AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

EVOLUTIONARY AND GENOMIC MICROBIOLOGY

Genomic Diversity and Recombination among *Xylella fastidiosa* Subspecies

Mathieu Vanhove,[a] Adam C. Retchless,[a]* Anne Sicard,[a] Adrien Rieux,[b] Helvecio D. Coletta-Filho,[c] Leonardo De La Fuente,[d] Drake C. Stenger,[e] Rodrigo P. P. Almeida[a]

[a]Department of Environmental Science, Policy and Management, University of California, Berkeley, California, USA
[b]CIRAD, UMR PVBMT, St. Pierre, Réunion, France
[c]Centro de Citricultura Sylvio Moreira, IAC, Sao Paulo, Brazil
[d]Department of Entomology and Plant Pathology, Auburn University, Auburn, Alabama, USA
[e]San Joaquin Valley Agricultural Sciences Center, Agricultural Research Service, U.S. Department of Agriculture, Parlier, California, USA

ABSTRACT *Xylella fastidiosa* is an economically important bacterial plant pathogen. With insights gained from 72 genomes, this study investigated differences among the three main subspecies, which have allopatric origins: *X. fastidiosa* subsp. *fastidiosa*, *multiplex,* and *pauca*. The origin of recombinogenic *X. fastidiosa* subsp. *morus* and *sandyi* was also assessed. The evolutionary rate of the 622 genes of the species core genome was estimated at the scale of an *X. fastidiosa* subsp. *pauca* subclade (7.62×10^{-7} substitutions per site per year), which was subsequently used to estimate divergence time for the subspecies and introduction events. The study characterized genes present in the accessory genome of each of the three subspecies and investigated the core genome to detect genes potentially under positive selection. Recombination is recognized to be the major driver of diversity in *X. fastidiosa*, potentially facilitating shifts to novel plant hosts. The relative effect of recombination in comparison to point mutation was calculated ($r/m = 2.259$). Evidence of recombination was uncovered in the core genome alignment; *X. fastidiosa* subsp. *fastidiosa* in the United States was less prone to recombination, with an average of 3.22 of the 622 core genes identified as recombining regions, whereas a specific clade of *X. fastidiosa* subsp. *multiplex* was found to have on average 9.60 recombining genes, 93.2% of which originated from *X. fastidiosa* subsp. *fastidiosa*. Interestingly, for *X. fastidiosa* subsp. *morus*, which was initially thought to be the outcome of genome-wide recombination between *X. fastidiosa* subsp. *fastidiosa* and *X. fastidiosa* subsp. *multiplex*, intersubspecies homologous recombination levels reached 15.30% in the core genome. Finally, there is evidence of *X. fastidiosa* subsp. *pauca* strains from citrus containing genetic elements acquired from strains infecting coffee plants as well as genetic elements from both *X. fastidiosa* subsp. *fastidiosa* and *X. fastidiosa* subsp. *multiplex*. In summary, our data provide new insights into the evolution and epidemiology of this plant pathogen.

IMPORTANCE *Xylella fastidiosa* is an important vector-borne plant pathogen. We used a set of 72 genomes that constitutes the largest assembled data set for this bacterial species so far to investigate genetic relationships and the impact of recombination on phylogenetic clades and to compare genome content at the subspecies level, and we used a molecular dating approach to infer the evolutionary rate of *X. fastidiosa*. The results demonstrate that recombination is important in shaping the genomes of *X. fastidiosa* and that each of the main subspecies is under different selective pressures. We hope insights from this study will improve our understanding of *X. fastidiosa* evolution and biology.

KEYWORDS *Xylella fastidiosa*, emerging disease, genomic diversity, recombination

Citation Vanhove M, Retchless AC, Sicard A, Rieux A, Coletta-Filho HD, De La Fuente L, Stenger DC, Almeida RPP. 2019. Genomic diversity and recombination among *Xylella fastidiosa* subspecies. Appl Environ Microbiol 85:e02972-18. https://doi.org/10.1128/AEM.02972-18.

Editor Emma R. Master, University of Toronto

Address correspondence to Rodrigo P. P. Almeida, rodrigoalmeida@berkeley.edu.

* Present address: Adam C. Retchless, Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

Received 11 December 2018
Accepted 19 April 2019
Accepted manuscript posted online 26 April 2019
Published 17 June 2019

Xylella fastidiosa is a plant pathogen responsible for damaging plant diseases that has emerged in the past few decades. Emerging diseases caused by the bacterium have been associated with long-distance introductions and geographical expansion of an insect vector, as well as host adaptation (1). The pathogen has a significant economic impact in multiple agroecosystems, causing a wide range of diseases worldwide, including Pierce's disease of grapevines (PD), citrus variegated chlorosis (CVC), leaf scorch diseases in almond (ALS), coffee (CLS), and oleander (OLS), and, more recently, olive quick decline syndrome (OQDS) (2).

To understand these epidemics, significant research advances regarding evolutionary origin and ecology of the pathogen have been obtained using a multilocus sequence typing (MLST) approach based on seven housekeeping genes (3). The repository of sequence data led to a consensus regarding the evolutionary history of this bacterium, and *X. fastidiosa* was divided into four subspecies, largely separated due to geographical constraints (4). While *X. fastidiosa* subsp. *sandyi* infects oleander in the southern regions of the United States, *X. fastidiosa* subsp. *multiplex* causes a wide range of diseases in commercial plants, including peach, blueberry, plum, and almond, as well as in trees, such as maple, elm, and sycamore, throughout North America (5, 6), and plum leaf scorch disease in South America (7). *X. fastidiosa* subsp. *fastidiosa* causes disease in grapevine, among other plants, and is thought to have originated in Central America (8). In South America, *X. fastidiosa* subsp. *pauca* causes disease in citrus and coffee; this subspecies was recently found to cause OQDS in southern Italy and in Southeast Brazil. Genotyping led to the characterization of a monophyletic *X. fastidiosa* subsp. *pauca* group in Italy known as sequence type 53 (ST53) (9, 10); in Brazil, the lineage associated with olive trees was ST16, first reported in coffee plants (7). These examples reflect the importance of human-mediated invasion. Other reports of introductions have been described, such as the identification of *X. fastidiosa* subsp. *fastidiosa* in Taiwan (11) and the presence of *X. fastidiosa* subsp. *pauca* in Central America (4). A fifth subspecies infecting mulberry has been reported (*X. fastidiosa* subsp. *morus*) and is thought to be the result of intersubspecific homologous recombination (IHR) between *X. fastidiosa* subsp. *fastidiosa* and *multiplex* (12). The current taxonomy of *X. fastidiosa* is based on a limited and biased data set, meaning that the small number of strains genotyped were collected mostly from important crop plants, and genotyping was done primarily using MLST (9, 13). This is likely to change with the advent of whole-genome sequencing data, so that phylogenic relationships can be investigated at a finer scale and genomic details such as recombination events can be unveiled.

X. fastidiosa entered the genomic era when the first genome of a plant-associated bacterium was sequenced in the year 2000, citrus variegated chlorosis (CVC) strain 9a5c (14). Two years later, the draft sequences of Dixon (*X. fastidiosa* subsp. *multiplex*) and Ann-1 (*X. fastidiosa* subsp. *sandyi*) (15) were released, soon followed by the etiological agent of PD in California (Temecula1, *X. fastidiosa* subsp. *fastidiosa*) (16). In 2006, a single-nucleotide polyphormism (SNP)-based study investigating genetic variations of four strains and showing variation in the rates of genome evolution among the subspecies was published (17). Other genomic studies used such data to study recent *X. fastidiosa* outbreaks, establishing the link between European and Central American strains (9, 18, 19). However, most research that has taken advantage of genomic data has focused on the study of gene function and how *X. fastidiosa* interacts with both plant and insect hosts (20). In addition, consideration has been given to the fact that *X. fastidiosa* lacks a type III secretion system and corresponding effectors. The role of horizontally transferred elements, including integrated prophages and plasmids, has also been evaluated, but their contributions to virulence or host specificity remain unknown (2, 14).

Estimates of genetic divergence between different subspecies have ranged from 1% to 3% at synonymous sites, and the divergence has been hypothesized to be the consequence of geographic isolation over the past 20,000 to 50,000 years (21, 22). Divergence analysis was performed on a restricted number of loci using an estimated mutation rate of 5.4×10^{-10} changes per site per generation derived from a study on

Escherichia coli (23) and an assumed number of generations per year. While these estimates have been useful, the values obtained may not be accurate due the limited number of loci available for analyses and the mutation rate derived from *E. coli*. As such, no molecular evolutionary rate has been estimated for *X. fastidiosa* by using a calibration-based approach. In addition, genetic mixing among *X. fastidiosa* subspecies has been demonstrated and may have affected MLST-based estimates, as recombination affects branch length and tree topology in phylogenies and, consequently, dating estimates (24). A clear example of recombination is based on the introduction of *X. fastidiosa* subsp. *multiplex* in South America, first associated with a plum disease (25). Several studies have provided evidence of recombination between the introduced *X. fastidiosa* subsp. *multiplex* and the endemic *X. fastidiosa* subsp. *pauca* (e.g., see references 7 and 26). Furthermore, experimental research has shown host pathogen specificity within subspecies, which is indicative of ecological isolation of specific clades within the different subspecies, and the presence of gene flow between otherwise divergent populations (26, 27). Therefore, while recombination is frequently observed in *X. fastidiosa* populations, it has so far not been analyzed at the genome level with representative phylogenetic coverage of currently known *X. fastidiosa* diversity.

Recombination is believed to play a major role in pathogen evolution, as bacteria acquire novel alleles or genes that may lead to host range expansion, increments in virulence, or adaptation to new ecological niches (28–30). For this pathogen, individual nucleotides have been estimated to be three times more likely to change due to recombination than due to point mutation, based on 10 genetic loci and a relatively small sample size (3). *X. fastidiosa* recombination has been observed both under laboratory conditions (31) and in the field (21) and has been proposed to be the main driver in the emergence of *X. fastidiosa* subsp. *morus* (12). A similar hypothesis has been proposed for the emergence of coffee and citrus diseases in South America, where the endemic subspecies became pathogenic via IHR from the introduction of *X. fastidiosa* subsp. *multiplex* in the region (6). Accordingly, a recent study aiming to characterize *X. fastidiosa* diversity in South America highlighted the presence of bidirectional allelic exchange in areas where both endemic *X. fastidiosa* subsp. *pauca* and invasive *X. fastidiosa* subsp. *multiplex* were present (7). As such, the use of whole-genome sequence data should improve the understanding of the various forces shaping *X. fastidiosa* populations.

