Differentiation of *Xylella fastidiosa* Strains via Multilocus Sequence Analysis of Environmentally Mediated Genes (MLSA-E)

Jennifer K. Parker,^a Justin C. Havird,^b and Leonardo De La Fuente^a

Department of Entomology and Plant Pathology, Auburn University, Auburn, Alabama, USA,^a and Department of Biological Sciences, Auburn University, Auburn, Alabama, USA^b

Isolates of the plant pathogen *Xylella fastidiosa* are genetically very similar, but studies on their biological traits have indicated differences in virulence and infection symptomatology. Taxonomic analyses have identified several subspecies, and phylogenetic analyses of housekeeping genes have shown broad host-based genetic differences; however, results are still inconclusive for genetic differentiation of isolates within subspecies. This study employs multilocus sequence analysis of environmentally mediated genes (MLSA-E; genes influenced by environmental factors) to investigate *X. fastidiosa* relationships and differentiate isolates with low genetic variability. Potential environmentally mediated genes, including host colonization and survival genes related to infection establishment, were identified *a priori*. The ratio of the rate of nonsynonymous substitutions to the rate of synonymous substitutions (dN/dS) was calculated to select genes that may be under increased positive selection compared to previously studied housekeeping genes. Nine genes were sequenced from 54 *X. fastidiosa* isolates infecting different host plants across the United States. Results of maximum likelihood (ML) and Bayesian phylogenetic (BP) analyses are in agreement with known *X. fastidiosa* subspecies clades but show novel within-subspecies differentiation, including geographic differentiation, and provide additional information regarding host-based isolate variation and specificity. dN/dS ratios of environmentally mediated genes, though <1 due to high sequence similarity, are significantly greater than housekeeping gene dN/dS ratios and correlate with increased sequence variability. MLSA-E can more precisely resolve relationships between closely related bacterial strains with low genetic variability, such as *X. fastidiosa* isolates. Discovering the genetic relationships between *X. fastidiosa* isolates will provide new insights into the epidemiology of populations of *X. fastidiosa*, allowing improved disease management in economically important crops.

Xylella fastidiosa is a Gram-negative plant-pathogenic bacterium that is adapted to live inside the xylem vessels of plant hosts and the foregut of leafhopper insect vectors (8). Biofilm formed inside the xylem vessels is believed to play a key role in disease development by obstructing water passage from the root to aerial parts of the plant (8). The bacterium infects several economically important agricultural crops, including grape, citrus, almond, and peach, as well as a variety of landscape trees and plants (38). Despite this broad host range, isolates of *X. fastidiosa* from different plant hosts are genetically very similar. Comparative genomic analyses between the most divergent *X. fastidiosa* complete genome strains (the Pierce's disease of grapevine type strain Temecula and the citrus variegated chlorosis type strain 9a5c) have shown that 98% of their genes are shared, and the average amino acid identity is 96% (85). Taxonomic analyses have grouped *X. fastidiosa* strains into several subspecies on the basis of DNA-DNA hybridization and phylogenetic studies: *X. fastidiosa* subsp. *fastidiosa*, *X. fastidiosa* subsp. *sandyi*, *X. fastidiosa* subsp. *multiplex*, and *X. fastidiosa* subsp. *pauca* (75, 76). Though isolates from these subspecies are genetically too similar for further taxonomic classification, studies on the biological traits of isolates from the same subspecies have indicated differences in infectivity (2, 3, 29, 37, 45, 51). Because of this, efforts to further characterize the diversity seen among *X. fastidiosa* isolates within the proposed subspecies are ongoing (91).

Diverse molecular techniques have been applied to genetically characterize *X. fastidiosa* isolates, including restriction fragment length polymorphism (RFLP) (10), random amplification of polymorphic DNA (RAPD) (3, 12, 32, 65), microsatellite/simple sequence repeat (SSR) analysis (2, 35, 48, 67), and 16S-23S rRNA

internal transcribed spacer (ITS) sequencing (32, 54, 56, 67). Results from strain typing, phylogenetic analyses, and other data comparisons in the aforementioned studies have shown broad host-based genetic differences between strains; however, phylogenetic results on differentiation of strains from within the same *X. fastidiosa* subspecies and those isolated from the same host plant species are still inconclusive. More recently, a multilocus sequence typing (MLST) scheme has been implemented (2, 74, 76, 91) to provide a standardized means of interpreting phylogenetic data from many isolates of *X. fastidiosa*. This MLST scheme relies on sequence analysis of seven housekeeping genes which are required for basic cellular maintenance. Housekeeping genes are usually relatively conserved, making them good candidates for use in evolutionary phylogenetic analyses of distantly related groups. One drawback to using housekeeping genes is that they may not provide enough variability to differentiate closely related isolates. The often-sequenced *gyrB* gene has 97.5 to 100% sequence similarity among *X. fastidiosa* isolates (69), exemplifying the limited variation found among *X. fastidiosa* housekeeping genes.

