidiosa* strains isolated from Brazil. *Microbiology* 161, 1018–1033.

Bergsma-Vlami M., J.L.J. van de Bilt, N.N.A. Tjou-Tam-Sin, C.M. Helderman, P.P.M.A. Gorkink-Smits, … and M. Westenberg, 2017. Assessment of the genetic diversity of *Xylella fastidiosa* in imported ornamental *Coffea arabica* plants. *Plant Pathology* 66, 1065–1074.

Besemer J., A. Lomsadze and M. Borodovsky, 2001. GeneMark S: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* 29, 2607–2618.

Bhattacharyya A., S. Stilwagen, G. Reznik, H. Feil, W. S. Feil, ... and N. Larsen, 2002. Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. *Genome Research* 12, 1556–1563.

Bleve G., G. Marchi, F. Ranaldi, A. Gallo, F. Cimaglia, ... and G. Surico, 2016. Molecular characteristics of a strain Salento-1) of *Xylella fastidiosa* isolated in Apulia (Italy) from an olive plant with the quick decline syndrome. *Phytopathologia Mediterranea* 55,139–146.

Bollettino Ufficiale della Regione Puglia, 2016. – No. 64 del 01-6-2016.

Bragard C., D. Caffier, T. Candresse, E. Chatzivassiliou, K. Dehnen-Schmutz, … and S. Winter, 2016. Statement on diversity of *Xylella fastidiosa* subsp. *pauca* in Apulia. *EFSA Journal* 14, 4542.

Bull C.T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, ... and Y. Takikawa, 2012. List of new names of plant pathogenic bacteria (2008–2010). *Journal of Plant Pathology* 94, 21–27.

Cariddi C., M. Saponari, D. Boscia, A. De Stradis, G. Loconsole, ... and G. P. Martelli, 2014. Isolation of a *Xylella fastidiosa* strain infecting olive and oleander in Apulia, Italy. *Journal of Plant Pathology* 96, 425–429.

Chatterjee S., R.P.P. Almeida and S. Lindow, 2008. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annual Review of Phytopathology* 46, 243–271.

Chin C.S., D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, ... and S.W. Turner, 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* 10, 563–569.

Coletta-Filho H.D., M.A. Takita, A.A. de Souza, C.I. Aguilar-Vildoso and M.A. Machado, 2001. Differentiation of strains of *Xylella fastidiosa* by a variable number of tandem repeat analysis. *Applied and Environmental Microbi*ology 67, 4091–4095.

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

Denancé N., B. Legendre, M. Briand, V. Olivier, C. Boisseson, … and M.A. Jacques, 2017. Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France. *Plant Pathology* 7, 1054–1064.

De La Puente-Redondo V.A., N.G. del Blanco, C.B. Gutierrez-Martin, F.J. Garcìa-Peña and E.R. Ferri, 2000. Comparison of different PCR approaches for typing of *Francisella tularensis* strains. *Journal of Clinical Microbiology* 38, 1016–1022.

Elbeaino T., F. Valentini, R. Abou Kubaa, P. Moubarak, T. Yaseen and M. Digiaro, 2014. Multilocus sequence typing of *Xylella fastidiosa* isolated from olive affected by "olive quick decline syndrome" in Italy. *Phytopathologia Mediterranea* 53, 533–542.

Euzéby J., 2009. In: List of New Names and New Combinations Previously Effectively, but not Validly, Published, List 127. *International Journal of Systematic and Evolutionary Microbiology* 59, 923–925.

Feil E.J., 2004. Small change: keeping pace with microevolution. *Nature Review of Microbiology* 2, 483–495.

Feil E.J. and B.G. Spratt, 2001. Recombination and the population structures of bacterial pathogens. *Annual Review of Microbiology* 55, 561–590.

Galili T., 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720.

Gevers D., F.M. Cohan, J.G. Lawrence, B.G. Spratt, T. Coenye, ... and J. Swings, 2005. Opinion: Re-evaluating prokaryotic species. *Nature Reviews Microbiology* 3, 733–739.

Giampetruzzi A., M. Chiumenti, M. Saponari, G. Donvito, A. Italiano, … and P. Saldarelli, 2015. Draft genome sequence of the *Xylella fastidiosa* CoDiRO strain. *Genome Announcements* 3, e01538-14.