Here, we present an analysis of 72 *X. fastidiosa* genomes, including representatives of the five currently described subspecies. The study adds 36 previously unpublished genomes to the 36 strains that were available when this study was initiated. The study aimed to characterize genomic similarities and differences between the main *X. fastidiosa* subspecies. The study uses the largest whole-genome data set for *X. fastidiosa* to date and has the ability to test hypotheses regarding the evolutionary history of the pathogen. The study explored phylogeny and population structure of the pathogen and investigated gene content differences in the pangenome among the three main subspecies. Due to the availability of isolation dates for multiple isolates in the data set, the study allowed the phylogenetic tree to be calibrated, which provided estimates of both the molecular evolutionary rate of this plant pathogen and the divergence time between subspecies while deducing the introduction time for multiple *X. fastidiosa* invasions. Finally, since recombination is recognized as a major driver of genetic diversification in *X. fastidiosa*, the study estimated the rate of recombination in relation to point mutations (*r*/*m*), as well as patterns of recombination in the core genome for the different subspecies, to describe the presence of recombination in *X. fastidiosa*-host associations.

RESULTS

Phylogeny, population structure, and pangenome analyses. A total 72 genomes were included in the study; the average length of the genomes was 2,536,946 bp, with a mean N_{50} and L_{50} of 441,508 bp and 6.9, respectively. A maximum-likelihood phylogeny was constructed (Fig. 1); the resulting topology demonstrated structure among

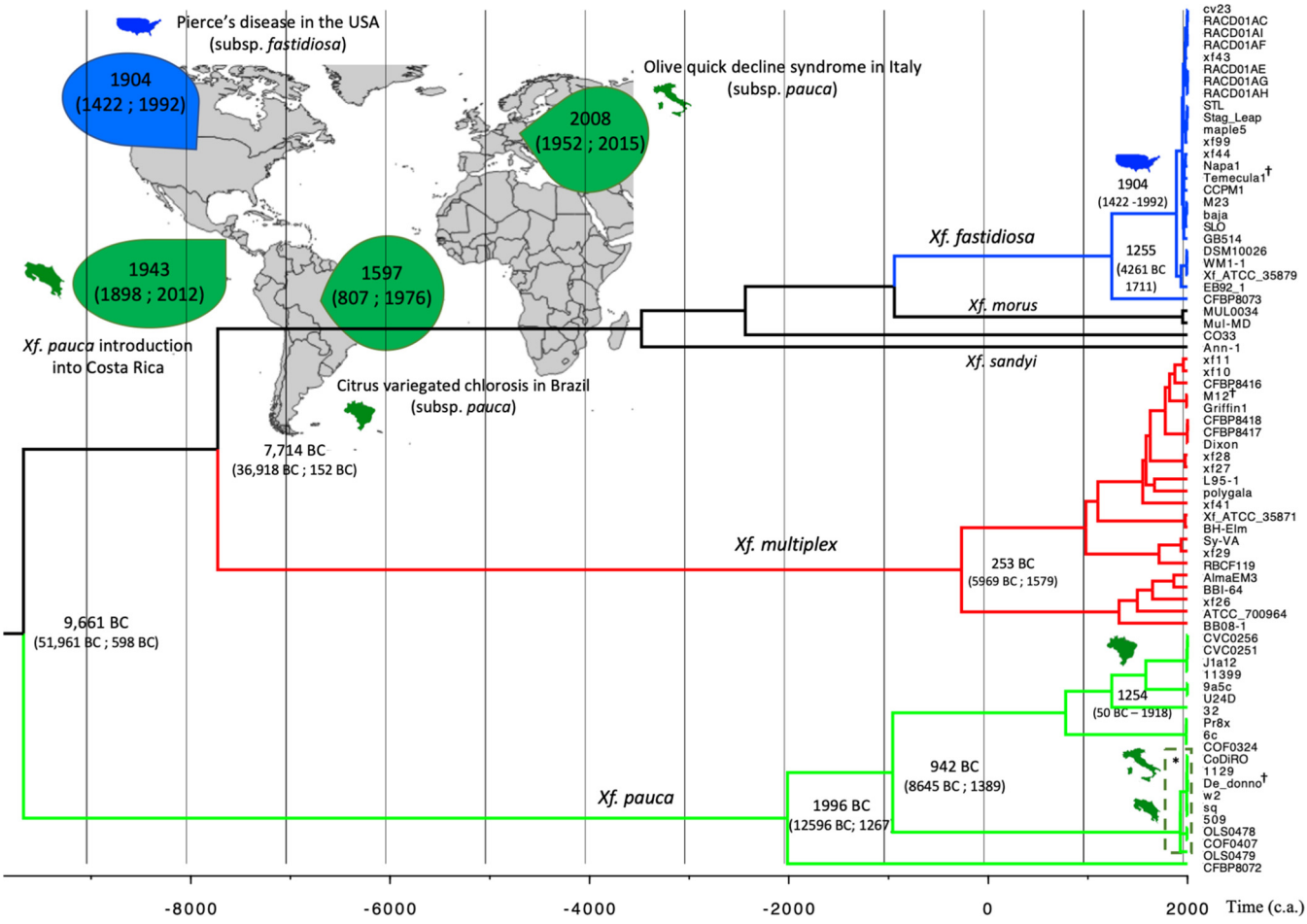

FIG 1 Clustering of *Xylella fastidiosa* (*Xf.*) isolates ($n = 72$). The three main subspecies are highlighted with blue, red, and green branches, and the estimated times of the principal introductions of the pathogen across the globe are summarized, with the confidence intervals provided. Daggers highlight the reference strains used to uncover SNPs. The broken rectangle (indicated by an asterisk) highlights the scale (i.e., clade) at which the temporal signal was found (and tip-dating performed). The map was generated with ArcGIS software.

strains. After removal of regions with evidence of recombination, a clustering analysis (Bayesian analysis of population structure [BAPS]) identified four lineages containing the three main subspecies and one lineage containing strains from *X. fastidiosa* subsp. *morus* and *sandyi*, as well as strain CO33 (see Fig. S1 and Table S2 in the supplemental material). This additional lineage is not monophyletic and contains only four highly divergent strains; the taxonomic status of these strains and lineage(s) may be better resolved with additional data. When investigating population structure without accounting for recombination, a subsequent division was observed within *X. fastidiosa* subsp. *pauca* between South American strains and strains from Italy and Central America (Fig. S1). In other words, *X. fastidiosa* taxonomy may vary depending on the data used and should be revisited in light of these findings.

Pangenome analyses identified 622 and 291 genes in core (present in 99% of isolates) and softcore (present in 95 to 99% of isolates) genomes at the species level (Table 1). *X. fastidiosa* subsp. *pauca* displayed the highest nucleotide diversity (73,906 SNPs, $\pi = 5.27 \times 10^{-3}$) among subspecies (Table 2). *X. fastidiosa* subsp. *fastidiosa* was the most clonal (14,200 SNPs; $\pi = 7.51 \times 10^{-4}$), with the highest relative effect of recombination to mutation ($r/m = 4.059$). *X. fastidiosa* subsp. *multiplex* had a total of 37,485 SNPs when mapped to the reference strain *X. fastidiosa* subsp. *multiplex* M12 and a genetic diversity (π) of 2.39×10^{-3}. Waterson's estimator (θ) was the lowest among *X. fastidiosa* subsp. *fastidiosa* isolates ($\theta = 1.52 \times 10^{-3}$) and the highest in *X. fastidiosa* subsp. *pauca* ($\theta = 4.26 \times 10^{-3}$) (Table 2). The number of prophage regions

TABLE 1 Genes in the *X. fastidiosa* core and accessory genomes, including the three main subspecies studied[a]

Clade	No. of genes present in:			
	Core	Softcore	Shell	Cloud
X. fastidiosa ($n = 72$)	615	285	2,926	6,996
X. fastidiosa subsp. *fastidiosa* ($n = 25$)	1,282	460	867	790
X. fastidiosa subsp. *multiplex* ($n = 23$)	828	399	2,250	2,978
X. fastidiosa subsp. *pauca* ($n = 20$)	1,516	143	2,096	1,123

[a]The core genome represents genes present in 99% of strains, the softcore genome represents genes present in 95% to 99% of strains, the shell includes those present in 15% to 94% of strains, and the cloud includes genes found in less than 15% of the strains.

was slightly higher in *X. fastidiosa* subsp. *fastidiosa* (9.54) and *multiplex* (9.33) than in *X. fastidiosa* subsp. *pauca* (8.55).

Pangenome analyses using Roary revealed variations in the accessory genomes of the three main subspecies (Table 1; Fig. S1). *X. fastidiosa* subsp. *multiplex* displayed the highest number of genes ($n = 6,465$), followed by *X. fastidiosa* subsp. *pauca* ($n = 4,747$) and *X. fastidiosa* subsp. *fastidiosa* ($n = 3,399$). *X. fastidiosa* subsp. *fastidiosa* ($n = 22$) was the most conserved species, with the core/softcore genome accounting for 51% of the pangenome ($n = 1,742$ genes). The accessory genome (shell and cloud genome) of *X. fastidiosa* subsp. *multiplex* was the largest, accounting for 81% of the pangenome ($n = 5,228$ genes).

The genomic sequences identified using Roary were then subjected to a BLAST search to identify genes in the accessory genomes of each subspecies. In *X. fastidiosa* subsp. *fastidiosa*, 59 genes were uniquely found in the subspecies. Gene Ontology (GO) analyses revealed gene functions (Table S8), and one alanine racemase gene, *alr,* was found in all *X. fastidiosa* subsp. *fastidiosa* isolates. PD_1249 encodes a hemagglutinin and was found in 18 strains, while PD_1802, a polysaccharide deacetylase gene, was found in 14 strains; finally, *birA*, which encodes a biotin ligase, was present in 8 *X. fastidiosa* subsp. *fastidiosa* strains. In *X. fastidiosa* subsp. *multiplex*, 457 genes were specifically found, including *coq7,* which was present in 20 of the 23 *X. fastidiosa* subsp. *multiplex* strains and is known to be involved in metal ion binding and oxidation-reduction processes. In 19 *X. fastidiosa* subsp. *multiplex* strains, the genes *tgpA* and *gluP*, two membrane component genes, and *mrdB,* involved in the regulation of cell shape, were detected. Thirteen strains were found to have PD_0310, a gene known to be involved in pathogenicity (32). Two genes, *isf* and *etfD,* were found in 20 and 19 strains, respectively, and are involved in oxidation-reduction activity. Six *X. fastidiosa* subsp. *multiplex* strains carried the *xadA* gene, which is involved in *X. fastidiosa* adhesion to surfaces (33). Ninety-six genes were found only in *X. fastidiosa* subsp. *pauca*, including 8 genes that were present in over half of the *X. fastidiosa* subsp. *pauca* strains used in the study: *mtgA, lptE, pilQ, rpmD, aat, rimP, egl,* and *pdxH*. Additionally, 4 genes were uniquely associated with *X. fastidiosa* subsp. *sandyi* (Table S8).