Similar problems detecting genetic variability, despite differences in infectivity characteristics, have frequently occurred in

Received 25 August 2011 Accepted 13 December 2011

Published ahead of print 22 December 2011

Address correspondence to Leonardo De La Fuente, lzd0005@auburn.edu.

Supplemental material for this article may be found at <http://aem.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AEM.06679-11

analyses of mammalian pathogen housekeeping genes. Researchers have resorted to analyses of alternate genes with greater variability to improve genetic differentiation of pathogen isolates and provide insight into recent evolutionary processes, including virulence genes (*Clostridium difficile*, *Chlamydomphila pecorum*, *Salmonella enterica* subsp. *enterica*, *Listeria monocytogenes*) (44, 90, 81, 92) and other genes related to pathogen-host interactions (*S. enterica* serovar Typhimurium, *Escherichia coli*, *S. enterica* subsp. *enterica*, *C. pecorum*) (28, 49, 50, 90). The recent *X. fastidiosa* MLST publication also included a nonhousekeeping gene related to motility (*pilU*) to increase phylogenetic resolution, and it provides most of the variation among nonrecombinant *X. fastidiosa* subsp. *fastidiosa* isolates, though it is excluded from the conventional MLST scheme (91). Such genes are likely subject to adaptive evolutionary pressure (positive selection), making them more suitable for population genetics study than for determining deep evolutionary relationships. Additionally, not all housekeeping genes are suitable for evolutionary phylogenies. A comparison of MLST schemes for 17 human pathogens showed that, though housekeeping genes were under purifying selection, on average, 13 of the species analyzed had evidence of punctual adaptive evolution within one or more loci (64). However, it has been argued that multilocus sequence analysis (MLSA) of genes subject to positive selection can still provide accurate predictions of short-term evolution, as long as the major clades identified with phylogenetic analysis of housekeeping genes are maintained in the analysis of the positively selected genes (31, 50).

An alternative for discriminating between closely related *X. fastidiosa* isolates is to perform MLSA of genes influenced by environmental factors, termed environmentally mediated genes. Environmentally mediated genes are involved in adaptation to environmental changes and are usually subject to positive selection pressure (35), so they should have greater sequence variability than conserved housekeeping genes under stabilizing selection pressure. Therefore, sequence analysis of environmentally mediated genes should provide increased phylogenetic differentiation and be more likely to yield genetic differences between recently diverged isolates, comparable to results from human bacterial pathogen MLST studies using virulence and other positively selected genes (44, 50, 81, 92). However, no previous studies have explored the diversity of virulence genes in a plant-pathogenic bacterium. Because *X. fastidiosa* is likely subject to strong selection pressure (76), it should have some positively selected, environmentally mediated genes with higher sequence variation. Previous studies have elucidated the essential roles of many specific adaptive genes involved in the proliferation of this xylem-limited pathogen in the host (e.g., adhesion, biofilm formation, motility, cell signaling, plant cell wall degradation, metal homeostasis, and toxin production genes; reviewed by Varani et al. [86]). It is hypothesized that analyzing these genes in *X. fastidiosa* will increase differentiation of genetically similar isolates, in particular, isolates infecting the same plant host.

Here, multilocus sequence analysis of environmentally mediated genes (MLSA-E) was applied to identify *X. fastidiosa* isolate relationships. MLSA-E was used for genes related to processes important for establishing *X. fastidiosa* infections and colonization of the insect vector, such as surface attachment, biofilm formation, virulence, and nutrient transport and utilization. These types of genes may be more relevant to host-based genetic variability. Genes of interest were identified from previous studies

and, based on the ratio of the rate of nonsynonymous substitutions to the rate of synonymous substitutions (dN/dS), an indication of positive selection. Target-specific PCR primers were designed from available *X. fastidiosa* whole-genome sequences. Target genes from *X. fastidiosa* isolates from a variety of host plants from several geographic regions in the United States were PCR amplified and sequenced. Phylogenetic analyses of the resulting sequence information show host-based and geographic origin-based genetic relationships. Discovering the genetic relationships between *X. fastidiosa* isolates will provide new insights into the epidemiology of populations of *X. fastidiosa*, which will allow improved disease management in a variety of economically important agricultural crops.