Giampetruzzi A., M. Saponari, R.P.P. Almeida, S. Essakhi, D. Boscia, … and P. Saldarelli, 2017. Complete Genome Sequence of the Olive-Infecting Strain *Xylella fastidiosa* subsp. *pauca* De Donno. *Genome announcements* 5, e00569-17.

Goris J., K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme and J.M. Tiedje, 2007. DNA-DNA hybridization values and their relation to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 57, 81–91.

Gouy M., S. Guindon and O. Gascuel, 2010. SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27, 221–224.

Hunt M., N. De Silva, T.D. Otto, J. Parkhill, J.A. Keane and S.R. Harris, 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology* 16, 294.

Jacques M.A., N. Denancé, B. Legendre, E. Morel, M. Briand, ... and D. Crouzillat, 2016. New coffee plant-infecting *Xylella fastidiosa* variants derived via homologous recombination. *Applied and Environmental Microbiology* 82, 1556–1568.

Jolley K.A. and M.C. Maiden, 2014. Using MLST to study bacterial variation: prospects in the genomic era. *Future Microbiology* 9, 623–630.

Krumsiek J., R. Arnold and T. Rattei, 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 8, 1026–1028.

Krzywinski M., J. Schein, I. Birol, J. Connors, R. Gascoyne, ... and M.A. Marra, 2009. Circos: an information aesthetic for comparative genomics. *Genome research* 19: 1639–1645.

Kung S.H. and R.P.P. Almeida, 2011. Natural competence and recombination in the plant pathogen *Xylella fastidiosa*. *Applied and Environmental Microbiology* 77, 5278–5284.

Loconsole G., M. Saponari, D. Boscia, G. D'Attoma, M. Morelli, … and R.P.P. Almeida, 2016. Intercepted isolates of *Xylella fastidiosa* in Europe reveal novel genetic diversity. *European Journal of Plant Pathology* 146, 85–94.

Louws F.J., D.W. Fulbright, C.T. Stephens and F.J. de Bruijn, 1994. Specific genomic fingerprints of phytopathogenic *Xanthomonas* and *Pseudomonas* pathovars and strains generated with repetitive sequences and PCR. *Applied and Environmental Microbiology* 60, 2286–2295.

Lowe T.M. and S.R. Eddy, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25, 955–964.

Maiden M.C., M.J.J. Van Rensburg, J.E. Bray, S.G. Earle, S.A. Ford, … and N.D. McCarthy, 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* 11, 728–736.

Marcelletti S. and M. Scortichini, 2016. *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.

Martelli G.P., D. Boscia, F. Porcelli and M. Saponari, 2016. The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. *European Journal of Plant*