Divergence time estimation at the species level. To investigate the temporal evolution of *X. fastidiosa*, an ML phylogeny was built using the core genome of 43 isolates (911,604 nucleotides [nt] and 1,384 SNPs) spanning 33 years (1983 to 2016) of evolution. Sufficient temporal signal (i.e., detectable amount of evolutionary change) was found to be present only at the scale of a small clade ($n = 9$ strains) within *X. fastidiosa* subsp. *pauca* (Fig. 1 and Fig. S3). Potential reasons for the lack

TABLE 2 Summary statistics for genomic diversity of *X. fastidiosa* and three subspecies[a]

Clade	No. of SNPs	Core (nt)	π	θ	*r/m*	Tajima's D
X. fastidiosa ($n = 72$)	311,435	498,968	1.29×10^{-2}	9.51×10^{-3}	2.259	1.279
X. fastidiosa subsp. *fastidiosa* ($n = 25$)	14,200	1,167,862	7.51×10^{-4}	1.52×10^{-3}	4.059	−2.085
X. fastidiosa subsp. *multiplex* ($n = 23$)	37,485	736,868	2.39×10^{-3}	2.62×10^{-3}	2.654	−0.495
X. fastidiosa subsp. *pauca* ($n = 20$)	73,906	1,512,832	5.27×10^{-3}	4.26×10^{-3}	3.727	1.010

[a]The entire data set was mapped against *X. fastidiosa* subsp. *fastidiosa* strain Temecula1; for each subspecies, the data were mapped to its own reference.

of temporal signal throughout the phylogeny may include, but are not limited to, the short time span of the collection dates for strains in comparison to the age of the most recent common ancestors (MRCAs) and limited data. Performing tip-dating inference using BEAST at this evolutionary scale allowed inference of a substitution rate of 7.6204×10^{-7} per site per year (95% confidence interval [CI], 1.024×10^{-7} to 1.454×10^{-6}). When extrapolating this evolutionary rate (and its uncertainty) across the whole species using BEAST, we obtained a value of ucld.stev (the standard deviation of the uncorrelated log-normal relaxed clock) of 0.517, suggesting little variation (ucld.stdev < 1) in rates among branches. When comparing the distribution rates for *X. fastidiosa* subsp. *pauca*, *multiplex*, and *fastidiosa* (the other subspecies could not be included in this comparison because of the small sample size), we did not find significant differences in rates among the subspecies (Fig. S3). By rate-dating the whole tree, divergence between *X. fastidiosa* subsp. *pauca* and the other subspecies was estimated to date back to 9661 Before Common Era (BCE) (CI, 51961 BCE to 598 BCE) and the split between *X. fastidiosa* subsp. *fastidiosa* and *multiplex* was estimated at 7714 BCE (CI, 31918 BCE to 152 BCE) (Fig. 1). Within *X. fastidiosa* subsp. *fastidiosa,* the introduction of *X. fastidiosa* subsp. *fastidiosa* associated with PD in the United States was estimated to date to ca. 1904 (CI, 1423 CE to 1993 CE). In the case of PD, the times to most common recent ancestor (TMRCA) were similar for isolates from the western and eastern United States ($TMRCA_{West} = 1963$ CE; $TMRCA_{East} = 1951$ CE). Within *X. fastidiosa* subsp. *pauca*, the branch that included strains associated with CVC was estimated to have originated in 1597 (CI, 807 CE to 1976 CE) and the TMRCA for strains associated with CVC and coffee leaf scorch (CLS) in Brazil dated to 1254 (CI, 50 BCE to 1918 CE). *X. fastidiosa* subsp. *pauca* isolates found in Costa Rica were dated to 1943 (CI, 1898 CE to 2012 CE), and the subsequent introduction into Apulia, Italy, was estimated to date to 2008 (CI, 1952 CE to 2015 CE).

Selection. *X. fastidiosa* subsp. *fastidiosa* and *multiplex* displayed strong overall negative Tajima's D values (-2.085 and -0.495, respectively), indicating signs of adaptation (Table 2). Selection forces affecting the different lineages were then investigated at the gene level using a phylogenetic gene-based approach (ratio of nonsynonymous to synonymous evolutionary changes [*dN*/*dS*]). Fourteen genes with an excess of nonsynonymous mutations in each of the subspecies were obtained (Table S3). Each subspecies had different genes under selection. The 15 genes with the highest *dN*/*dS* values for each subspecies were almost exclusively restricted to each clade (Table 3). The number of genes found with a *dN*/*dS* ratio of >1 uniquely associated with a lineage was higher in *X. fastidiosa* subsp. *pauca* (90 genes), followed by *X. fastidiosa* subsp. *fastidiosa* (76 genes) and *multiplex* (43 genes). Only PD_1285 was found in the top 15 extreme *dN*/*dS* values for both *X. fastidiosa* subsp. *fastidiosa* and *multiplex*; PD_1285 has a putative conserved RseA domain, a negative regulator of σ^E activity, which suppresses transcription of heat shock response promoters. Four genes with high *dN*/*dS* values were associated with GO terms: *pilY1* (PD_1611) was associated with a pilus tip-associated protein (34), PD_1506 with a hemolysin-type calcium binding protein involved in pathogenesis (GO:0009405), *frpC* with a peptidase; and PD_1671 with a GGDEF domain-containing response regulator (35).

The Gene Ontology analysis revealed that genes under selection were involved in different functions (Table S6). For *X. fastidiosa* subsp. *fastidiosa*, pilus assembly, N-terminal protein amino acid acetylation, response to oxidative stress, and manganese ion transmembrane transport were commonly found. In *X. fastidiosa* subsp. *multiplex*, genes involved in lipoprotein transport or response to antibiotics (*ponB* and *trxB*) were identified. Some pathogenesis-related genes were found in both *X. fastidiosa* subsp. *fastidiosa* (*frpC*) and *multiplex* (*hsf*). The dominating GO terms found in *X. fastidiosa* subsp. *pauca* were associated with responses to nitroactive stress (*hmp*), biofilm formation (*lpdA*), or oxygen transport.

TABLE 3 Fifteen highest *dN*/*dS* values for each of the three main subspecies, highlighting potential differences in selective pressures for each clade

Subspecies and gene	*dN*/*dS*
X. fastidiosa subsp. *fastidiosa*	
PD_1472	9.172
PD_0299	8.586
frpC	7.182
PD_0511	3.735
bcp	3.638
pspA_3	3.361
PD_1100	3.013
mrkD	2.646
pspB	2.546
PD_1285	2.45
tnp	2.449
pilB	2.334
PD_1138	2.222
PD_1128	2.198
rsuA	2.096
X. fastidiosa subsp. *multiplex*	
xerC	7.884
PD_0518	4.766
mucD	4.605
PD_1347	3.299
PD_0768	2.575
PD_1731	2.390
hemC	2.258
tolA	2.141
PD_1352	2.108
PD_1669	1.971
PD_1112	1.951
recD	1.850
PD_1285	1.785
PD_1671	1.750
PD_1003	1.710
X. fastidiosa subsp. *pauca*	
minE	6.499
PD_1593	5.345
PD_0978	4.738
PD_1843	4.109
rsmG	3.746
ctp	3.120
PD_2025	3.030
PD_1717	3.017
PD_0312	2.510
PD_1173	2.404
PD_1275	2.382
tonB_3	2.104
PD_0965	2.076
PD_1249	1.949
nadA	1.939

Recombination at the species level. Recombination analyses used the core genome alignment of the 72 strains (498,968 bp) (Fig. 2). The fastGEAR algorithm detected genomic segments that had undergone recent homologous recombination events among and within the different lineages. A total of 1,035 recombining segments were identified and were subsequently searched using BLAST; 237 genes were mapped to the reference genome (Table S7). To compare the performance of the fastGEAR software, a similar analysis was performed in ClonalFrameML, which identified 370 genes. A total of 204 genes, which represents 86.0% of the genes uncovered using the fastGEAR algorithm, were found in both analyses (Fig. S4). Therefore, fastGEAR appeared more conservative, and the subsequent analysis used the fastGEAR algorithm, as it permitted identification of the origin of recombination events. *X. fastidiosa* subsp.