MATERIALS AND METHODS

Isolation of *X. fastidiosa*. Few *X. fastidiosa* isolates from the southeastern United States, with the exception of isolates from Florida, have been collected and genetically analyzed. To improve representation of *X. fastidiosa* isolates from the southeastern United States, isolates were collected from Georgia and Alabama. Leaves showing *X. fastidiosa* infection symptoms of marginal leaf scorching were collected from host plants in these states and stored at 4°C. Within 24 h, leaves were processed according to a modified bacterial isolation protocol (36). Briefly, each leaf was surface sterilized with 70% ethanol (2 min), followed by 1.2% sodium hypochlorite (5 min), and then rinsed 3 times with sterile deionized water. The lower half of the main vein and the petiole (with tip trimmed and discarded) were removed with a sterile blade. The tissue was cut along the length of the vein, sliced into small pieces, and added to 3 ml PD2 (17) liquid medium. Medium with leaf tissue was incubated with shaking at 28°C for 2 h. PD2 samples of 100 μ l were spread plated on PW agar plates (16) and incubated at 28°C for up to 1 month, until presumptive *X. fastidiosa* colonies appeared. Colonies were restreaked for isolation, confirmed by PCR (58), and stored in PD2 with 20% glycerol at -80°C.

***X. fastidiosa* isolate collection and DNA extraction.** *X. fastidiosa* isolates ($n = 49$) were isolated as discussed above ($n = 8$) or acquired from other laboratories ($n = 41$). Isolates were originally collected from 33 sampling sites in the following U.S. states: Alabama ($n = 1$), California ($n = 13$), Florida ($n = 11$), Georgia ($n = 18$), and Texas ($n = 6$) (see Fig. S1 and Table S1 in the supplemental material). Isolates were cultured for 1 to 2 weeks on PW agar (16). Bacteria were then scraped from the agar plates, resuspended in 200 μ l molecular-grade water, and DNA extracted using a modified cetyltrimethylammonium bromide protocol (22). Sequences for isolates with complete genomes deposited in GenBank ($n = 5$) were also included in this study (see isolates identified with crosses in Table S1 in the supplemental material). Of the combined isolates analyzed here ($n = 54$), the majority were isolated from grape ($n = 28$), followed by blueberry ($n = 7$), almond ($n = 3$), oak ($n = 3$), lupine ($n = 2$), sycamore ($n = 2$), and elderberry, elm, oleander, orange, plum, ragweed, redbud, sunflower, and sunflower ($n = 1$ for each).

MLSA-E gene selection. Genes with the following characteristics were initially chosen for MLSA (Table 1): (i) experimental proof from previous studies regarding their role in the infection process and indicating the possibility that they are environmentally mediated and (ii) a high probability of sequence variability, based on comparisons of the five *X. fastidiosa* complete genome sequences deposited in GenBank (Temecula, accession number [AE009442](#); GB514, accession number [CP002165](#); M23, accession number [CP001011](#); M12, accession number [CP000941](#); 9a5c, accession number [AE003849](#)). Genes related to processes important for establishing *X. fastidiosa* infections, such as surface attachment, biofilm formation, virulence, and nutrient transport/utilization, were initially selected *a priori* from the literature, with preference given to genes in each category with the most well elucidated roles. A BLAST search was conducted to identify the candidate genes with the greatest sequence variation among the five complete genome sequences. Ten genes representing sev-