Pathology 144, 235–243.
Meier-Kolthoff J.P., A.F. Auch, H.P. Klenk and M. Göker, 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14, 60.
Miyamoto M., D. Motooka, K. Gotoh, T. Imai, K. Yoshitake, ... and M. Kasahara, 2014. Performance comparison of second and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics* 15, 699.
Nawrocki E.P. and S.R. Eddy, 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935.
Nunney L., X. Yuan, R. Bromley, J. Hartung, M. Montero-Astúa, ... and R. Stouthamer, 2010. Population genomic analysis of a bacterial plant pathogen: novel insight into the origin of Pierce's disease of grapevine in the US. *PLoS One* 5: e15488.
Nunney L., X. Yuan, R.E. Bromley and R. Stouthamer, 2012. Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. *Applied and Environmental Microbiology* 78, 4702–4714.
Nunney L., D.B. Vickerman, R.E. Bromley, S.A. Russell, J.R. Hartman, … and R. Stouthamer, 2013. Recent evolutionary radiation and host plant specialization in the *Xylella fastidiosa* subspecies native to the United States. *Applied and Environmental Microbiology* 79, 2189–2200.
Nunney L., D.L. Hopkins, L.D. Morano, S.E. Russell and R. Stouthamer, 2014a. Intersubspecific recombination in *Xylella fastidiosa* strains native to the United States: infection of novel hosts associated with an unsuccessful invasion. *Applied and Environmental Microbiology* 80, 1159–1169.
Nunney L., E.L. Schuenzel, M. Scally, R.E. Bromley and R. Stouthamer, 2014b. Large-scale intersubspecific recombination in the plant-pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. *Applied and Environmental Microbiology* 80, 3025–3033.
Nunney L., B. Ortiz, S.A. Russell, R.R. Sánchez and R. Stouthamer, 2014c. The complex biogeography of the plant pathogen *Xylella fastidiosa*: genetic evidence of introductions and subspecific introgression in Central America. *PloS one* 9, e112463.
Olmo D., A. Nieto, F. Adrover, A. Urbano, O. Beidas, ... and M. Montes-Borrego, 2017. First detection of *Xylella fastidiosa* on cherry (*Prunus avium*) and *Polygala myrtifolia* plants, in Mallorca Island, Spain. *Plant Disease*, 101, 1820.
Purcell A.H., 1997. *Xylella fastidiosa*, a regional problem or global threat? *Journal of Plant Pathology* 79, 99–105.
Purcell A., 2013. Paradigms: examples from the bacterium *Xylella fastidiosa*. *Annual Review of Phytopathology* 51, 339–356.
R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Rhoads A. and K.F. Au, 2015. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 13, 278–289.
Richter M. and R. Rosselló-Móra, 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences* 106, 19126–19131.
Rodrigues J.L., M.E. Silva-Stenico, J.E. Gomes, J.R. S. Lopes and S.M. Tsai, 2003. Detection and diversity assessment of *Xylella fastidiosa* in eld-collected plant and insect samples by using 16S rRNA and gyrB sequences. *Applied and Environmental Microbiology* 69, 4249–4255.
Salipante S.J., D.J. SenGupta, L.A. Cummings, T.A. Land, D.R. Hoogestraat and B.T. Cookson, 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *Journal of Clinical Microbiology* 53, 1072–1079.
Saponari M., D. Boscia, G. Altamura, G.D'Attoma, V. Cavalieri, ... and O. Potere, 2016. Pilot project on *Xylella fastidiosa* to reduce risk assessment uncertainties. *EFSA Supporting Publications*, DOI: 10.2903/sp.efsa.2016.EN-1013
Scally M., E.L. Schuenzel, R. Stouthamer and L. Nunney, 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Applied and Environmental Microbiology* 71, 8491–8499.
Schaad N.W., E. Postnikova, G. Lacy, M.B. Fatmi and C. Chung-Jan, 2004. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. *piercei*, subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov.. *Systematic and Appl ied Microbiology* 27, 290–300.
Schuenzel E.L., M. Scally, R. Stouthamer, and L. Nunney, 2005. A multigene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. *Applied and Environmental Microbiology* 71, 3832–3839.
Sicard A., M. Saponari, M. Vanhove, S. Essakhi, A. Giampetruzzi, … and R.P.P. Almeida, 2017. Identifying *Xylella fastidiosa* host adaptation candidate genes: the case of *X. fastidiosa* subsp. *pauca* isolates and olive trees in Italy. In: *Abstracts, European conference on Xylella 2017*: *finding answers to a global problem*, November 13–15, 2017, Palma de Mallorca, Spain, 43–44.
Spratt B.G., 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Current Opinion in Microbiology* 2, 312–316.
Tang P., M.A. Croxen, M.R. Hasan, W.W. Hsiao and L.M. Hoang, 2017. Infection control in the new age of genomic epidemiology. *American Journal of Infection Control* 45, 170–179.
Wells J.M., B.C. Raju, H.Y. Hung, W.G. Weisburg, L. Mandelco-Paul, and D.J. Brenner, 1987. *Xylella fastidiosa* gen. nov., sp. nov: gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. *International Journal of Systematic and Evolutionary Microbiology* 37, 136–143.
Welsh J. and M. McClelland, 1991. Genomic fingerprints produced by PCR with consensus tRNA gene primers. *Nucleic Acids Research* 19, 861–866.
Yuan X., L. Morano, R. Bromley, S. Spring-Pearson, R. Stouthamer and L. Nunney, 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander Leaf Scorch in the United States. *Phytopathology* 100, 601–611.

Accepted for publication: December 22, 2017
Published online: February 14, 2018