

Homologous Recombination and *Xylella fastidiosa* Host–Pathogen Associations in South America

Helvécio D. Coletta-Filho, Carolina S. Francisco, João R. S. Lopes, Christiane Muller, and Rodrigo P. P. Almeida

First and second authors: IAC, Centro de Citricultura Sylvio Moreira, Cordeirópolis, São Paulo, Brazil; second author: UNESP, Universidade Estadual Paulista, Campus de Jaboticabal, Graduate Program in Genetics and Plant Breeding, São Paulo, Brazil; third and fourth authors: Departamento de Entomologia, ESALQ, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; and fifth author: Department of Environmental Science, Policy, and Management, University of California, Berkeley.

Current address of C. S. Francisco: Plant Pathology, Institute of Integrative Biology, ETH Zürich, CH-8092 Zürich, Switzerland.

Accepted for publication 3 November 2016.

ABSTRACT

Homologous recombination affects the evolution of bacteria such as *Xylella fastidiosa*, a naturally competent plant pathogen that requires insect vectors for dispersal. This bacterial species is taxonomically divided into subspecies, with phylogenetic clusters within subspecies that are host specific. One subspecies, *pauca*, is primarily limited to South America, with the exception of recently reported strains in Europe and Costa Rica. Despite the economic importance of *X. fastidiosa* subsp. *pauca* in South America, little is known about its genetic diversity. Multilocus sequence typing (MLST) has previously identified six sequence types (ST) among plant samples collected in Brazil (both subsp. *pauca* and *multiplex*). Here, we report on a survey of *X. fastidiosa* genetic diversity (MLST based) performed in six regions in Brazil and two in Argentina, by sampling five different plant species. In addition to the six previously reported ST, seven new subsp. *pauca* and two new subsp. *multiplex* ST were identified. The

presence of subsp. *multiplex* in South America is considered to be the consequence of a single introduction from its native range in North America more than 80 years ago. Different phylogenetic approaches clustered the South American ST into four groups, with strains infecting citrus (subsp. *pauca*); coffee and olive (subsp. *pauca*); coffee, hibiscus, and plum (subsp. *pauca*); and plum (subsp. *multiplex*). In areas where these different genetic clusters occurred sympatrically, we found evidence of homologous recombination in the form of bidirectional allelic exchange between subsp. *pauca* and *multiplex*. In fact, the only strain of subsp. *pauca* isolated from a plum host had an allele that originated from subsp. *multiplex*. These signatures of bidirectional homologous recombination between endemic and introduced ST indicate that gene flow occurs in short evolutionary time frames in *X. fastidiosa*, despite the ecological isolation (i.e., host plant species) of genotypes.

Understanding the drivers associated with emerging diseases is of paramount importance in the development of risk reduction strategies. Long-distance dispersal mediated by anthropogenic activities, land use changes, abiotic or biotic modifications affecting host biology or ecology, pathogen evolution, as well as other factors such as climate alterations all have been shown to drive emergence of animal, plant, and human diseases (Daszak et al. 2000; Hacker and Carniel 2001; Hufbauer et al. 2012; Murray and Daszak 2013). Within this context, however, the emergence of diseases caused by related pathogens is usually associated with a limited number of drivers. For example, flu pandemics are linked to virus reassortments (Garten et al. 2009) and bird migrations (Verhagen et al. 2015), while land use and climate change affect the occurrence of Lyme disease in North America (Simon et al. 2014). Therefore, broadly observed patterns serve as starting points when evaluating factors associated with emerging diseases caused by related pathogens. Quantitative analyses of databases and published literature have provided important insights on factors, as well as pathogen groups and potential geographical hotspots, associated with emerging animal and human diseases (Murray et al. 2015). A very limited number of similar studies are available for plants (Anderson et al. 2004) but plant diseases emerge due to a variety of factors, from breakdown of host plant resistance via pathogen adaptation (Croll and McDonald

2012) to the introduction of new insect vectors (Almeida and Nunney 2015).

The bacterium *Xylella fastidiosa* is the etiological agent of some of the most important plant diseases that have emerged during the last few decades. Epidemics of *X. fastidiosa* have been associated with long-distance introductions, geographical expansion of an invasive vector, as well as host adaptation possibly, via homologous recombination (Almeida and Nunney 2015). Significant advances in the understanding of ecological and evolutionary processes associated with *X. fastidiosa* epidemics have been recently obtained after the advent of a public multilocus sequence typing (MLST) database, a repository of sequence data for most of the known *X. fastidiosa* diversity (Scally et al. 2005; Schuenzel et al. 2005; Yuan et al. 2010). Currently, there are five *X. fastidiosa* subspecies, largely based on DNA-DNA hybridization and MLST data (Almeida and Nunney 2015). Recent data based on field surveys and experimental research have shown that, within subspecies, there is host–pathogen specificity, especially when symptoms are concerned (Almeida et al. 2008; Harris and Balci 2015; Nunney et al. 2013). Within *X. fastidiosa* subsp. *pauca*, specifically, cross inoculation of citrus- and coffee-infecting isolates, which are genetically distinct, does not lead to disease even under greenhouse conditions (Almeida et al. 2008).

X. fastidiosa is naturally competent, with high homologous recombination rates being achieved under laboratory conditions (Kung and Almeida 2011) as well as in the field (Nunney et al. 2013). Because recombination affects phylogenetic analyses, a general understanding of *X. fastidiosa* genetic diversity has not been available until recently. The case of the apparently recombinogenic origin of a genotype infecting mulberry (*X. fastidiosa* subsp. *morus*) (Nunney et al. 2014b), and similar but less dramatic gene flow

Corresponding author: H. D. Coletta-Filho;
E-mail address: helvecio@centrodecitricultura.br

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

molecular signatures associated with emerging diseases in South America by *X. fastidiosa* subsp. *pauca* (Almeida et al. 2008, Nunney et al. 2012), illustrate the impact of recombination on inference of phylogenetic relationships within *X. fastidiosa*. In addition, some recently published data emphasized that there is additional diversity within this species that has not been described, including within subsp. *pauca* (Loconsole et al. 2016).