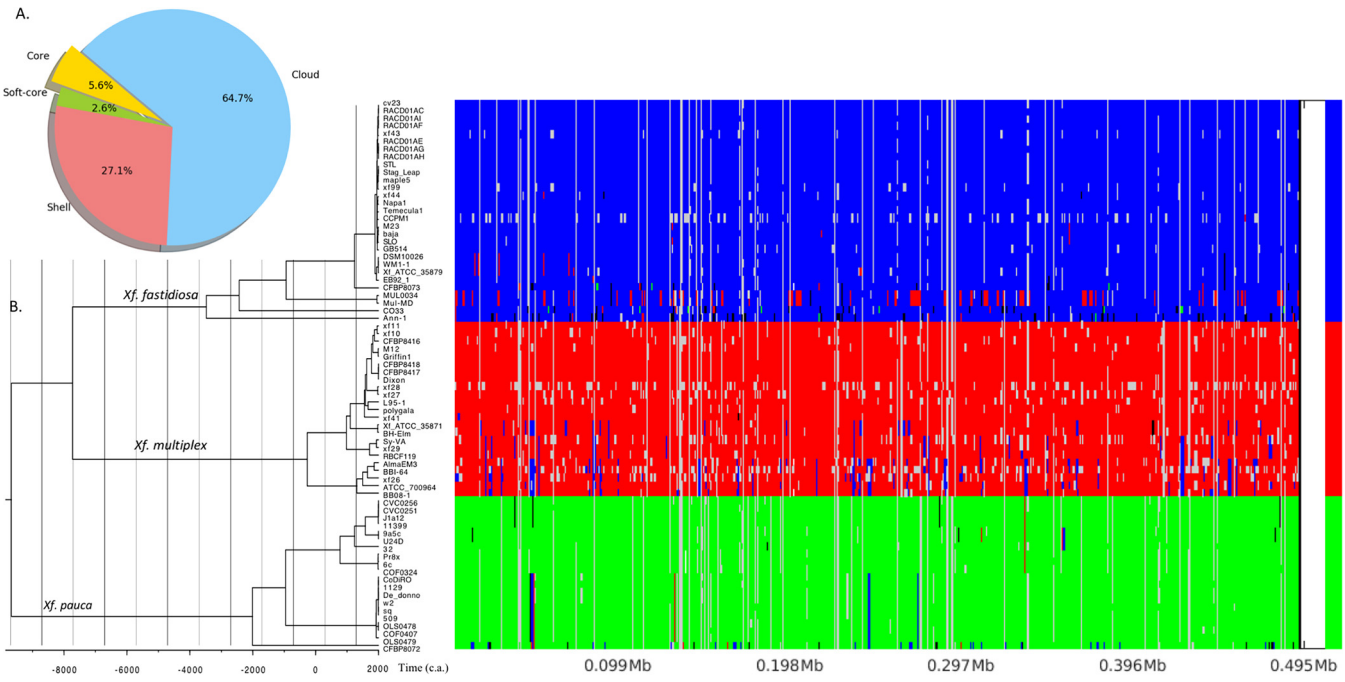

FIG 2 (A) Proportion of core and accessory genes in the *X. fastidiosa* genomes analyzed. (B) Recombination events estimated using FastGEAR. Three lineages corresponding to *X. fastidiosa* subsp. *fastidiosa* (blue), *pauca* (red), and *multiplex* (green) were obtained. *X. fastidiosa* subsp. *pauca* was composed of two genetic clusters, ST53 and Brazilian citrus/coffee strains. Similarly, in the lineage that includes *X. fastidiosa* subsp. *fastidiosa*, two genetic clusters were found, one incorporating *X. fastidiosa* subsp. *fastidiosa* strains and one including the two *X. fastidiosa* subsp. *sandyi* strains and the two *X. fastidiosa* subsp. *morus* strains.

multiplex exhibited the highest number of recombination events (221 genes), with 93.2% originating from *X. fastidiosa* subsp. *fastidiosa* (206 genes) (Table S7). In *X. fastidiosa* subsp. *fastidiosa*, 10 strains presented recombining regions (40%) for a total of 85 recombining segments, with 44% of the observed recombination events ($n = 37$ segments, 2.43% of the core genome) occurring in the Mexican strain CFBP8073.

Seventeen of the 23 *X. fastidiosa* subsp. *multiplex* strains (73.3%) had recombining regions, with an average of 9.61 genes per genome (average of 2.76% of the core genome), whereas among the 9 recombining *X. fastidiosa* subsp. *fastidiosa* strains from the United States, the average was 3.22 genes per genome (0.32% of the core genome). Within *X. fastidiosa* subsp. *multiplex*, 8 strains recombined with the *X. fastidiosa* subsp. *fastidiosa* lineage (Fig. 2). These strains were obtained in the United States from blueberry ($n = 3$), *Polygala myrtifolia* ($n = 1$), *Platanus occidentalis* ($n = 1$), *Sycamore* sp. ($n = 1$), *Ambrosia* sp. ($n = 1$), and an unknown host ($n = 1$). Genes with evidence of recombination were involved in iron ion transport, dephosphorylation, or response to stress. A recombinant *lexA* gene was detected in *X. fastidiosa* subsp. *multiplex* ATCC 35871, a strain from plum in Georgia, USA. This gene also was found to be under positive selection among *X. fastidiosa* subsp. *fastidiosa* strains and was found to have originated from the *X. fastidiosa* subsp. *fastidiosa* lineage. The *lexA* gene functions as a repressor of the SOS response to DNA damage. In *X. fastidiosa* subsp. *fastidiosa*, a recombinant region found in strain WM1-1 included *htpG*, involved in stress response, but most GO terms that were associated with the subspecies were involved in copper ion homeostasis, such as *copB* or PD_2039, part of a multicopper oxidase family.

The ST53 *X. fastidiosa* subsp. *pauca* cluster presented only five recombining segments, including two genes from the *X. fastidiosa* subsp. *fastidiosa* lineage that were found in all strains of the ST53 cluster (as observed in Fig. 2): *cyoA,* encoding ubiquinol oxidase subunit II involved in copper ion binding activity (GO:0005507), and *ilvD*, coding for a dehydratase involved in metal ion binding activity. In the second *X. fastidiosa* subsp. *pauca* genetic cluster (the CVC cluster), 34 recombining segments were detected: 32.4% ($n = 11$) originating from *X. fastidiosa* subsp. *multiplex* and 17.6%

($n = 6$) originating from *X. fastidiosa* subsp. *fastidiosa*. GO term analysis revealed that these genes were involved in spermidine synthesis, protein transport, cell wall organization, and oxidation-reduction processes (Table S7).

X. fastidiosa subsp. *morus* and *sandyi* had high numbers of recombining segments, with 75.5 segments per strain for *X. fastidiosa* subsp. *morus* (15.30% of the core genome) and 81.0 for *X. fastidiosa* subsp. *sandyi* (9.06% of the core genome). The origin of the recombination seems to be different for the two subspecies, as most recombining elements in *X. fastidiosa* subsp. *morus* originated from *X. fastidiosa* subsp. *multiplex* (84.1%), compared to only 19.8% in *X. fastidiosa* subsp. *sandyi*, in which the majority of recombination elements originated from unsampled lineages (72.8%). The transcriptional regulator gene, *lexA*, was also identified in *X. fastidiosa* subsp. *morus* strains as part of the recombining segments, as well as other genes involved in cell division, dephosphorylation, or protein synthesis. In *X. fastidiosa* subsp. *sandyi*, spermidine biosynthesis, copper ion homeostasis, and phosphate ion transmembrane transport elements were also subject to recombination.

Strain CO33 recombined as much as *X. fastidiosa* subsp. *sandyi* ($n = 73$; 6.54% of the core genome), again primarily with an unsampled genetic cluster (58.1%). Genes involved in response to stress, such as *dps*, encoding a DNA starvation stationary-phase protection protein, were detected. Some overlap was observed between recombination and selection analyses, indicative of the relevance of a particular gene in the adaptation of the pathogen. For instance, PD_1974 (hypothetical protein) and PD_1613 (pilus assembly) were under selection in the *X. fastidiosa* subsp. *fastidiosa* lineage and also present as recombinant regions in strains BH-Elm and ATCC_700964 (*X. fastidiosa* subsp. *multiplex*), respectively.

Case study: recombination between coffee- and citrus-infecting strains in *X. fastidiosa* subsp. *pauca*. A similar approach was taken to identify exchanges among *X. fastidiosa* subsp. *pauca* strains from coffee and citrus by mapping recombining segments to the *X. fastidiosa* subsp. *pauca* 9a5c reference. The analyses included 19 *X. fastidiosa* subsp. *pauca* strains and revealed three genetic lineages within the subspecies (Fig. 3). The first lineage was composed of three strains from Brazil (coffee, $n = 2$; plum, $n = 1$). The second lineage was composed of two subsequent genetic clusters, one including three recombining strains: 9a5c (citrus), U24D (citrus), and 32 (coffee). The second genetic cluster of this lineage was composed of four nonrecombining strains from citrus. The remaining lineage (lineage 3) was composed of the ST53 strains from Italy and Costa Rica and showed no sign of recombination within *X. fastidiosa* subsp. *pauca*. Among recombining lineages, two strains from citrus (9a5c and U24D) had 167 genetic elements identified as recombinant (Table S4), 70 (41.8%) from lineage 1 (coffee, $n = 2$; plum, $n = 1$), 33 (19.4%) from lineage 3 (olive, Italy, $n = 6$; oleander, Costa Rica, $n = 2$; coffee, Costa Rica, $n = 1$) and 64 (38.2%) from unsampled lineages (Table 4). The other *X. fastidiosa* subsp. *pauca* recombination strain, 32 (coffee), contained 167 recombinant genes (Table S5), 62 (37.8%) from lineage 1, 35 (21.2%) from lineage 3, and 70 (42.4%) from other lineages. Among recombining genes mapped to the 9a5c reference, 54 were present in both coffee and citrus strains originating principally from lineage 1, including genes involved in putrescine biosynthesis or heme transport (Fig. S3).

DISCUSSION

Phylogeny, population structure, and pangenome analyses. Strong population structure was observed, and after removal of recombining segments, BAPS properly assigned the three main subspecies to three lineages. Based on this analysis, *X. fastidiosa* subsp. *morus* and *sandyi* and the CO33 strain constituted a different lineage. Interestingly, BAPS and admixture analysis identified two groups within *X. fastidiosa* subsp. *pauca,* one South American cluster and the same previously described monophyletic clade that included Italian strains and other ST53 strains from Costa Rica (4). Additional analyses investigating *X. fastidiosa* subsp. *pauca* population structure found