TABLE 1 Genes selected for final *X. fastidiosa* MLSA-E analysis

Gene	Locus tag (Temecula)	Gene size (bp)	Category	Description	Forward primer (5'→3')	Reverse primer (5'→3')	Amplicon size (bp)	Reference(s)
<i>acvB</i>	PD_1902	900	Virulence	Virulence protein: suggested to regulate pathogenicity and disease symptoms	ACAGTA TCGCCG TC GAAG TGATGA	CATGCA TACRGC GATGYT TCCGAT	743	34
<i>copB</i>	PD_0101	930	Resistance	Copper resistance protein B precursor: encodes a copper resistance cation-translocating ATPase induced by copper exposure	ATGAAC ACCCGT AC CTGG TTCGTA	ATTTAG TCTCCA CCATGA GCCGCA	607–862	68, 82
<i>cvaC^a</i>	PD_0215	309	Toxin	Colicin V precursor: encodes a bacteriocin precursor proposed to be a defense mechanism	TGCGTG AATTRA CA TTGA CCG	CCTAGT CTGCGG CTTAAG CAGATT	330	20, 62
<i>fimA</i>	PD_0062	555	Attachment	Fimbrial subunit precursor: component of type I pili important for biofilm formation and aggregation	CCCAGT GCGTCG TTATCG ATTATT GGT	TTTYGY ACTCTC AA GCAT CGCATC	557	18–20, 24, 25, 46
<i>gaa</i>	PD_0315	1,992	Toxin	Glutaryl-7-aminocapthosporanic acid acylase precursor: encodes a member of the newly described family of β -lactam antibiotic acylases	TGAGAG CTGCGY ATGTTC CAATGA	ACAGCT TCTGGC AAGAAC AAGCAC	1129	4, 5
<i>pglA^a</i>	PD_1485	1,635	Cell wall degradation	Polygalacturonase precursor: needed for degrading host plant cell walls to allow colonization	TAGTGC TGGCCT AA CGAT GTYGGT	CCGTAT CAGCAA CCACAT GGAAGT	828/829	63, 71
<i>pilA</i>	PD_1924	447	Motility	Fimbrial protein: major structural protein of type IV pili used for twitching motility and upstream migration	ATCGCK CTGCCY AT GTAC CAAA	CAGCAT TGATCG TRTTCG TGRTG	405	57
<i>rpfF</i>	PD_0407	873	Pathogenicity	Regulator of pathogenicity factors: involved in biosynthesis of diffusible signal factor (DSF) to mediate cell-to-cell signaling for biofilm formation	GCGCTC CATAGT TC GGAG TGATTT	ATGTCC GCTGTA CATCCC ATTCCT	825	8, 9, 60
<i>xadA^a</i>	PD_0731	2,994	Attachment	Outer membrane afimbrial adhesin: may contribute to initial cell binding to surfaces	TGGGAG GTCAAA GYACTG CCATCA	GCATTG GCAGCA ACACTC GAATCA	1087/1108	25

^a Genes selected on the basis of *dN/dS*.

eral categories of potential environmental interactions were initially selected (Table 1; see Table S2 in the supplemental material). General sequence variability for each gene was assessed by quantifying the number of unique haplotypes on the basis of preliminary results using 28 *X. fastidiosa* isolates and these 10 genes.

The *dN/dS* between the Pierce's disease of grapevine type strain (*X. fastidiosa* strain Temecula) and the citrus variegated chlorosis type strain (*X. fastidiosa* strain 9a5c) was calculated for the 10 MLSA-E genes using the DnaSP program, version 5 (47). *dN/dS* is an indicator of selection pressure and was calculated to determine if this ratio could be used to predict additional MLSA-E genes with greater sequence variability and, if so, what range of *dN/dS* ratios would be expected for high-variability sequences. Because *X. fastidiosa* isolates are genetically very similar and strains Temecula and 9a5c have the most dissimilar complete genome sequences available, genes with a higher Temecula/9a5c *dN/dS* may have greater sequence variability among more closely related isolates. To test this hypothesis, Temecula/9a5c *dN/dS* ratios for the 10 initial MLSA-E genes were plotted against the number of alleles found among the 28 *X. fastidiosa* isolates initially analyzed. Results from this (see below) led to selection of three additional genes for the final MLSA-E (Table 1, asterisks) on the basis of the following new criteria: (i) possibility of being environmentally mediated on the basis of the findings of previous studies and (ii) a Temecula/9a5c *dN/dS* of >0.15, an indication of positive selection pressure. Finally, Temecula/9a5c *dN/dS* ratios for all 13 genes regions

were plotted against the number of alleles found among the 28 initially analyzed *X. fastidiosa* isolates to determine the correlation between *dN/dS* and sequence variability from the larger data set. Due to lower sequence variability and number of alleles, as predicted by the Temecula/9a5c *dN/dS*, four of the initial genes were eliminated from the final MLSA-E of all 54 isolates (see Table S2 in the supplemental material).