X. fastidiosa subsp. *pauca*, limited until recently to South America, is the etiological agent of economically important diseases, including citrus variegated chlorosis and coffee leaf scorch. These diseases emerged in the last 30 years; MLST data indicate the presence of homologous recombination events with another subspecies of *X. fastidiosa* (subsp. *multiplex*), which was earlier introduced into South America and causes disease in plum in Argentina, Paraguay, and southern Brazil (French and Kitajima 1978). A proposal has been made suggesting that loci or alleles donated by this introduced genotype to endemic *X. fastidiosa* subsp. *pauca* led to emergence of the citrus and coffee diseases mentioned above, which is possible yet not conclusively demonstrated based on available data (Nunney et al. 2012). In addition to being the etiological agent of diseases in South America, *X. fastidiosa* subspecies *pauca* has now been reported from Costa Rica (Nunney et al. 2014a) and Europe (Loconsole et al. 2016; Saponari et al. 2013). Despite the growing relevance due to its increasing geographical distribution, as well as crops to which it causes disease, the genetic diversity of subsp. *pauca* remains poorly characterized. The goal of this study was to generate novel information on the genetic diversity of this subspecies using MLST data from isolates collected from different regions and hosts in South America. Strains were isolated from previously unknown hosts as well as from those for which data were already available. New *X. fastidiosa* subsp. *pauca* sequence types (ST) were found in citrus, coffee, plum, and other hosts, while homologous recombination was found to be important in shaping the diversity of subsp. *pauca* in short evolutionary time frames.

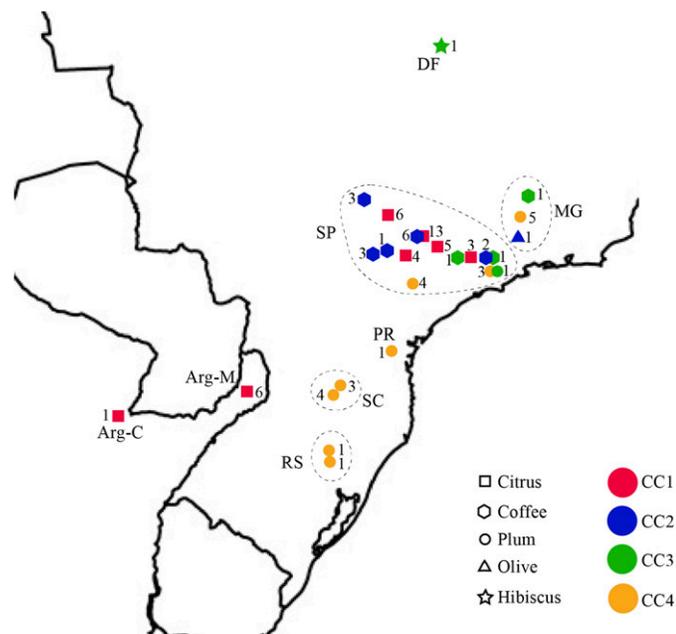


Fig. 1. Geographical distribution of sample collection sites. The map includes the host plant and number of samples collected in each site, as well as the clonal complex assigned to samples. Brazilian States are circled or labeled: RS = Rio Grande do Sul, SC = Santa Catarina, PR = Paraná, SP = São Paulo, MG = Minas Gerais, and DF = Distrito Federal; regions in Argentina are labeled as Arg-M = Misiones, and Arg-C = Corrientes. The location of symbols is approximate for clarity; in addition, a few sites required two overlapping symbols, one of which was slightly moved.

***X. fastidiosa* sampling and culturing.** In all, 102 samples from infected and symptomatic host plants were collected from different sites in Brazil (São Paulo State: citrus, coffee, and plum; Minas Gerais State: coffee, plum, and olive; Paraná, Santa Catarina, and Rio Grande do Sul: plum; and Brasília, Distrito Federal: hibiscus) and Argentina (citrus) (Supplementary Table S1; Fig. 1). Infected plant tissue from citrus (*Citrus sinensis*), coffee (*Coffea arabica*), plum (*Prunus domestica*), hibiscus (*Hibiscus rosa-sinensis*), and olive (*Olea europaea*) was transported to the laboratory and isolates were cultured and triple-cloned on buffered charcoal yeast extract (BCYE) agar medium (Wells et al. 1981) so that only single clones were genotyped. Identity of the clones (i.e., as *X. fastidiosa*) was determined with the specific RST31-33 primer set (Minsavage et al. 1994). All isolates were frozen at -80°C for later use.

DNA extraction, polymerase chain reaction assay, and sequencing. Total DNA was extracted from isolates using the Wizard DNA purification kit (Promega Corporation). The polymerase chain reaction (PCR) amplification of seven MLST loci (*leuA*, *petC*, *malF*, *cysG*, *holC*, *nuoL*, and *gltT*) was done following protocols described by Yuan et al. (2010). After confirmation of amplicon size in agarose gels (1%), amplicons were purified using the QIAquick PCR purification Kit (Qiagen) and directly sequenced on both strands (ABI 3730 sequencer; Life Technology). The forward and reverse sequences of each isolate were assembled by the CLC Genomics Workbench platform (Qiagen). Single-nucleotide polymorphisms were manually checked in the electropherograms.

MLST analysis. Reference sequences for each MLST locus was downloaded from the *X. fastidiosa* MLST database (<http://pubmlst.org/xfastidiosa/>) and aligned with the new data using ClustalW software in the MEGA package (Tamura et al. 2013). The MLST characterization of each isolate followed guidelines described by Nunney et al. (2013) and the *X. fastidiosa* MLST databases (<http://pubmlst.org/xfastidiosa/>). When necessary, the identification of a new allele for the MLST gene was done by using the next number of those present in the preexisting database. The same procedure was done for a new allelic profile of an MLST gene for which an ST number is assigned. The ST were grouped in clonal complexes (CC) by using the software e-BURST (<http://eburst.mlst.net/>), using as criterion the sharing of more than five of seven alleles among ST into a CC, as proposed by Yuan et al. (2010). All information has been deposited in the *X. fastidiosa* MLST database, which is curated by Len Nunney (University of California, Riverside).

Genetic distance and phylogeny. Concatenated sequences from the seven MLST loci were used to run both genetic distance (GD) and phylogenetic models for an overall comparison between all *X. fastidiosa* previously known and novel ST obtained in this work. Although the GD model may fit better to the dataset than the phylogenetic approach used, due to the presence of recombination, analyses using both models should give better confidence for the results. The GD model was performed in the Phylip 3.61 package (Felsenstein 1989) using the programs Seqboot adopting 1,000 bootstraps, Dnadist, Neighbor, and Consense. The phylogenetic reconstruction was done based on Bayesian inference using the Markov Chain Monte Carlo with 100,000 iterations and 1,000 bootstraps using MrBayes v.3.2 (Huelsenbeck and Ronquist 2001). The best-fit model of DNA evolution was the generalized time-reversible, as determined by Paup v.4 beta software (Swofford 2002). Fig. Tree 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to generate trees from the output of the MrBayes and Phylip analyses (>50% consensus tree); trees were midpoint rooted. SplitsTree v.4 (Huson and Bryant 2006) was used to build a NeighborNet phylogenetic network representing the genetic relationships among ST. The genetic diversity index within each *X. fastidiosa* subspecies was estimated based on the Jukes-Cantor model with 1,000 bootstraps by using the DnaSP package. For this analysis, sequences of alleles for each one of seven MLST loci of ST

previously clustered into each subspecies (*pauca*, *morus*, *fastidiosa*, and *multiplex*) were used.