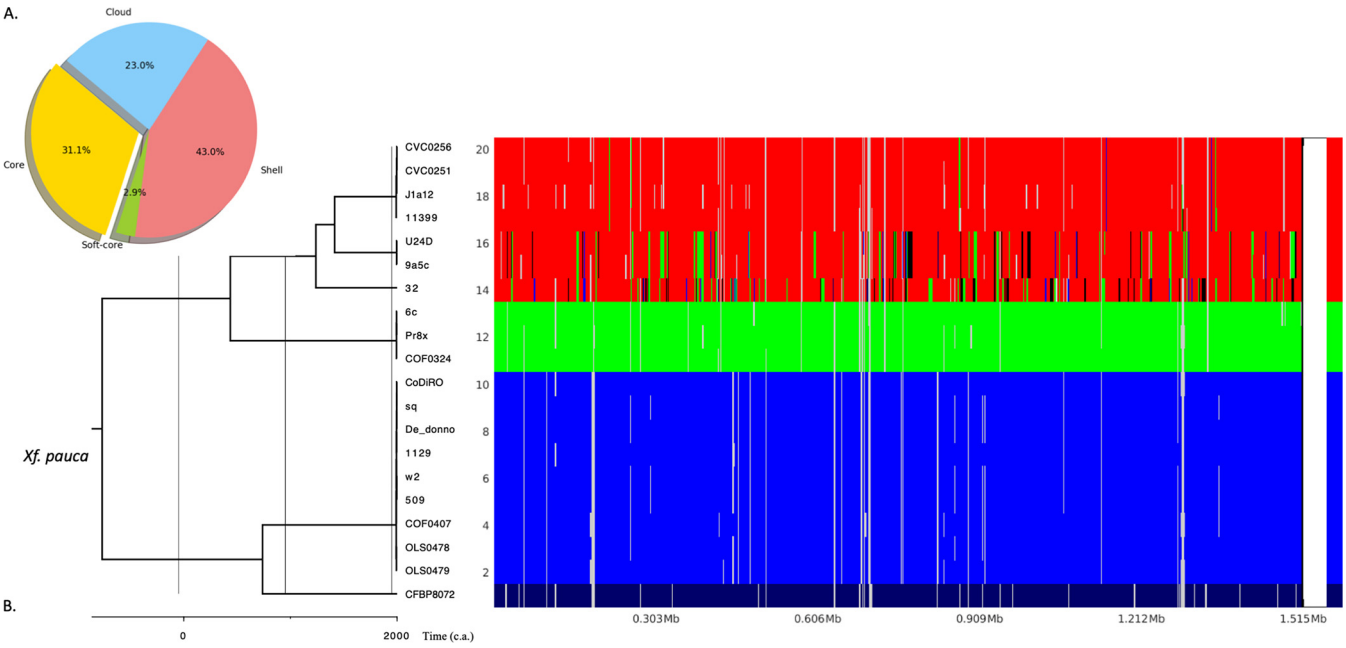

FIG 3 (A) Proportion of core and accessory genes in the *X. fastidiosa* subsp. *pauca* genomes analyzed. (B) Recombination events within *X. fastidiosa* subsp. *pauca* estimated using FastGEAR. Three lineages were obtained: lineage 1 (green) was composed of strains found on coffee ($n = 2$) and plum ($n = 1$); lineage 2 (red) was composed of two subsequent genetic clusters, one consisting of three recombining strains, i.e., *X. fastidiosa* 9a5c (citrus), *X. fastidiosa* U24D (citrus), and *X. fastidiosa* 32 (coffee); and lineage 3 (blue) was the ST53 genetic cluster (olive, Italy [$n = 6$]; oleander, Costa Rica [$n = 2$]; coffee, Costa Rica [$n = 1$]). Additionally, the CFBP8072 strain belongs to this lineage but constituted a different genetic cluster (dark blue).

four genetic clusters in the subspecies and, within the South America monophyletic clade, identified three recombining strains: 9a5c (citrus), UD24 (citrus), and 32 (coffee).

The *X. fastidiosa* core/softcore genomes were composed of only 900 genes, highlighting the importance of the accessory genome within *X. fastidiosa* as a species. Genetic diversity was higher in *X. fastidiosa* subsp. *pauca* ($\pi = 5.27 \times 10^{-3}$) and lower in *X. fastidiosa* subsp. *fastidiosa* ($\pi = 7.51 \times 10^{-2}$). These results were corroborated by the high number of SNPs found among *X. fastidiosa* subsp. *pauca* strains (73,906 SNPs) and the limited number of polymorphic sites identified in *X. fastidiosa* subsp. *fastidiosa* (14,200 SNPs). *X. fastidiosa* subsp. *multiplex* harbored 37,485 SNPs, but its core genome appeared smaller than those of the other two subspecies (736,868 nt) while its accessory genome comprised 81% of the genes found within the subspecies. The extensive accessory genome of *X. fastidiosa* subsp. *multiplex* might be a consequence of its extensive host range. The presence of genes such as *coq7, isf*, and *etfD,* which are involved in oxidation reduction processes, might contribute to host adaptation. For instance, six strains were found to have the *xadA* gene, which is important in *X. fastidiosa* adhesion to surfaces and has been thought to play a role in *X. fastidiosa* colonization of plants (32, 33). We note that while this study attempted to improve the representation of *X. fastidiosa* phylogenetics at the genome level, the data are still limited and subject to sampling biases; furthermore, the use of an automated annotator introduces its own biases and limitations into the performed analyses.

TABLE 4 Donor lineage and recipient strain of recombining fragments in *X. fastidiosa* subsp. *pauca*

Recipient strain(s)	No. of recombining fragments in:			Total no.
	Donor lineage 1	Donor lineage 3	Other lineage	
9a5c & U24D (citrus)	70	33	64	167
32 (coffee)	62	35	70	167
Shared	25	8	17	50

Divergence time estimation at the species level. The *X. fastidiosa* subsp. *pauca* subclade was the only group displaying a significant temporal signal, allowing for the first estimation of the molecular evolutionary rate (7.6204×10^{-7} substitutions per site per year; CI, 1.024×10^{-7} to 1.454×10^{-6}) for *X. fastidiosa* using a tip-calibrated approach. This finding represents an improvement in the study of *X. fastidiosa* evolutionary history, as previous dating estimates were performed using the mutation rate of *E. coli* and an arbitrary number of generations per year (22). The rate obtained is comparable to those obtained for several human and wildlife pathogens, spanning the range from 10^{-5} to 10^{-8} nucleotide substitutions per site per year (36). Analyses were then extended to the entire phylogenic tree, assuming a constant rate of evolution for the different lineages. Molecular dating revealed that the split between *X. fastidiosa* subsp. *pauca* and the other two subspecies occurred around calendar year (ca.) 10000 BCE; the split was previously estimated to have occurred 60,000 years ago based on MLST (6). The present estimates are probably more precise, as this study contains more isolates and is based on a significantly greater proportion of the genome. The division between *X. fastidiosa* subsp. *fastidiosa* and *multiplex* was estimated at 8000 BCE; this estimate is slightly more recent than the ones using MLST data (6, 22), which estimated that this split occurred between 15,000 and 30,000 years ago. Similarly, our estimates suggest that divergence occurred between *X. fastidiosa* subsp. *fastidiosa* and *sandyi* ca. 3400 BCE versus 20,000 to 40,000 years ago, as previously estimated. Although our results suggest a very low level of rate heterogeneity among subspecies, the dating estimates presented here based on the extrapolation of a substitution rate estimated at the scale of an *X. fastidiosa* subsp. *pauca* subclade should be taken with precaution. We hope that these estimates will be refined by additional studies in which the molecular rate for each subspecies will be intrinsically estimated. Additional data also will improve inferences and likely reduce confidence intervals, providing more accurate information.

Selection. Selective signatures within the genome are imposed by local biotic and abiotic pressures and may contribute to fitness variation among *X. fastidiosa* subspecies. Here, we provide evidence of variation among these allopatric lineages. Strong overall negative Tajima's D values in *X. fastidiosa* subsp. *fastidiosa* (−2.085) and *X. fastidiosa* subsp. *multiplex* (−0.495) might be the reflection of adaptation. *X. fastidiosa* subsp. *fastidiosa* displayed a strong negative value that might be the result of recent population expansion or migration. Using the gene-based approach, an excess of nonsynonymous mutations was observed within all three subspecies in 14 genes. When the 15 genes with the highest *dN*/*dS* values are examined, it is apparent that *X. fastidiosa* subspecies are under different selective pressures, as no overlap is observed except with PD_1285, an RseA domain gene, which is present in both *X. fastidiosa* subsp. *fastidiosa* and *multiplex*. In *E. coli,* RseA is an inner membrane protein that binds to σ^E, acting as a negative regulator of the signal response to extracytoplasmic stress such as heat shock controlled by σ^E (37). It is worth noting that when *dN*/*dS* methods are applied to microbial populations, several issues are presented (reviewed in reference 38). Patterns of polymorphism are expected to appear more neutral than patterns of fixation, and as a consequence, some genes under positive selection might not display an increased *dN*/*dS* ratio.

When the ontology of genes under selection in *X. fastidiosa* subsp. *pauca*, *fastidiosa,* and *multiplex* was investigated, 17 GO terms were commonly found for genes involved in cell metabolism, including oxidation-reduction processes, ATP and zinc binding activity, cell division, translation, and proteolysis. Different functions were identified in the three subspecies. For instance, among *X. fastidiosa* subsp. *fastidiosa* strains, *hemY* was under selection. This gene is involved the heme metabolic process and is essential for oxidization of coproporphyrinogen III (39). A glutathione peroxidase gene, *gpo,* involved in response to oxidative stress, was also specific to *X. fastidiosa* subsp. *fastidiosa*. In *X. fastidiosa* subsp. *multiplex*, genes under selection are involved in cell wall organization (*ponB*), ATP-binding cassette transport (*msbA*), or the removal of

superoxide radical (*trxB*). In *X. fastidiosa* subsp. *pauca*, gene ontology analysis revealed genes involved in various metabolic pathways, notably *hmp*, a nitric oxide dioxygenase gene which is known to provide resistance to nitrosative stress (40). A dihydrolipoamide dehydrogenase gene, *lpdA,* was also identified as under selection and is highly expressed during biofilm formation (41).

Recombination at the subspecies level. Recombination in *X. fastidiosa* has been hypothesized to be the main driver of host adaptation, although data used in prior studies have been based on a small number of housekeeping genes (i.e., MLST data). We were able to investigate homologous recombination in the core/softcore genome of the 72 *X. fastidiosa* genomes. Recombination contributed more to genetic diversity than mutation, with an overall *r*/*m* value of 2.259, a relatively high value compared to that of other bacteria (29). Our *r*/*m* value at the species level was slightly lower than previous estimations made by Scally et al. (3), which were based on 10 genes. The frequency of recombination between subspecies differs among strains, even within subspecies. This could reflect geographic isolation, where recombination between subspecies has occurred only as strains have been introduced to new regions where other subspecies were already established. Nine strains of *X. fastidiosa* subsp. *fastidiosa* isolated in the United States were found to have recombining segments. The average length of these recombining regions accounted for an average of 0.32% of the core genome. These strains were found predominantly in California, where one introduced genotype likely spread in grapevines through the region. The low level of recombination observed within this U.S. *X. fastidiosa* subsp. *fastidiosa* strain suggests rapid clonal expansion and ecological isolation. Recombination of *X. fastidiosa* subsp. *fastidiosa* with other subspecies might be occurring in Central America, where more genetic diversity can be found (12). In this study, the most recombining *X. fastidiosa* subsp. *fastidiosa* strain originated in Mexico, CFBP8073, and accounted for 44% of the total recombination within the subspecies (2.43% of the core genome). Additionally, a significant proportion of recombining segments originated from unsampled genotypes, indicating that the diversity of *X. fastidiosa* subsp. *fastidiosa* is unlikely to have been fully described.