Detection of positive selection. *dN/dS* ratios were calculated for the set of housekeeping gene regions described in a previous MLST study of *X. fastidiosa* (91) to determine if their *dN/dS* ratios were significantly different from those used in the final MLSA-E. All gene regions (of both the final MLSA-E and MLST genes) were designated either housekeeping or nonhousekeeping genes on the basis of the literature. *gyrB*, an additional housekeeping gene previously used for phylogenetic analysis of *X. fastidiosa* (59, 69), was also included. *dN/dS* ratios from the housekeeping genes were compared with *dN/dS* ratios from the nonhousekeeping genes using a Wilcoxon rank-sum test ($\alpha = 0.05$) in Statistix software, version 8.0 (Analytical Software).

PCR amplification and DNA sequencing. PCR primers (Table 1; see Table S2 in the supplemental material) were designed for the selected genes using PrimerQuest software (Integrated DNA Technologies, Coralville, IA), with the exception of those for the *malF* primers (91). Isolate DNA was amplified by PCR on an S1000 thermocycler (Bio-Rad Laboratories, Hercules, CA) in reaction mixtures (50 μ l) containing the following components: 1.25 U PerfectTaq DNA polymerase (5Prime, Gaithers-

burg, MD), 1× PerfectTaq PCR buffer, 0.2 mM deoxynucleoside triphosphates, 0.2 μM each primer, and 1 μl DNA template. Cycling parameters for all genes were 94°C for 3 min, followed by 25 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 45 s, with a final step of 72°C for 10 min. PCR products were verified by gel electrophoresis. DNA sequencing was conducted on an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA) by Lucigen Corporation (Middleton, WI).

Phylogenetic analyses. Geneious software, version 5.4 (23), was used to trim low-quality sequence end reads, create consensus gene sequences, translate putative protein sequences, and align sequences (using the Geneious ClustalW [43] plug-in). For the *pglA* gene, the alignment was trimmed to exclude the noncoding region at the beginning of the PCR amplicon and the region after the frameshift mutation that has been described in *X. fastidiosa* subsp. *pauca* (85). For the *copB* gene, 1 bp was deleted from the GB514 isolate sequence in GenBank (locus tag XFLM_05775) because it was suspected to be a sequencing error (see below).

To reconstruct the evolutionary relationships of the *X. fastidiosa* isolates, phylogenies were generated for each individual gene alignment and for the concatenated sequence alignment (see File S1 in the supplemental material). Bayesian phylogenetic (BP) analyses were conducted using the MrBayes program, version 3.1.2 (70), with two chains of 2 million generations each. Partitions were created for each codon position of each gene, and models of nucleotide substitution were applied to each partition, as determined by the Modeltest program, version 3.7 (66), according to the Akaike information criterion (1). Trees were sampled every 200 generations, and the first 5,000 trees (10%) were discarded as burn-in for each chain, prior to generating the extended-majority-rule consensus tree. Maximum likelihood (ML) analyses were performed using the RAxML program, version 7.0.3 (79), and the rapid BS algorithm (80) with 1,000 bootstraps. For ML analyses, each codon position was treated as a separate partition, and the GTR + I + Γ model was implemented for each position using the GTRGAMMAI option. Trees were initially rooted with *Xanthomonas campestris* pv. *campestris* strain ATCC 33913 as the outgroup, but final trees were rooted with *X. fastidiosa* isolate 9a5c (orange) as the outgroup, based on initial results.

To compare topologies between the single-gene ML trees and the concatenated ML tree, log likelihood differences ($\Delta \ln L$) were computed and Shimodaira-Hasegawa (SH) tests (78) were performed between each gene tree and the concatenated gene tree using the *-f h* option in RAxML. The nine MLSA-E genes were also divided into groups of high-*dN/dS* genes and groups of low-*dN/dS* genes, based on the Temecula/9a5c MLSA-E gene region *dN/dS*. ML phylogenies generated from data sets with increasing numbers of high- or low-*dN/dS* genes were compared to the phylogeny generated from the complete nine-gene data set using SH tests to determine the minimum number of genes of each type needed to obtain a tree not significantly worse than the nine-gene tree. An SH test was also used to determine if genes with evidence of recombination (see methods below) have a significant effect on the nine-gene ML tree. This was done by removing potentially recombinant genes from the alignment, producing an ML tree with the remaining genes, and using the SH test to compare it to the original tree.