Intragenic recombination analysis. Recombination analyses focused only on the novel alleles (*holC*, *nuoL*, and *leuA* loci) with occurrence in *X. fastidiosa* subsp. *pauca* in South America. Estimates for recombination sites was done by using the ZZ statistic (Kelly 1997), which analyzed the linkage disequilibrium between polymorphic sites in relation to physical distance, as implemented in DnaSP v.5 (Librado and Rozas 2009) using 1,000 coalescent simulations. The genetic algorithm recombination detection (GARD) was used to screen multiple-sequence alignments for recombination sites, and it was run by using the DataMonkey software (<http://www.datamonkey.org/>) (Kosakovsky Pond et al. 2006). Signals of recombination were also searched for with the RDP v.4 package (Martin et al. 2015) using eight different nonparametric recombination detection tests simultaneously. Phylogenetic reconstruction for each locus (*holC*, *nuoL*, and *leuA*) was done with MrBayes using the same procedure described above for concatenated sequences. The phylogenetic trees were visualized with Fig. Tree and midpoint rooted.

Nucleotide sequence accession numbers. All of the MLST data are available from the *X. fastidiosa* MLST database. The gene sequences for the novel MLST alleles obtained in the present study are available at GenBank under the following accession numbers: *holC*, allele 22 = KU521510 and 23 = KU521511; *leuA*, allele 14 = KU521512; and *nuoL*, allele 17 = KU521513.

RESULTS

MLST profiles. Five *X. fastidiosa* subsp. *pauca* ST were previously reported from Brazil (ST11 to -14 and -16) (Nunney et al. 2012); only one of those (ST12) was not found in our survey. Therefore, all but one of the known subsp. *pauca* ST previously reported in Brazil were detected. Seven novel ST from subsp. *pauca* and two from subsp. *multiplex* (ST63 and -67) were identified, as well as one previously reported genotype from subsp. *multiplex* (ST26) (Table 1). New alleles were identified for loci *leuA* (allele 14), *holC* (alleles 22 and 23), and *nuoL* (allele 17). These 15 ST, which represent all currently known MLST-based diversity from South America, clustered in four CC (Table 1). All isolates obtained from citrus clustered in CC1, although the seven isolates from Argentina belonged to ST69 exclusively, which did not include any isolates from Brazil. Isolates in CC2 were obtained from coffee trees; the exception was one isolate from an olive tree that belonged to ST16. Isolates in CC4 were only obtained from plum plants and belong to subsp. *multiplex*, while one genotype from plum clustered in CC3 (ST71) and had a substantially different allelic profile (i.e., subsp. *pauca*) from those ST in CC4.

Novel allelic profiles and intragenic recombination. Novel alleles were found for loci *leuA*, *holC*, and *nuoL*. ST70, from a hibiscus plant, carried two novel alleles (*leuA* number 14 and *holC* number 22), both of which were phylogenetically related to coffee isolates (Fig. 2). For the remaining loci of ST70, one allele was found in citrus isolates, three in coffee isolates, and one in isolates from both hosts. Based on two statistical analyses (ZZ and GARD), both novel alleles show signals of intragenic recombination as observed among alleles *leuA* 14, 7, and 8 (Supplementary Table S2) and among alleles of *holC* (Supplementary Table S3). However, no recombination signal was detected when RDP4 was used. Nevertheless, these data suggest that ST70 is not an ancestral genotype of subsp. *pauca*. Instead, ST70 likely represents a more recent case of recombination. All isolates from citrus trees collected in Argentina belonged to ST69, which had two novel alleles (*holC* number 23 and *nuoL* number 17), both of which were phylogenetically related to subsp. *pauca* alleles (Fig. 2). Sequences of *holC* show two highly differenced groups in subsp. *pauca* but also include signals of intragenic recombination among alleles 10, 11, 12, 22, and 23 as identified with the ZZ statistic and GARD but not when RDP4 was used to analyze the data. In the case of *nuoL* 17, no evidence of intrasubspecific recombination events was observed

(Supplementary Table S4). It is notable that, although two novel ST were isolated from citrus trees in Brazil, no new alleles were identified, suggesting that recombination among ST within CC1 and not mutation (due to the number of substitutions involved) led to the emergence of these genotypes (Table 1).

Phylogenetic relationship among ST. Three distinct analyses were performed due to the presence of intragenic and allelic recombination in the *X. fastidiosa* MLST data; the results are similar and, therefore, summarized together. Within subsp. *pauca*, the genotypes ST53, ST73, and ST74 have not been detected in South America but are present in Costa Rica and Italy (ST53) or were intercepted in coffee plants introduced into Europe (ST73 and ST74), and form a clade associated with subsp. *pauca*. The allelic profile and phylogenetic placement of ST53, ST73, and ST74 suggest that these genotypes originated in South America, despite the fact that the clade has no reported ST in the region. Three CC found in South America belong to subsp. *pauca*, while the fourth CC belongs to subsp. *multiplex*, based on the MLST profile and distinct phylogenetic analyses (Fig. 3). ST26 is a subsp. *multiplex* type almost exclusively associated with *Prunus* spp. in North America (Nunney et al. 2013). ST63 is nested within this group but the placement of ST67 is not as clear due to the recombination events among subsp. *multiplex* and *pauca*. The other three CC identified in this study were monophyletic regardless of the method used to infer relationships among each other and other *X. fastidiosa* genotypes in South America; these CC were also previously identified by Nunney et al. (2012). All isolates from citrus trees belonged to CC1, which formed a sister clade with CC2 (primarily found in coffee plants), with the exception of one isolate from a symptomatic olive tree. The last group (CC3) included isolates from coffee, hibiscus, and plum. Less stringent parameters (≥ 4 shared alleles of the seven loci) than used here placed the ST from CC1 and CC2 into a unique group but CC3 and CC4 remained unaltered. More stringent parameters (≥ 6 shared alleles of the seven loci) resulted in several singletons (data not shown). It is possible that, once more of the *X. fastidiosa* genetic diversity is described, more