This study supports the hypothesis that IHR is associated with shifts to novel plant hosts and that recombination is important in *X. fastidiosa* evolution. *X. fastidiosa* subsp. *multiplex* displayed the highest number of recombination events (221 genes) originating mainly from the *X. fastidiosa* subsp. *fastidiosa* lineage (93.2%). *X. fastidiosa* subsp. *multiplex* appears as the lineage that is most prone to recombination, with 17 recombining strains and an average of 2.76% of recombining elements in the core genome (six strains with $\geq$4%). *X. fastidiosa* subsp. *multiplex* genotypes prone to recombination formed defined phylogenetic groups that have been previously described by Nunney et al. (21). All *X. fastidiosa* subsp. *multiplex* strains present in the study were strains from the United States, where *X. fastidiosa* subsp. *fastidiosa* also occurs; however, recombination patterns could not be linked to a particular host.

Ann-1 (*X. fastidiosa* subsp. *sandyi*) displayed 9.06% of recombining regions in the core genome. This finding was similar in proportion to that observed in CO33 strains (6.54%). The two *X. fastidiosa* subsp. *morus* strains showed a high number of segments coming from the *X. fastidiosa* subsp. *multiplex* lineage (84.10%), confirming the association of IHR with the ability of the bacterium to adapt to new hosts. Studies based on MLST estimated recombination based on 8 loci to have occurred in 50% of the investigated loci (12). In light of these analyses, these recombination events in *X. fastidiosa* subsp. *morus* accounted for 15.30% of the core genome. While it seems likely that the accessory genome also played a central role in adaptation to mulberry, the data do not support that *X. fastidiosa* subsp. *morus* is the result of large-scale genome-wide recombination between *X. fastidiosa* subsp. *fastidiosa* and *multiplex*.

Insights on *X. fastidiosa* subsp. *fastidiosa*. One isolate from Mexico (CFBP8073) did not cluster within *X. fastidiosa* subsp. *fastidiosa*. Divergence of this strain in relation to the rest of the clade was estimated to have occurred in 1255 CE. It should be noted

that Central America harbors a higher level of *X. fastidiosa* subsp. *fastidiosa* genetic diversity than North America, but only MLST data are available (6, 12). It is likely that additional strains from Central America and Mexico will improve phylogenetic coverage of *X. fastidiosa* diversity, particularly for *X. fastidiosa* subsp. *fastidiosa*. Based on the current data set, it appears clear that *X. fastidiosa* subsp. *fastidiosa* was introduced into the United States ca. 1904 (CI, 1423 CE to 1993 CE). PD was first reported in 1892 (42). Additional sampling may decrease the confidence interval and allow a better estimation of the age of the *X. fastidiosa* clade causing PD in the United States, but these data reinforce the hypothesis that PD emerged from an introduction from Central America into the United States (43).

Insights on *X. fastidiosa* subsp. *pauca* in South America. The first record of *X. fastidiosa* in South America is associated with the emergence of CVC in Brazil in 1987 (44). However, the TMRCA for these isolates was estimated to be 1597 CE (CI, 807 CE to 1976 CE). Moreover, the TMRCA among *X. fastidiosa* subsp. *pauca* strains infecting citrus and strains affecting coffee plants in Brazil was estimated to be 1254 CE. *X. fastidiosa* subsp. *pauca* is thought to have recently come in contact and acquired its pathogenic potential via homologous recombination with *X. fastidiosa* subsp. *multiplex*, which was proposed to have been introduced into the continent 80 years ago (25, 45). Unfortunately, the data set does not contain any *X. fastidiosa* subsp. *multiplex* isolates from South America, and therefore, we were unable to date the introduction. Within the CVC and CLS cluster, isolates had genetic segments acquired from both *X. fastidiosa* subsp. *fastidiosa* (17.6%) and *multiplex* (32.4%). These results confirm that *X. fastidiosa* subsp. *pauca* had been in contact with *X. fastidiosa* subsp. *multiplex* but that it also acquired genes from *X. fastidiosa* subsp. *fastidiosa*. Additionally, the data set contained one isolate from Ecuador (CFBP8072). TMRCA between this isolate and other strains in the clade date to 1996 BCE (CI, 12596 BCE to 1267 CE), indicating that *X. fastidiosa* subsp. *pauca* is of South American origin, as previously suggested (6).

This study supported the monophyletic clade of *X. fastidiosa* subsp. *pauca* infecting citrus trees in South America (7, 26). When recombination was investigated among *X. fastidiosa* subsp. *pauca*, four genetic lineages were identified. Three recombining strains (9a5c, citrus; U24D, citrus; 32, coffee) belonging to the same lineage as the four nonrecombining strains from citrus were identified. *X. fastidiosa* was first reported from citrus in Brazil in 1987 (46); our results provide additional evidence of IHR in citrus strains, as most of the recombination observed originated from coffee strains. *X. fastidiosa* seems to have been present in South America for centuries, and strains associated with CVC and CLS formed monophyletic clades with a distant origin.

Insights into the ST53 clade. Molecular dating allowed reconstruction of the introduction history of the ST53 clade that is responsible for a large outbreak of OQDS in the Apulia region in southern Italy (19). We estimated the introduction of the ST53 clade into Costa Rica to date back to 1943 CE (CI, 1898 CE to 2012 CE), when it came into contact with *X. fastidiosa* subsp. *fastidiosa*. The ST53 clade shared an exclusive locus encoding a putative ATP-binding protein with *X. fastidiosa* subsp. *fastidiosa* (18). Recombination analyses confirmed that the clade recombined with *X. fastidiosa* subsp. *fastidiosa* strains and acquired two genes, *cyo1* and *ilvD*. Fifty years later, *X. fastidiosa* subsp. *pauca* was subsequently introduced into Europe ca. 2008 CE (CI, 1952 CE to 2015 CE).

Summary. In conclusion, the present data set provides insight into the molecular epidemiology of this important plant pathogen. To date, only a limited number of genomic studies regarding *X. fastidiosa* have been published, and more recently they have been focused on the introduction of *X. fastidiosa* subsp. *pauca* into Italy from Central America (9, 18, 19). A molecular substitution rate for *X. fastidiosa* was computed while refining the evolutionary history of the three main subspecies. Gene differences among the subspecies were identified, and recombination analyses revealed recent interactions between the different subspecies. The study highlighted the genomic variability and plasticity of *X. fastidiosa* and corroborates previous work suggesting that IHR may be associated with host adaptation. The study revealed that the *X. fastidiosa*

subsp. *morus* genome does not appear to be the result of large-scale genome-wide recombination between *X. fastidiosa* subsp. *fastidiosa* and *multiplex*, although more data would assist in addressing this question in detail. The emergence of new genotypes as a product of recombination appears possible due to the high adaptive potential and genomic plasticity of this plant pathogen, which continues to expand its geographic range and to be associated with emerging diseases of significant economic impact.

MATERIALS AND METHODS

Strains, whole-genome sequencing, and SNP calling. A total of 72 genomes were included in the study. Genomes of 36 strains were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/genome/genomes/173) in FASTA format. An equal number were added to the study; these strains were selected to increase the phylogenetic representation of poorly characterized clades. Eight were sequenced on Illumina HiSeq 2000, 12 on the Illumina MiSeq platform, 13 using PacBio, and 3 on 454 GS Titanium. Samples were sequenced at the University of California, Berkeley, Vincent J. Coates Genomics Sequencing Laboratory (California Institute for Quantitative Biosciences; QB3) and the Center for Genomic Sciences, Allegheny Singer Research Institute, Pittsburgh, PA. All raw reads and information regarding each strain have been submitted to the NCBI (BioProject accession no. PRJNA488805). To obtain the number of single-nucleotide polymorphisms (SNPs) for the whole data set and for *X. fastidiosa* subsp. *fastidiosa*, reads were aligned to *X. fastidiosa* subsp. *fastidiosa* Temecula1 (ASM724v1). The number of SNPs for *X. fastidiosa* subsp. *multiplex* isolates was obtained by aligning reads to *X. fastidiosa* subsp. *multiplex* M12 (SAMN02598402), and reads for *X. fastidiosa* subsp. *pauca* isolates were mapped to *X. fastidiosa* subsp. *pauca* 9a5c (SAMN02603773) (Table 1). All reference genomes can be found on the EnsemblBacteria server. Alignments were performed using the long-read or short-read alignment component of Burrows-Wheeler Aligner (BWA) 0.7.15 aln (47), using a quality threshold of 15 as described by Rhodes et al. (48). PacBio reads were aligned using the BWA mem algorithm specific for PacBio. FastQs were converted to SAM format using BWA, converted to BAM files, and then sorted and indexed with SAMTOOLS version 1.3.1 (49). Duplicate reads were marked with PICARD TOOLS (v.2.4.1). The resulting BAM files were recalibrated around insertions or deletions (indels) using the GATK RealignerTargetCreator and IndelRealigner (50). Detection of single-nucleotide polymorphisms and indels was called using GATK UnifiedGenotyper version 3.6 in haploid mode (51, 52). SNPs and indels were filtered to call only high-confidence variants, using a minimum threshold of 80% of reads. Resulting variants were mapped to genes using VCF-annotator (Broad Institute, Cambridge, MA) on the latest release of *X. fastidiosa* Temecula1 (ASM724v1.36). Mapped reads for each strain are given in Table S1 in the supplemental material.

Whole-genome assembly, recombination, and pangenome analyses. Illumina sequences for each strain were assembled using SPAdes 3.6.0 (53) using the careful parameter. PacBio reads were assembled using canu.v.1.3 (54). Quality check was performed using Quast.4.6.3 (55). There were small differences in total contig length among strains (Table S1). Each contig was then ordered using progressive Mauve (56) and annotated using Prokka (57). The pangenomes of the data set and each of the main three *X. fastidiosa* subspecies were built using Roary (58) to identify the core and accessory genes present in these populations. To investigate recombination events, fastGEAR software was used to identify recombination events between lineages (ancestral) affecting all strains within the lineage and recombination within lineages (recent) (59). The software is designed to detect possible origins of recombination from other lineages as well as unknown origins. The number of substitutions introduced by recombination relative to point mutations (r/m) was estimated in ClonalFrameML. The method is known to undercount recombination events that introduce a small number of polymorphisms (60). A comparison between the two approaches to estimate recombination events, fastGEAR and ClonalFrameML, was performed to evaluate agreement between the two methods. Phage were detected using PHAST (61). Harvestv1.1.2 (62) was used to reconstruct the phylogeny of the newly sequenced genomes using the *X. fastidiosa* reference strain Temecula1.