Haplotype networks. To visualize relationships between haplotypes, networks based on the nine-gene alignment were created (see File S1 in the supplemental material) using the TCS program, version 1.21 (13), which uses the previously described cladogram estimator algorithm (84). The default settings were used (i.e., branches represent 95% parsimoniously plausible connections between haplotypes). Loops (reticulations) in the network, which represent ambiguous and implausible connections, were broken according to previous recommendations (15, 83).

Analysis of sequence diversity, recombination, and linkage disequilibrium. Sequence diversity of the nine final single-gene alignments was described by calculating nucleotide diversity per site (π) using DnaSP software, version 5 (47). Single-gene alignments were also tested for evidence of recombination using both the RDP4 (Recombination Detection

FIG 1 Correlation between Temecula/9a5c *dN/dS* ratios for the 13 initially analyzed genes and the number of alleles identified for each gene among a subset ($n = 28$) of the *X. fastidiosa* isolates. The dashed line represents the *dN/dS* threshold of 0.15 used to select genes for final MLSA-E. Genes with *dN/dS* ratios above this threshold (with the exception of *pilJ*) were retained for final MLSA-E analyses, while genes with *dN/dS* ratios below this threshold were eliminated.

Program 4) Beta program, version 4.11 (52), and PHI (pairwise homoplasy index) test (6). RDP tests the alignment with a suite of nine nonparametric recombination detection methods that provide the number, breakpoints, and statistical probability of recombination events, while the PHI test scans the alignment for overall evidence of recombination. Significance of all recombination events was tested at a P value of 0.05. Alignments were also analyzed for linkage disequilibrium using the ZZ statistic (72), as implemented in DnaSP, version 5 (47).

***X. fastidiosa* MLST scheme comparative analyses.** To compare the *X. fastidiosa* MLST scheme (74, 91) with the MLSA-E scheme developed here, haplotype diversity was investigated using the seven MLST housekeeping genes versus the nine MLSA-E genes. Data from the online *X. fastidiosa* MLST database were compared against data for the *X. fastidiosa* isolates used in this study, and 12 isolates analyzed were identified to be in common: 9a5c, ALS6, Conn Creek, EB92-3, Georgia Plum, Hopland, L95-1, L95-2, M12, M23, Preston, and Temecula (see Table S1 in the supplemental material). Sequence data for the MLST genes were downloaded from the website, concatenated alignments were created, and number of haplotypes was determined.

Nucleotide sequence accession numbers. The GenBank accession numbers for the 441 new sequences generated here are JQ361912 to JQ362352.

RESULTS

Gene selection for MLSA-E. A positive correlation between Temecula/9a5c *dN/dS* ratios of the initial 10 MLSA-E genes and the number of alleles among the 28 *X. fastidiosa* isolates initially analyzed was observed. On the basis of this, genes with *dN/dS* ratios of >0.15 were postulated to have sufficient variability for MLSA-E because they were predicted by the trend line to have a greater number of alleles and the potential to be under positive selection. To select additional genes with high *dN/dS* ratios ($dN/dS > 0.15$), the Temecula/9a5c *dN/dS* was calculated for numerous potential environmentally mediated genes identified in the literature (data not shown). This led to selection of three additional MLSA-E genes with the highest *dN/dS* ratios of the genes surveyed. A significant positive correlation ($R^2 = 0.66$, $P = 0.0007$ for regression analysis) was found between Temecula/9a5c *dN/dS* ratios for the 13 target MLSA-E gene regions sequenced and the number of alleles identified among the 28 *X. fastidiosa* isolates initially analyzed (Fig. 1). On the basis of these results, nine genes with *dN/dS* ratios of >0.15 were chosen for final MLSA-E of the 54 *X. fastidiosa* isolates. The *dN/dS* ratios of the nonhousekeeping genes in

TABLE 2 Temecula/9a5c *dN/dS* for housekeeping genes, nonhousekeeping genes, and the final MLSA-E gene regions used in this study^a