TABLE 1. Profiles of multilocus sequence type (MLST) from *Xylella fastidiosa* isolates from various hosts species in Brazil and Argentina

ST ^b	MLST allele numbers ^a							N ^c
	<i>leuA</i>	<i>petC</i>	<i>malF</i>	<i>cysG</i>	<i>holC</i>	<i>nuoL</i>	<i>gltT</i>	
CC1								
11	7	7	7	9	10	8	8	25
12	7	7	7	9	13	8	8	0
13	7	6	7	9	10	7	8	1
64*	7	7	7	9	10	7	8	2
65*	7	6	7	9	10	8	8	2
69*	7	6	7	9	23**	17**	8	7
CC2								
16	7	6	8	10	11	8	8	15
66*	7	8	8	10	11	8	8	1
CC3								
14	8	8	8	11	12	9	9	2
68*	14**	8	8	11	12	9	8	1
70*	14**	7	8	11	22**	9	8	1
71*	<u>5</u>	8	8	11	12	9	9	1
CC4								
26	5	3	3	3	6	3	5	21
63*	5	<u>6</u>	3	3	6	3	5	1
67*	5	3	<u>8</u>	3	<u>12</u>	3	5	1

^a Two asterisks (**) indicate new alleles (*leuA* 14, nearest to allele 8; *holC* 22, nearest to allele 12; *holC* 23, nearest to allele 11; and *nuoL* 17, nearest to allele 8). Numbers in bold and underlined are alleles typical of *X. fastidiosa* subsp. *multiplex*. Numbers underlined are allele transfers between subsp. *pauca* and *multiplex* of *X. fastidiosa*.

^b CC = clonal complex. One asterisk (*) indicates new sequence type (ST). Information on representative isolates belonging to different ST is presented in Supplementary Table S1.

^c Number of isolates.

stringent parameters than those used here could generate closer connections between CC and host plants.

Allelic recombination. Bidirectional allelic recombination between subsp. *pauca* and *multiplex* was observed among the ST. ST71 has one locus (*leuA* 5) that clusters phylogenetically and is characteristic of the introduced CC4 (subsp. *multiplex*) (Fig. 4). Transfer of loci from subsp. *pauca* to plum isolates in CC4 that have a subsp. *multiplex* profile was observed for ST63 (*petC* 6) and ST67 (*malF* 8 and *holC* 12). Based on this study, the transfer of alleles to the introduced subsp. *multiplex* was not associated with the detection of this CC in host plants other than plum, whereas evidence of transfer from subsp. *multiplex* to *pauca*, as indicated by ST71, was correlated with the association of that ST with plum hosts.

Subspecies level nucleotide diversity. Nucleotide diversity within each *X. fastidiosa* subspecies (except subsp. *sandyi*, because it has only one ST) was compared by using available MLST data (Table 2). It should be noted that this comparison includes alleles originating from intragenic and intergenic recombination. Regardless, subsp. *pauca* appears to have more diversity than the other three subspecies analyzed, despite the fact that recombination has been previously reported to affect allelic diversity of all *X. fastidiosa* subspecies.

DISCUSSION

The use of seven housekeeping genes for genotyping *X. fastidiosa* strains (i.e., MLST) has led to a robust taxonomic framework for the species, in addition to insights into processes shaping diversity. Although not appropriate for the typing of closely related strains (Yuan et al. 2010), MLST has been useful at the subspecies level of phylogenetic resolution (Almeida et al. 2008; Yuan et al. 2010). The portability of the system has allowed for fast identification of epidemic strains (Loconsole et al. 2016), as well as host–pathogen

associations among sympatric ST (Harris and Balci 2015). MLST was used to identify novel genetic diversity in South America, expanding the number of ST from 6 to 15. The gathered data suggest a well-supported and monophyletic clade of *X. fastidiosa* subsp. *pauca* infecting citrus trees in South America (CC1), while strains from symptomatic coffee plants belonged to two groups (CC2 and CC3) which also had strains isolated from hibiscus, plum, and olive trees (Almeida et al. 2008; Nunney et al. 2012). The main ST infecting plum (ST26) is a typical subsp. *multiplex* genotype from North America, supporting the now-established proposal that subsp. *multiplex* was introduced into South America, probably over 80 years ago, via contaminated plant material from North America (French and Kitajima 1978; Kitajima et al. 1975; Nunes et al. 2003). The results show that, within the time frame since this introduction, homologous recombination has led to bidirectional gene flow between the endemic subsp. *pauca* and introduced subsp. *multiplex*, which currently occur sympatrically in South America, as previously suggested with a smaller data set (Nunney et al. 2012).

Although it is now evident that *X. fastidiosa* subspecies and clusters therein have high degrees of host plant species–bacterial genotype (MLST-based) specificity, taxonomic delimitation of genetic clusters that cause disease in one host but not another remains unclear. A limited set of alleles was identified in CC1, associated with citrus only, whereas ST in two other CC were obtained from coffee plants (CC2 and CC3) but those groups included genotypes from olive, hibiscus, and plum as well. This observation is not limited to subsp. *pauca*; in the case of subsp. *multiplex*, isolates obtained from trees in Washington, DC alone, there are examples of specific ST infecting sycamore or oak, while elm may be infected with two ST (Harris and Balci 2015). On the other hand, almond trees in California may be infected with two distinct subsp. *multiplex* ST and strains from subsp. *fastidiosa* (Almeida and Purcell 2003; Nunney et al. 2013). Although a