Phylogeny, population assignment, and molecular dating. Whole-genome SNP files were converted to Nexus and Phylip format. RAxML, executing 1,000 bootstrap inferences and a generalized time-reversible substitution matrix (63), was used to generate maximum-likelihood trees over 1,000 bootstrap replicates, which were visualized in FigTree version 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). Population structure was assessed using hierarchical Bayesian analysis of population structure (BAPS) (64), and each strain was assigned to a particular genetic cluster.

The nonrecombining core genome of strains with known dates of isolation ($n = 43$), spanning 33 years of evolution (1983 to 2016), was used to investigate the presence of a temporal signal in the data set. For this purpose, the linear regression between sample age and root-to-tip distances at every internal node of the tree was computed using a custom R script sourcing the distRoot function of the adephylo R package (65). Internal nodes, at which both the linear regression was statistically significant and the slope was positive, were assumed to contain detectable evolutionary change and were considered thorough evolutionary scales for tip-dating to be performed.

Tip-dating inferences were performed using BEAST v1.8.4 (66) within clades at which a sufficient temporal signal was detected. The nucleotide substitution rate was simulated with a general time-reversible (GTR) substitution model of evolution. Rate variation among sites was modeled with a discrete gamma distribution with four rate categories. An uncorrelated log-normal relaxed clock was assumed to

account for rate variation among lineages. To minimize prior assumptions about demographic history, an extended Bayesian skyline plot (EBSP) approach was adopted to integrate data over different coalescent histories. The tree was calibrated using tip-dates only. To do so, flat priors (i.e., uniform distributions) for the substitution rate (1×10^{-12} to 1×10^{-2} substitutions/site/year), as well as for the age of any internal node in the tree, were applied. Five independent chains in which samples were drawn every 8,000 Markov chain Monte Carlo (MCMC) steps from a total of 80,000,000 steps were run, after a discarded burn-in of 8,000,000 steps. Convergence to the stationary distribution and sufficient sampling and mixing were checked by inspection of posterior samples (effective sample size, >200). Parameter estimation was based on the samples combined from the different chains. The best-supported tree was estimated from the combined samples by using the maximum clade credibility method implemented in TreeAnnotator.

In a second step, the rate of substitution was estimated using tip-dates to calibrate the whole tree containing the 72 genomes. This procedure (called "rate-dating") was performed with BEAST by incorporating the posterior of the tip-dating inference as prior for the rate of substitution, with all other parameters being unchanged from the previous step. This allowed inference of divergence times of the different subspecies. This analysis also allowed testing of the rate heterogeneity among lineages. This was done first by looking at the posterior for the standard deviation of the uncorrelated log-normal relaxed clock (the parameter ucld.stdev). If this parameter is small (close to 0), there is little variation in rates among branches, whereas values greater than 1 are indicative of substantial rate heterogeneity among lineages. In a second step, rate values for each single branch of the whole tree were extracted using a custom python script and used to compare the distributions of substitution rates among *X. fastidiosa* subspecies.

Population genetics and detection of regions under positive selection. General statistics were generated for each of the three main *X. fastidiosa* subspecies (*fastidiosa*, *multiplex,* and *pauca*): nucleotide diversity (π), Waterson's estimator (θ), and Tajima's D, using VariScan v.2.0 (67) on the core genome alignment and for each subspecies (Table 1). Nucleotide diversity (π) measures the degree of polymorphism in a population (68). The Waterson estimator estimates population mutation rate and decreases when the sample size or recombination rate increases (69). Gene clusters were generated by Roary using gene annotation from Prokka (57). Each gene cluster was then aligned using mafft v.7.305b (70), and a phylogenetic tree was generated using RAxML. The *dN*/*dS* (ω) ratio also was estimated using *de novo* assemblies and Codeml from the PAML4.1 package (71). For each gene, Codeml (runmode 0, model 0) was used to estimate *dN*/*dS,* assuming constant *dN*/*dS* ratios over all branches of the phylogenetic tree. Gene ontology (GO) terms were investigated and summarized using Blast2GO (72) to identify gene pathways which are part of the accessory genome and which were involved in either selection or recombination and might play a role in host adaptation.

Data availability. All raw reads and information regarding each strain have been submitted to the NCBI (BioProject accession no. PRJNA488805).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02972-18.

SUPPLEMENTAL FILE 1, PDF file, 2.2 MB.

ACKNOWLEDGMENTS

We acknowledge and are thankful to the following colleagues for providing or allowing us to sequence the bacterial strains used in this study: Lisa Morano, Harald Scherm, C. J. Chang, Don Hopkins, Maria Saponari, Annalisa Giampetruzzi, and Rodrigo Krugner.

Helvecio D. Coletta-Filho received CNPq research fellowships (proc. no. 313676/2017-8). This research was funded by awards from the California Department of Food and Agriculture, Horizon 2020 (XF-ACTORS project and European Union's Horizon 2020 Research and Innovation Program under Marie Skłodowska-Curie grant agreement no. 707013). Adrien Rieux thanks the European Union (European Regional Development Fund, ERDF contract no. GURDT I2016-1731-0006632), the Conseil Régional de La Réunion, the French Agropolis Foundation (Labex Agro–Montpellier, E-SPACE project no. 1504-004), and the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) for their financial support. The work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, which is supported by instrumentation grant NIH S10 OD018174.

REFERENCES

1. Almeida RPP, Nunney L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? Plant Dis 99:1457–1467. https://doi.org/10.1094/PDIS-02-15-0159-FE.
2. Sicard A, Zeilinger AR, Vanhove M, Schartel TE, Beal DJ, Daugherty MP, Almeida R. 2018. *Xylella fastidiosa*: insights into an emerging plant pathogen. Annu Rev Phytopathol 56:181–202. https://doi.org/10.1146/annurev-phyto-080417-045849.
3. Scally M, Schuenzel EL, Stouthamer R, Nunney L. 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity.