Housekeeping genes		Nonhousekeeping genes		Final MLSA-E gene regions	
Gene	<i>dN/dS</i>	Gene	<i>dN/dS</i>	Gene	<i>dN/dS</i>
<i>cysG</i>	0.19	<i>acvB</i>	0.29	<i>acvB</i>	0.26
<i>glrT</i>	0.09	<i>copB</i>	0.28	<i>copB</i>	0.26
<i>gyrB^b</i>	0.06	<i>cvaC</i>	0.34	<i>cvaC</i>	0.35
<i>holC</i>	0.11	<i>exbD1</i>	0.11		
<i>leuA</i>	0.07	<i>fimA</i>	0.40	<i>fimA</i>	0.37
<i>malF</i>	0.10	<i>gaa</i>	0.23	<i>gaa</i>	0.22
<i>nuoL</i>	0.10	<i>pglA</i>	0.21	<i>pglA</i>	0.21
<i>petC</i>	0.28	<i>pilA</i>	0.24	<i>pilA</i>	0.30
<i>pyrE^c</i>	0.08	<i>pilJ</i>	0.10		
		<i>rpjF</i>	0.19	<i>rpjF</i>	0.19
		<i>xadA</i>	0.30	<i>xadA</i>	0.36
Mean	0.12		0.24		0.28
Median	0.10		0.24		0.26

^a Unless noted, housekeeping genes are from the MLST study (91). Nonhousekeeping genes are from this study. The *dN/dS* ratios of the nonhousekeeping genes were significantly higher than the *dN/dS* ratios of the housekeeping genes (Wilcoxon rank-sum test, two-tailed $P = 0.0013$).

^b *gyrB* used in other *X. fastidiosa* phylogenetic analyses (59, 69).

^c *pyrE* used in initial MLSA-E analyses in this study.

this study (mean = 0.24) were significantly higher than those of the housekeeping genes used in the MLST *X. fastidiosa* study (91) and this study (mean = 0.12) (Wilcoxon rank-sum test, two-tailed $P = 0.0013$) (Table 2).

MLSA-E gene sequences. Sequencing results indicate specific characteristics for several genes analyzed which represent novel findings. In agreement with previous reports (85), the *pglA* frame-shift mutation seen in *X. fastidiosa* subsp. *pauca* isolates is absent in the non-*X. fastidiosa* subsp. *pauca* isolates analyzed in this study. There is a 1-bp insertion in the GB514 strain *copB* sequence in GenBank that likely represents a sequencing error. The insertion produces a stop codon that results in truncation of the gene, whereas the gene is translated in other complete genome sequences and *X. fastidiosa* isolates from grape in this study (including another isolate from Texas). *copB* functions in copper homeostasis in *X. fastidiosa* (68), and inactivation of the homologous *copAB* operon in *Xanthomonas axonopodis* pv. *citri* rendered the bacterium incapable of growing in medium containing copper (82), so this gene is probably necessary for *X. fastidiosa* survival. Also, *copB* contains a region with a 15-bp (minisatellite)-coding variable number tandem repeat (VNTR). When oriented in the coding direction, a repeat of 5'-ACACCCAGATGGATC-3' occurs (described here for the first time), followed by a repeat which is a previously described variation of this sequence: 5'-ACACAGGGATGGATC-3' (48). The two substitutions that account for the difference in the two repeats represent a nonsynonymous mutation, whereby glutamine (CAG) versus glycine (GGG) is translated in the first versus second repeat, respectively. It is also notable that *xadA* contains a 21-bp indel. The deletion is present in all unique *X. fastidiosa* subsp. *fastidiosa* isolate groups except for haplotype groups GR_GAFL (9 Georgia grape and 2 Florida grape isolates) and EB_LUPINE (1 Florida elderberry isolate and 1 Florida lupine isolate). The insertion is present in all other isolates examined here.

Haplotype diversity. A total of 6,026 bp from nine genes was sequenced or obtained from public databases for 54 *X. fastidiosa* isolates. Analysis of the final concatenated nine-gene alignment for all 54 *X. fastidiosa* isolates (49 isolates plus data from four full-genome sequences) revealed 29 unique haplotypes (see Table S1 in the supplemental material). Each haplotype was represented by 1 to 11 isolates. The unique haplotypes demonstrate host-based differences between isolates. Grape and blueberry isolates are the largest groups from a single host presented here. Ten unique haplotypes were identified from 28 grape isolates, and three unique haplotypes were identified from 7 blueberry isolates. This confirms that genetic variability is present among isolates obtained from the same host plant species.

Of the 10 grape isolates collected from Georgia, all had identical MLSA-E genes, except for CCPM1, which was acquired from plant tissue collected in the same county on the same day as four of the other isolates, but from a different cultivar (Petit Manseng). Using the 12 isolates shared by both MLST and MLSA-E schemes, 10 unique haplotypes were identified by MLSA-E analysis, while only 8 were identified using the MLST genes (data not shown).