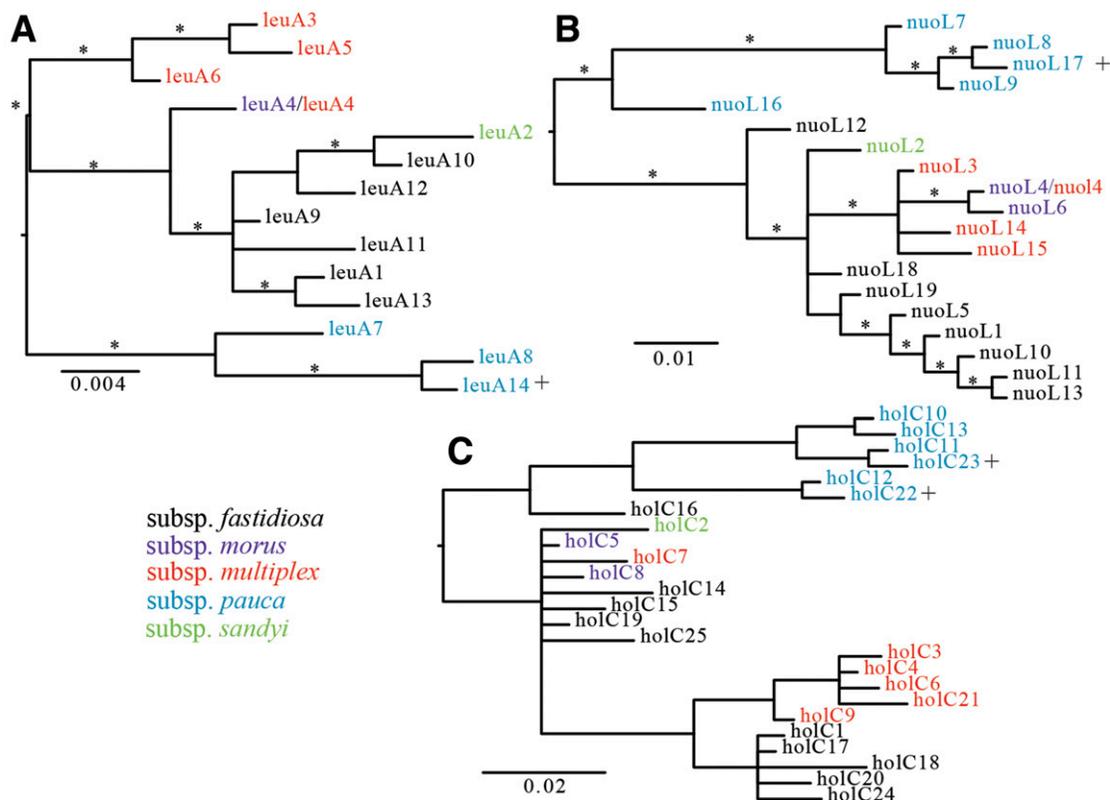


Fig. 2. Single-locus phylogenetic trees of the new *Xylella fastidiosa* subsp. *pauca* alleles for loci **A**, *leuA*; **B**, *holC*; and **C**, *nuoL* in relation to all alleles reported for *X. fastidiosa*. Asterisks represent branches with $\geq 90\%$ bootstrap support; symbol + after an allele number indicates novel subsp. *pauca* alleles. Alleles are color-coded based on *X. fastidiosa* subspecies of origin.

significant amount of *X. fastidiosa* diversity has not been experimentally tested for host plant range, the consistency of field associations indicates host specificity requirements for expression of disease symptoms. However, it is clear that MLST typing alone, despite its usefulness, is insufficient for precise inferences on the host range of

strains, especially when grouping of CC is considered, as demonstrated here in the case of ST71, which was recovered from a symptomatic plum tree but has only one allele typical of plum-infecting isolates, while all others originate from South American subsp. *pauca*.

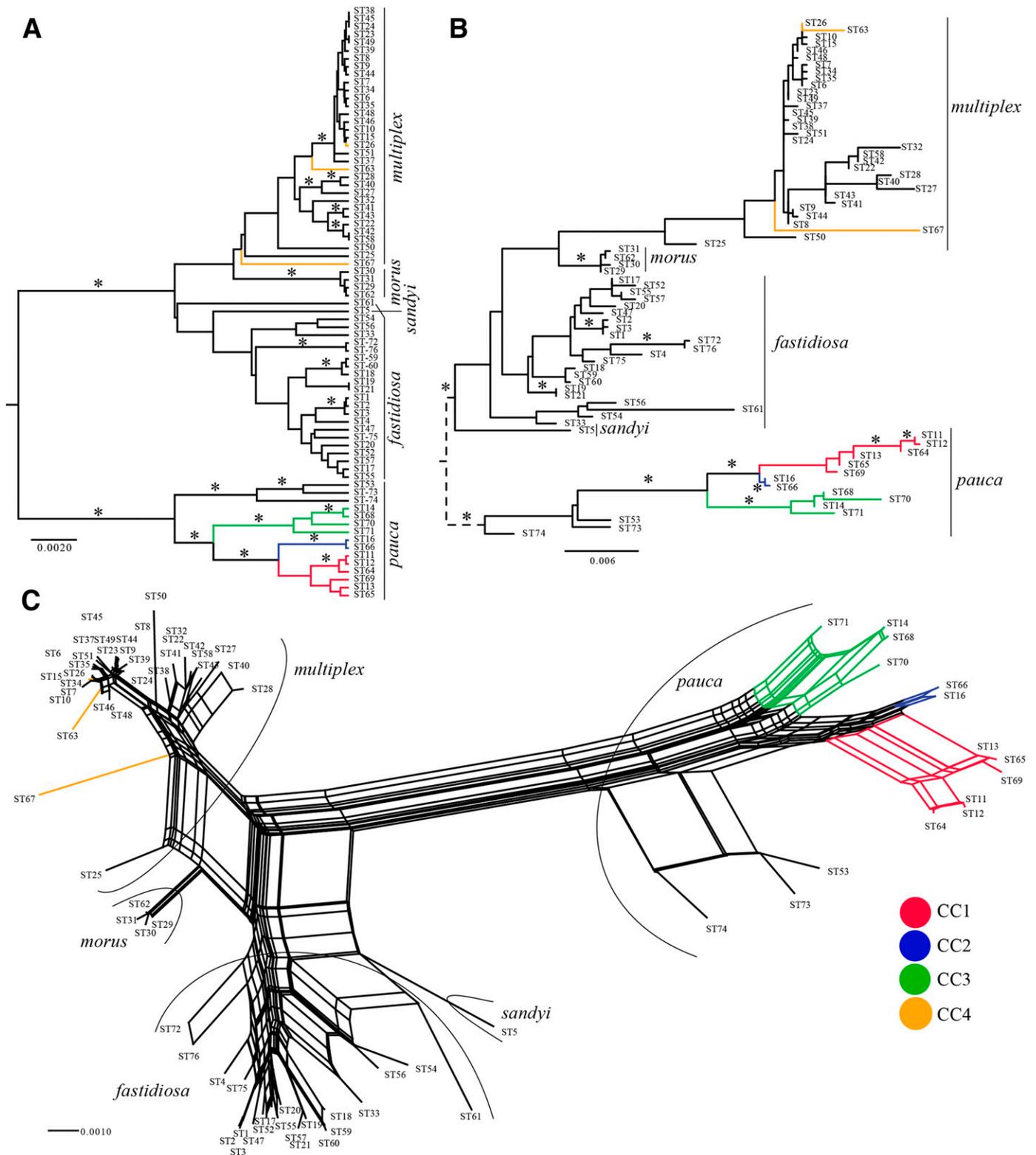


Fig. 3. Phylogenetic placement of *Xylella fastidiosa* sequence types (ST) from South America in relation to all currently known ST. **A**, Dendrogram showing the genetic distance among *X. fastidiosa* ST. **B**, Bayesian inference for the phylogeny of *X. fastidiosa*. Asterisks in **A** and **B** represent branches with $\geq 90\%$ bootstrap support. **C**, Neighbor-net reconstruction of relationships among *X. fastidiosa* ST. e-BURST was used to define clonal complexes (CC) whereas the ST in a CC share six or more alleles. The same CC identified by e-BURST were color-labeled in all trees.