Appl Environ Microbiol 71:8491–8499. https://doi.org/10.1128/AEM.71.12.8491-8499.2005.
4. Nunney L, Ortiz B, Russell SA, Sánchez RR, Stouthamer R. 2014. The complex biogeography of the plant pathogen *Xylella fastidiosa*: genetic evidence of introductions and subspecific introgression in Central America. PLoS One 9:e112463. https://doi.org/10.1371/journal.pone.0112463.
5. Almeida RPP, Purcell AH. 2003. Biological traits of *Xylella fastidiosa* strains from grapes and almonds. Appl Environ Microbiol 69:7447–7452. https://doi.org/10.1128/AEM.69.12.7447-7452.2003.
6. Nunney L, Yuan X, Bromley RE, Stouthamer R. 2012. Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. Appl Environ Microbiol 78:4702–4714. https://doi.org/10.1128/AEM.01126-12.
7. Coletta-Filho HD, Francisco CS, Lopes JRS, Muller C, Almeida R. 2017. Homologous recombination and *Xylella fastidiosa* host–pathogen associations in South America. Phytopathology 107:305–312. https://doi.org/10.1094/PHYTO-09-16-0321-R.
8. Nunney L, Yuan X, Bromley R, Hartung J, Montero-Astúa M, Moreira L, Ortiz B, Stouthamer R. 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. https://doi.org/10.1371/journal.pone.0015488.
9. Loconsole G, Saponari M, Boscia D, D'Attoma G, Morelli M, Martelli GP, Almeida R. 2016. Intercepted isolates of *Xylella fastidiosa* in Europe reveal novel genetic diversity. Eur J Plant Pathol 146:85–94. https://doi.org/10.1007/s10658-016-0894-x.
10. Saponari M, Boscia D, Altamura G, Loconsole G, Zicca S, D'Attoma G, Morelli M, Palmisano F, Saponari A, Tavano D. 2017. Isolation and pathogenicity of *Xylella fastidiosa* associated to the olive quick decline syndrome in southern Italy. Sci Rep 7:17723. https://doi.org/10.1038/s41598-017-17957-z.
11. Su C-C, Deng W-L, Jan F-J, Chang C-J, Huang H, Chen J. 2014. Draft genome sequence of *Xylella fastidiosa* pear leaf scorch strain in Taiwan. Genome Announc 2:e00166-14.
12. Nunney L, Schuenzel EL, Scally M, Bromley RE, Stouthamer R. 2014. Large-scale intersubspecific recombination in the plant-pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. Appl Environ Microbiol 80:3025–3033. https://doi.org/10.1128/AEM.04112-13.
13. Yuan X, Morano L, Bromley R, Spring-Pearson S, Stouthamer R, Nunney L. 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. Phytopathology 100:601–611. https://doi.org/10.1094/PHYTO-100-6-0601.
14. Simpson AJG, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LMC, Araya JE, Baia GS, Baptista CS, Barros MH, Bonaccorsi ED, Bordin S, Bové JM, Briones MRS, Bueno MRP, Camargo AA, Camargo LEA, Carraro DM, Carrer H, Colauto NB, Colombo C, Costa FF, Costa MCR, Costa-Neto CM, Coutinho LL, Cristofani M, Dias-Neto E, Docena C, El-Dorry H, Facincani AP, Ferreira AJS, Ferreira VCA, Ferro JA, Fraga JS, França SC, Franco MC, Frohme M, Furlan LR, Garnier M, Goldman GH, Goldman MHS, Gomes SL, Gruber A, Ho PL, Hoheisel JD, Junqueira ML, Kemper EL, Kitajima JP, Krieger JE, Kuramae EE, Laigret F, Lambais MR, Leite LCC, Lemos EGM, Lemos MVF, Lopes SA, Lopes CR, Machado JA, Machado MA, Madeira AMBN, Madeira HMF, Marino CL, Marques MV, Martins EAL, Martins EMF, Matsukuma AY, Menck CFM, Miracca EC, Miyaki CY, Monteiro-Vitorello CB, Moon DH, Nagai MA, Nascimento ALTO, Netto LES, Nhani A, Nobrega FG, Nunes LR, Oliveira MA, de Oliveira MC, de Oliveira RC, Palmieri DA, Paris A, Peixoto BR, Pereira GAG, Pereira HA, Pesquero JB, Quaggio RB, Roberto PG, Rodrigues V, de M Rosa AJ, de Rosa VE, de Sá RG, Santelli RV, Sawasaki HE, da Silva ACR, da Silva AM, da Silva FR, Silva WA, da Silveira JF, Silvestri MLZ, Siqueira WJ, de Souza AA, de Souza AP, Terenzi MF, Truffi D, Tsai SM, Tsuhako MH, Vallada H, Van Sluys MA, Verjovski-Almeida S, Vettore AL, Zago MA, Zatz M, Meidanis J, Setubal JC. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature 406:151. https://doi.org/10.1038/35018003.
15. Bhattacharyya A, Stilwagen S, Ivanova N, D'Souza M, Bernal A, Lykidis A, Kapatral V, Anderson I, Larsen N, Los T, Reznik G, Selkov E, Walunas TL, Feil H, Feil WS, Purcell A, Lassez J-L, Hawkins TL, Haselkorn R, Overbeek R, Predki PF, Kyrpides NC. 2002. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. Proc Natl Acad Sci 99:12403–12408. https://doi.org/10.1073/pnas.132393999.
16. Van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB, Miyaki CY, Furlan LR, Camargo LEA, da Silva ACR, Moon DH, Takita MA, Lemos EGM, Machado MA, Ferro MIT, da Silva FR, Goldman MHS, Goldman GH, Lemos MVF, El-Dorry H, Tsai SM, Carrer H, Carraro DM, de Oliveira RC, Nunes LR, Siqueira WJ, Coutinho LL, Kimura ET, Ferro ES, Harakava R, Kuramae EE, Marino CL, Giglioti E, Abreu IL, Alves LMC, do Amaral AM, Baia GS, Blanco SR, Brito MS, Cannavan FS, Celestino AV, da Cunha AF, Fenille RC, Ferro JA, Formighieri EF, Kishi LT, Leoni SG, Oliveira AR, Rosa VE, Sassaki FT, Sena JAD, de Souza AA, Truffi D, Tsukumo F, Yanai GM, Zaros LG, Civerolo EL, Simpson AJG, Almeida NF, Setubal JC, Kitajima JP. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J Bacteriol 185:1018–1026. https://doi.org/10.1128/JB.185.3.1018-1026.2003.
17. Doddapaneni H, Yao J, Lin H, Walker MA, Civerolo EL. 2006. Analysis of the genome-wide variations among multiple strains of the plant pathogenic bacterium *Xylella fastidiosa*. BMC Genomics 7:225. https://doi.org/10.1186/1471-2164-7-225.
18. Giampetruzzi A, Saponari M, Loconsole G, Boscia D, Savino VN, Almeida R, Zicca S, Landa B, Chacon Diaz C, Saldarelli P. 2017. Genome-wide analysis provides evidence on the genetic relatedness of the emergent *Xylella fastidiosa* genotype in Italy to isolates from Central America. Phytopathology 107:816–827. https://doi.org/10.1094/PHYTO-12-16-0420-R.
19. Marcelletti S, Scortichini M. 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. https://doi.org/10.1099/mic.0.000388.
20. Chatterjee S, Almeida RPP, Lindow S. 2008. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. Annu Rev Phytopathol 46:243–271. https://doi.org/10.1146/annurev.phyto.45.062806.094342.
21. Nunney L, Vickerman DB, Bromley RE, Russell SA, Hartman JR, Morano LD, Stouthamer R. 2013. Recent evolutionary radiation and host plant specialization in the *Xylella fastidiosa* subspecies native to the United States. Appl Environ Microbiol 79:2189–2200. https://doi.org/10.1128/AEM.03208-12.
22. Schuenzel EL, Scally M, Stouthamer R, Nunney L. 2005. A multigene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. Appl Environ Microbiol 71:3832–3839. https://doi.org/10.1128/AEM.71.7.3832-3839.2005.
23. Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.
24. Rieux A, Balloux F. 2016. Inferences from tip-calibrated phylogenies: a review and a practical guide. Mol Ecol 25:1911–1924. https://doi.org/10.1111/mec.13586.
25. Kitajima EW, Bakarcic M, Fernandez-Valiela MV. 1975. Association of rickettsialike bacteria with plum leaf scald disease. Phytopathology 65:476–479. https://doi.org/10.1094/Phyto-65-476.
26. Almeida RPP, Nascimento FE, Chau J, Prado SS, Tsai C-W, Lopes SA, Lopes J. 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. Appl Environ Microbiol 74:3690–3701. https://doi.org/10.1128/AEM.02388-07.
27. Francisco CS, Ceresini PC, Almeida RPP, Coletta-Filho HD. 2017. Spatial genetic structure of coffee-associated *Xylella fastidiosa* populations indicates that cross infection does not occur with sympatric citrus orchards. Phytopathology 107:395–402. https://doi.org/10.1094/PHYTO-08-16-0300-R.
28. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484:186–194. https://doi.org/10.1038/nature10947.
29. Giraud T, Koskella B, Laine A. 2017. Introduction: microbial local adaptation: insights from natural populations, genomics and experimental evolution. Mol Ecol 26:1703–1710. https://doi.org/10.1111/mec.14091.
30. Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J 3:199. https://doi.org/10.1038/ismej.2008.93.
31. Kung SH, Almeida R. 2011. Natural competence and recombination in the plant pathogen *Xylella fastidiosa*. Appl Environ Microbiol 77:5278–5284. https://doi.org/10.1128/AEM.00730-11.
32. Voegel TM, Doddapaneni H, Cheng DW, Lin H, Stenger DC, Kirkpatrick BC, Roper MC. 2013. Identification of a response regulator involved in surface attachment, cell-cell aggregation, exopolysaccharide production and virulence in the plant pathogen *Xylella fastidiosa*. Mol Plant Pathol 14:256–264. https://doi.org/10.1111/mpp.12004.

33. Lindow SE, Feil H. 2005. Effects of fimbrial (FimA, FimF) and afimbrial (XadA, HxfB) adhesins on the adhesion of *Xylella fastidiosa* to surfaces, p 173–176. *In* Proceedings of the 2005 Pierce's Disease Research Symposium. California Department of Food and Agriculture, Sacramento, CA.
34. Cruz LF, Parker JK, Cobine PA, De La Fuente L. 2014. Calcium-enhanced twitching motility in *Xylella fastidiosa* is linked to a single PilY1 homolog. Appl Environ Microbiol 80:7176–7185. https://doi.org/10.1128/AEM.02153-14.
35. Cursino L, Athinuwat D, Patel KR, Galvani CD, Zaini PA, Li Y, De La Fuente L, Hoch HC, Burr TJ, Mowery P. 2015. Characterization of the *Xylella fastidiosa* PD1671 gene encoding degenerate c-di-GMP GGDEF/EAL domains, and its role in the development of Pierce's disease. PLoS One 10:e0121851. https://doi.org/10.1371/journal.pone.0121851.
36. Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. 2016. Genome-scale rates of evolutionary change in bacteria. Microb Genom 2:e000094. https://doi.org/10.1099/mgen.0.000094.
37. De Las Peñas A, Connolly L, Gross CA. 1997. The σE-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two negative regulators of σE. Mol Microbiol 24:373–385. https://doi.org/10.1046/j.1365-2958.1997.3611718.x.
38. Hedge J, Wilson DJ. 2016. Practical approaches for detecting selection in microbial genomes. PLoS Comput Biol 12:e1004739. https://doi.org/10.1371/journal.pcbi.1004739.
39. Hansson M, Hederstedt L. 1994. *Bacillus subtilis* HemY is a peripheral membrane protein essential for protoheme IX synthesis which can oxidize coproporphyrinogen III and protoporphyrinogen IX. J Bacteriol 176:5962–5970. https://doi.org/10.1128/jb.176.19.5962-5970.1994.
40. Gardner PR, Gardner AM, Martin LA, Dou Y, Li T, Olson JS, Zhu H, Riggs AF. 2000. Nitric oxide dioxygenase activity and function of flavohemoglobins: sensitivity to nitric oxide and carbon monoxide inhibition. J Biol Chem 275:31581–31587. https://doi.org/10.1074/jbc.M004141200.
41. de Souza AA, Takita MA, Coletta-Filho HD, Caldana C, Yanai GM, Muto NH, de Oliveira RC, Nunes LR, Machado MA. 2004. Gene expression profile of the plant pathogen *Xylella fastidiosa* during biofilm formation in vitro. FEMS Microbiol Lett 237:341–353. https://doi.org/10.1016/j.femsle.2004.06.055.
42. Pierce NB. 1892. The California vine disease: a preliminary report of investigations. USDA Division of Vegetable Pathology bulletin no. 2. Government Printing Office, Washington, DC.
43. Hewitt WB. 1958. The probable home of Pierce's disease virus. Plant Dis Rep 42:211–215.
44. Chang CJ, Garnier M, Zreik L, Rossetti V, Bové JM. 1993. Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. Curr Microbiol 27:137–142. https://doi.org/10.1007/BF01576010.
45. French WJ, Kitajima EW. 1978. Occurrence of plum leaf scald in Brazil and Paraguay. Plant Dis Rep 62:1035–1038.
46. Rossetti V, Garnier M, Bové JM, Beretta M-J-G, Teixeira ARR, Quaggio JA, de Negri JD. 1990. Présence de bactéries dans le xylème d'orangers atteints de chlorose variégée, une nouvelle maladie des agrumes au Brésil. C R Acad Sci III 310:345–349.
47. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
48. Rhodes J, Beale MA, Fisher MC. 2014. Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of Cryptococcus neoformans using Illumina HiSeq. PLoS One 9:e113501. https://doi.org/10.1371/journal.pone.0113501.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
50. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.
51. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. https://doi.org/10.1038/ng.806.
52. Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformat 43:11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.
53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
54. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
55. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
56. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
57. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
58. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
59. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. 2017. Efficient inference of recent and ancestral recombination within bacterial populations. Mol Biol Evol 34:1167–1182. https://doi.org/10.1093/molbev/msx066.
60. Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol 11:e1004041. https://doi.org/10.1371/journal.pcbi.1004041.
61. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.
62. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:524. https://doi.org/10.1186/s13059-014-0524-x.
63. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. https://doi.org/10.1093/bioinformatics/btl446.
64. Corander J, Waldmann P, Marttinen P, Sillanpää MJ. 2004. BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20:2363–2369. https://doi.org/10.1093/bioinformatics/bth250.
65. Jombart T, Dray S. 2010. adephylo: exploratory analyses for the phylogenetic comparative method. Bioinformatics 26:1–21. https://doi.org/10.1093/bioinformatics/btq292.
66. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. https://doi.org/10.1186/1471-2148-7-214.
67. Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. 2005. VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics 21:2791–2793. https://doi.org/10.1093/bioinformatics/bti403.
68. Nei M, Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269–5273. https://doi.org/10.1073/pnas.76.10.5269.
69. Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276. https://doi.org/10.1016/0040-5809(75)90020-9.
70. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
71. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088.
72. Conesa A, Götz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:619832. https://doi.org/10.1155/2008/619832.