The novel allelic diversity found in this survey comprised alleles that cluster with other alleles from subsp. *pauca*. Evidence of recombinogenic origins of some of these alleles suggests intragenic recombination among genotypes. However, results from different recombination detection approaches were not consistent, as previously found by Nunney et al. (2012). The frequency and role of intragenic homologous recombination in *X. fastidiosa* was not a focus of this study but should be addressed at the species level, because it may provide insights into the evolution of host–pathogen associations. On the other hand, a comparison of allelic genetic diversity among *X. fastidiosa* subspecies showed a substantial amount of variability among loci as well as between subspecies. These results suggest that more genetic diversity likely exists within subsp. *pauca* compared with other subspecies. Because ST are based on unique sets of alleles, and because subsp. *pauca* is represented by a small fraction of ST when compared with subsp. *fastidiosa* and *multiplex*, the data suggest that subsp. *pauca* is more diverse than the other subspecies. However, the number of symptomatic plant hosts associated with subsp. *pauca* is smaller than for the other two abovementioned subspecies (Nunney et al. 2013; Yuan et al. 2010). This discrepancy may be biologically relevant; it may also be a consequence of large amounts of homologous recombination in *X. fastidiosa*, which would affect overall genetic diversity comparisons among these groups.

Genetic diversity studies based on field-collected samples, in which only information on host plant of origin is available for each isolate, have indicated that recombination within and between *X. fastidiosa* subspecies is frequent (Nunney et al. 2013; Yuan et al. 2010). In addition, *X. fastidiosa* is naturally competent (Kung and Almeida 2011), recombination can occur at frequencies orders of magnitude higher than mutation, and alleles can be transferred efficiently without selection (Kung and Almeida 2014). Horizontal gene flow substantially affects inferences on the evolutionary history of bacteria (Doroghazi and Buckley 2010; Fraser et al. 2007; Spratt et al. 2001), so much so that accurate placement of individual genotypes (or ST) on bifurcating phylogenetic trees is problematic. Three approaches were used to infer the clustering of the 15 South American *X. fastidiosa* ST

in relation to all presently available data. CC1, CC2, and CC3, together with ST53, -73, and -74 (reported from Costa Rica and Europe), are a monophyletic group currently considered as subsp. *pauca*. The placement of the clade with ST53, -73, and -74 raises important questions about *X. fastidiosa* diversity and biogeography. *X. fastidiosa* subsp. *pauca* is considered to be of South American origin but none of these ST have been detected in that region. It is possible that these ST are endemic to Central America but the allelic profile indicates that is not the case; we suggest that these were successful invasions from South America. The three CC in South America are statistically supported, indicating that clades within the subspecies exist as previously observed with other *X. fastidiosa* subspecies (Nunney et al. 2010, 2013). Within this context, all isolates obtained from citrus fall within CC1, while those from coffee trees are in CC2 and CC3. All ST were associated with only one host plant species, with the exception of ST16, which was recovered from coffee and olive trees. Similar observations have been found in the past for isolates of subsp. *fastidiosa* that infect grapevines and almond trees (Almeida and Purcell 2003). These observations raise questions about the fluidity of *X. fastidiosa* genotype–host plant species associations.

TABLE 2. Nucleotide diversity within four subspecies of *Xylella fastidiosa* based on the Jukes-Cantor genetic diversity index^a

Locus	Index for <i>X. fastidiosa</i> subsp.			
	<i>pauca</i>	<i>morus</i>	<i>fastidiosa</i>	<i>multiplex</i>
<i>cysG</i>	0.029	0.002	0.011	0.011
<i>gltT</i>	0.012	0.000	0.006	0.004
<i>holC</i>	0.036	0.003	0.026	0.023
<i>leuA</i>	0.014	0.000	0.004	0.007
<i>malF</i>	0.019	0.000	0.007	0.009
<i>nuoL</i>	0.030	0.002	0.006	0.007
<i>petC</i>	0.015	0.002	0.009	0.014
Loci concatenated	0.012	0.001	0.007	0.003

^a Subspecies *sandyi* was not included due to lack of diversity with the multilocus sequence type loci.

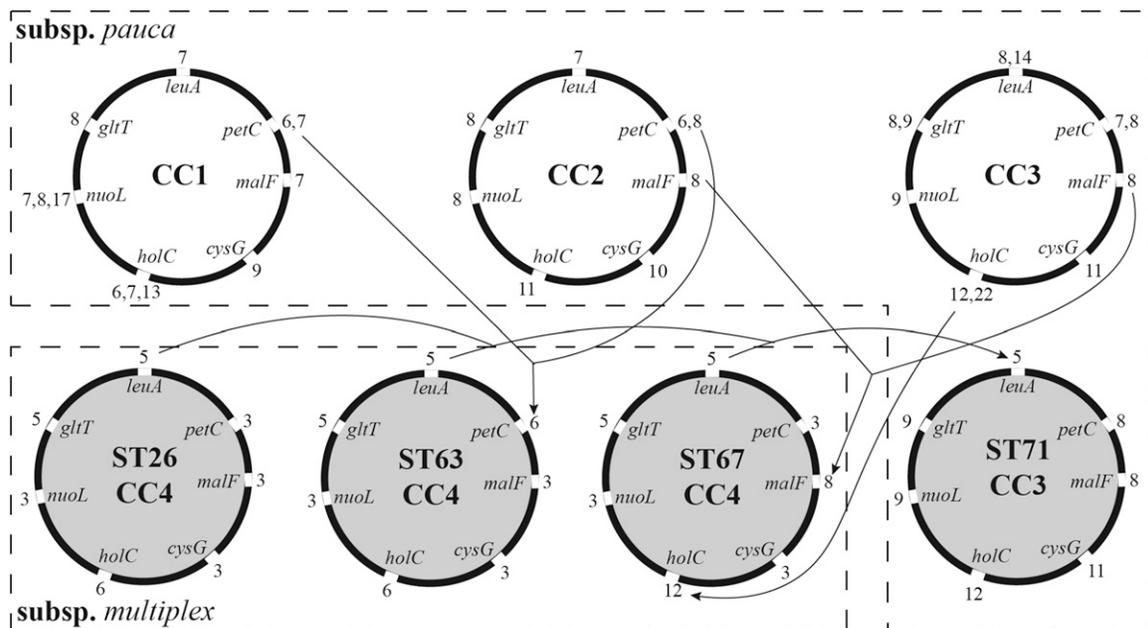


Fig. 4. Summary of the clonal complexes (CC) identified in South America in this study, which includes all sequence types (ST) currently known from the region. Subspecies assignments (*pauca* or *multiplex*) were based on the species-level phylogenies and allelic composition of typed strains; each locus includes the numbers for all alleles identified, and the placement of loci on each circle was arbitrary. ST isolated from plum hosts are shown individually in gray circles. Arrowheads illustrate the directionality of detected recombination events between subsp. *pauca* and *multiplex* ST. ST63 has a *petC* 6 allele originating from CC1 or CC2, while the *holC* and *malF* loci of ST67 originate from CC3 and CC2 or CC3, respectively. These events indicate gene flow from endemic to introduced genotypes. On the other hand, ST71 is a subsp. *pauca* genotype within CC3 with a *leuA* 5 allele that is characteristic of subsp. *multiplex*, indicating that this plum-isolated genotype has acquired at least one locus from the introduced CC4.

The placement of plum-associated ST (CC4 and ST71 in CC3) highlights the consequences of recombination on phylogenetic tree topology. ST71 clusters within CC3 but as a more basal taxon, primarily due to the presence of a *leuA* allele originating from subsp. *multiplex* (CC4). On the other hand, the three ST within CC4 are placed at disparate branches on the three trees that were generated, which is notable because that CC is likely to have originated from a single introduction (Nunes et al. 2003). ST26, a typical subsp. *multiplex* from North America, clusters within that group, while ST63 has the same profile but one allele originating from subsp. *pauca* and, therefore, is slightly divergent. However, ST67 has two loci with alleles from subsp. *pauca* and, therefore, its placement is basal within subsp. *multiplex*. An obvious conclusion derived from these observations is that single-locus genotyping of *X. fastidiosa* is inadequate for phylogenetic inferences or inferences of potential host plant species range.

In the case of *X. fastidiosa* in South America, Nunes et al. (2003) provided initial evidence of gene flow between the introduced subsp. *multiplex* and the endemic subsp. *pauca*. Almeida et al. (2008) and Nunney et al. (2012) showed further evidence of intragenic recombination between subsp. *pauca* and North American genotypes. Data presented here provide additional support to these observations, as well as expand the role of recombination in the evolution of *X. fastidiosa* and its associations with different host plants in South America. The introduced subsp. *multiplex* genotype had been initially reported from southern areas of South America (Argentina, Paraguay, and southern Brazil) associated with plum (Kitajima et al. 1975). In these regions, it appears to have remained genetically isolated from subsp. *pauca* whereas, in its northern distribution, it has exchanged alleles with genotypes infecting citrus (CC1) and probably coffee (CC2 and CC3). However, thus far, isolates belonging to CC4 subsp. *multiplex* have not been found associated with diseases in host plants other than plum. On the other hand, the contact between these two subspecies led to the detection of a subsp. *pauca* genotype that is capable of colonizing plum (ST71). Because no other subsp. *pauca* strains were found in plum, we propose that this occurrence supports the hypothesis that recombination is one of the drivers leading to novel *X. fastidiosa* diseases, as proposed by Almeida and Nunney (2015). In conclusion, our results corroborate proposals that recombination is a major element in *X. fastidiosa* evolution, occurring at short time frames and being associated with the emergence of new host–pathogen associations. Furthermore, although genetic distance among sequences reduces transformation efficiency (Kung et al. 2013), the results support the proposal that geographical isolation is the main factor limiting gene flow among *X. fastidiosa* genotypes, as previously suggested (Coletta-Filho et al. 2014). Ecological isolation via host plant specialization and vector species appears to allow for maintenance of genotypic clusters but not limit gene flow, as previously suggested (Almeida et al. 2008).

ACKNOWLEDGMENTS

This work was funded by FAPESP (São Paulo Research Foundation project numbers 2011/14479-0 and 2011/13803-9) and the California Agriculture Station. H. D. Coletta-Filho and J. R. S. Lopes received CNPq research fellowships (project numbers 306230/2013-5 and 309883/2011-3, respectively).

LITERATURE CITED

Almeida, R. P. P., Nascimento, F. E., Chau, J., Prado, S. S., Tsai, C.-W., Lopes, S. A., and Lopes, J. R. S. 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Appl. Environ. Microbiol.* 74:3690-3701.

Almeida, R. P. P., and Nunney, L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? *Plant Dis.* 99:1457-1467.

Almeida, R. P. P., and Purcell, A. H. 2003. Biological traits of *Xylella fastidiosa* strains from grapes and almonds. *Appl. Environ. Microbiol.* 69:7447-7452.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. 2004. Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19:535-544.

Coletta-Filho, H. D., Francisco, C. S., and Almeida, R. P. P. 2014. Temporal and spatial scaling of the genetic structure of vector-borne plant pathogen. *Phytopathology* 104:120-125.

Croll, D., and McDonald, B. A. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog.* 8:e1002608.

Daszak, P., Cunningham, A. A., and Hyatt, A. D. 2000. Emerging infectious diseases of wildlife: Threats to biodiversity and human health. *Science* 287:443-449.

Doroghazi, J. R., and Buckley, D. H. 2010. Widespread homologous recombination within and between *Streptomyces* species. *ISME J.* 4:1136-1143.

Felsenstein, J. 1989. Phylip—Phylogeny inference package (version 3.2). *Cladistic* 5:164-166.

Fraser, C., Hanage, W. P., and Spratt, B. G. 2007. Recombination and the nature of bacterial speciation. *Science* 315:476-480.

French, W. J., and Kitajima, E. W. 1978. Occurrence of plum leaf scald in Brazil and Paraguay. *Plant Dis. Rep.* 62:1035-1038.

Garten, R. J., Davis, C. T., Russell, C. A., Shu, B., Lindstrom, S., Balish, A., Sessions, W. M., Xu, X., Skepner, E., Deyde, V., Okomo-Adhiambo, M., Gubareva, L., Barnes, J., Smith, C. B., Emery, S. L., Hillman, M. J., Rivaller, P., Smagala, J., de Graaf, M., Burke, D. F., Fouchier, R. A. M., Pappas, C., Alpuche-Aranda, C. M., López-Gatell, H., Olivera, H., López, I., Myers, C. A., Faix, D., Blair, P. J., Yu, C., Keene, K. M., Dotson, P. D., Jr., Boxrud, D., Sambol, A. R., Abid, S. H., St. George, K., Bannerman, T., Moore, A. L., Stringer, D. J., Blevins, P., Demmler-Harrison, G. J., Ginsberg, M., Kriner, P., Waterman, S., Smole, S., Guevara, H. F., Belongia, E. A., Clark, P. A., Beatrice, S. T., Donis, R., Katz, J., Finelli, L., Bridges, C. B., Shaw, M., Jernigan, D. B., Uyeki, T. M., Smith, D. J., Klimov, A. I., and Cox, N. J. 2009. Antigenic and genetic characteristics of Swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 325:197-201.

Hacker, J., and Carniel, E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep.* 2:376-381.

Harris, J. L., and Balci, Y. 2015. Population structure of the bacterial pathogen *Xylella fastidiosa* among street trees in Washington D.C. *PLoS One* 10:e0121297.

Huelsensbeck, J., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.

Hufbauer, R. A., Facon, B., Ravigné, V., Turgeon, J., Foucaud, J., Lee, C. E., Rey, O., and Estoup, A. 2012. Anthropogenically induced adaptation to invade (AIAI): Contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evol. Appl.* 5:89-101.

Huson, D. H., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23:254-267.

Kelly, J. K. 1997. A test of neutrality based on interlocus associations. *Genetics* 146:1197-1206.

Kitajima, E. W., Bakarcic, M., and Fernandez-Valiela, M. 1975. Association of rickettsialike bacteria with plum leaf scald disease. *Phytopathology* 65:476-479.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C., and Frost, S. D. W. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23:1891-1901.

Kung, S. H., and Almeida, R. P. P. 2011. Natural competence and recombination in the plant pathogen *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 77:5278-5284.

Kung, S. H., and Almeida, R. P. P. 2014. Biological and genetic factors regulating natural competence in a bacterial plant pathogen. *Microbiology* 160:37-46.

Kung, S. H., Retchless, A. C., Kwan, J. Y., and Almeida, R. P. P. 2013. Effects of DNA size on transformation and recombination efficiencies in *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 79:1712-1717.

Librado, P., and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Loconsole, G., Saponari, M., Boscia, D., D'Attoma, G., Morelli, M., Martinelli, G. P., and Almeida, R. P. P. 2016. Intercepted isolates of *Xylella fastidiosa* in Europe reveal novel genetic diversity. *Eur. J. Plant Pathol.* 146:85-94.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003.

Minsavage, G., Thompson, C., Hopkins, D., Leite, R. M. V. B. C., and Stall, E. 1994. Development of a polymerase chain reaction protocol for detection of *Xylella fastidiosa* in plant tissue. *Phytopathology* 84:456-461.

Murray, K. A., and Daszak, P. 2013. Human ecology in pathogenic landscapes: Two hypotheses on how land use change drives viral emergence. *Curr. Opin. Virol.* 3:79-83.

Murray, K. A., Preston, N., Allen, T., Zambrana-Torrel, C., Hosseini, P. R., and Daszak, P. 2015. Global biogeography of human infectious diseases. *Proc. Natl. Acad. Sci. USA* 112:12746-12751.

- Nunes, L. R., Rosato, Y. B., Muto, N. H., Yanai, G. M., da Silva, V. S., Leite, D. B., Gonçalves, E. R., de Souza, A. A., Coletta-Filho, H. D., Machado, M. A., Lopes, S. A., and de Oliveira, R. C. 2003. Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements. *Genome Res.* 13:570-578.
- Nunney, L., Ortiz, B., Russell, S. A., Ruiz Sánchez, R., and Stouthamer, R. 2014a. The complex biogeography of the plant pathogen *Xylella fastidiosa*: Genetic evidence of introductions and subspecific introgression in Central America. *PLoS One* 9:e112463.
- Nunney, L., Schuenzel, E. L., Scally, M., Bromley, R. E., and Stouthamer, R. 2014b. Large scale intersubspecific recombination in the plant pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. *Appl. Environ. Microbiol.* 80:3025-3033.
- Nunney, L., Vickerman, D. B., Bromley, R. E., Russell, S. A., Hartman, J. R., Morano, L. D., and 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.
- Nunney, L., Yuan, X., Bromley, R., Hartung, J., Montero-Astúa, M., Moreira, L., Ortiz, B., and Stouthamer, R. 2010. Population genomic analysis of a bacterial plant pathogen: Novel insight into the origin of Pierce's disease of grapevine in the U.S. *PLoS One* 5:e15488.
- Nunney, L., Yuan, X., Bromley, R. E., and Stouthamer, R. 2012. Detecting genetic introgression: High levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. *Appl. Environ. Microbiol.* 78:4702-4714.
- Saponari, M., Boscia, D., Nigro, F., and Martelli, G. P. 2013. Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond, and olive trees exhibiting leaf scorching symptoms in Apulia (southern Italy). *J. Plant Pathol.* 95:659-668.
- Scally, M., Schuenzel, E. L., Stouthamer, R., and 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.
- Schuenzel, E. L., Scally, M., Stouthamer, R., and Nunney, L. 2005. A multi-gene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 71:3832-3839.
- Simon, J. A., Marrotte, R. R., Desrosiers, N., Fiset, J., Gaitan, J., Gonzalez, A., Koffi, J. K., Lapointe, F., Leighton, P. A., Lindsay, L. R., Logan, T., Milord, F., Ogden, N. H., Rogic, A., Roy-Dufresne, E., Suter, D., Tessier, N., and Millien, V. 2014. Climate change and habitat fragmentation drive the occurrence of *Borrelia burgdorferi*, the agent of Lyme disease, at the north-eastern limit of its distribution. *Evol. Appl.* 7:750-764.
- Spratt, B. G., Hanage, W. P., and Feil, E. J. 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr. Opin. Microbiol.* 4:602-606.
- Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4. Sinauer Associates, Sunderland, MA.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30:2725-2729.
- Verhagen, J. H., Herfst, S., and Fouchier, R. 2015. How virus travels the world. *Science* 347:616-617.
- Wells, J. M., Raju, B. C., Nyland, G., and Lowe, S. K. 1981. Medium for isolation and growth of bacteria associated with plum leaf scald and phony peach diseases. *Appl. Environ. Microbiol.* 42:357-363.
- Yuan, X., Morano, L., Bromley, R., Spring-Pearson, S., Stouthamer, R., and Nunney, L. 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. *Phytopathology* 100:601-611.