Phylogenetic analyses. Phylogenetic analyses were conducted on the concatenated nine-gene alignment of the 29 unique haplotypes (see File S1 and Table S1 in the supplemental material). Trees rooted using *Xanthomonas campestris* as the outgroup did not have relationships among unique *X. fastidiosa* haplotypes noticeably different from those of trees rooted with the *X. fastidiosa* subsp. *pauca* isolate (9a5c), so *X. fastidiosa* subsp. *pauca* was used to produce a tree with a shorter root for easier viewing and to facilitate ease of use of the sequence alignments for additional analyses. Tree topologies were similar for BP and ML phylogenetic analyses of the 29 unique haplotypes. Therefore, both posterior probabilities (PPs) and bootstrap support (BS) values are presented on the 50% majority rule consensus Bayesian topology (Fig. 2). Support values for most clades were generally high (>85% BS and >95% PP), and the tree is well resolved, with the exception of polyphyletic groups among the most-derived, closely related isolates from each lineage. However, in the ML analysis, BS values for a single *X. fastidiosa* subsp. *multiplex* clade were low (<50%), with *X. fastidiosa* subsp. *multiplex* clade A and clade B and the isolate from elm forming three clades as polytomies. In the BP analysis, the elm isolate was also the most basal in the single *X. fastidiosa* subsp. *multiplex* clade, suggesting that it may represent a rogue taxon (73) with a highly variable phylogenetic position.

Another difference between the BP and ML analyses occurs in terminal leaves within *X. fastidiosa* subsp. *fastidiosa*. Within this clade, most of the variability found in the sequences analyzed here is due to insertion/deletion events in the genes *copB* and *xadA*. Under most standard phylogenetic implementations (including RAXML and MrBayes), gaps and missing data are treated as missing data (26; Alexandros Stamatakis, Heidelberg Institute for Theoretical Studies, personal communication). Thus, gaps such as those found in this study do not contribute any phylogenetic information. However, it appears that RAXML is more influenced by these missing data, which can be attributed to the different strategies deployed by ML versus BP to infer/integrate over trees (Stamatakis, personal communication). As a result, whereas MrBayes produces a large polytomy for *X. fastidiosa* subsp. *fastidiosa* isolates, RAXML produces a more defined topology with variable (55 to 100%) BS support values (Fig. 3).

All single-gene trees (data not shown) have significantly lower

FIG 2 Phylogeny of the 29 unique haplotypes found among all *X. fastidiosa* isolates analyzed ($n = 54$) as inferred by BP analysis of the nine concatenated gene sequences. Nodes with $<50\%$ support were collapsed. BP posterior probabilities and maximum likelihood bootstrap values are presented above and below each node, respectively. Asterisks indicate 100% support values, while dashes indicate relationships not supported on the basis of ML analysis. See Fig. 3 for the expanded *X. fastidiosa* subsp. *fastidiosa* subtree.

Δln *L* values than the concatenated gene tree based on SH tests (Table 3). Additionally, the nine MLSA-E genes were further divided into groups of high-*dN/dS* genes and groups of low-*dN/dS* genes on the basis of the Temecula/9a5c MLSA-E gene region

dN/dS (Table 2). By comparing ML trees from different numbers of the highest-*dN/dS*-ratio genes using SH tests, it was determined that the six highest-*dN/dS*-ratio genes are needed to obtain a tree similar to the nine-gene tree ($\Delta \ln L$, -1.7). However, comparison of the ML tree from the eight lowest-*dN/dS*-ratio genes (only *fimA* was excluded) with the nine-gene tree showed that it was still significantly worse according to the SH test ($\Delta \ln L$, -145).

Haplotype networks. Isolates in the haplotype networks (Fig. 4) group together mostly along the lines of the five major lineages seen in phylogenetic analyses (Fig. 2): *X. fastidiosa* subsp. *pauca*, *X. fastidiosa* subsp. *multiplex* clade A, *X. fastidiosa* subsp. *multiplex* clade B, *X. fastidiosa* subsp. *sandyi*, and *X. fastidiosa* subsp. *fastidiosa*. Two additional haplotype networks were also recovered. In this analysis, *X. fastidiosa* subsp. *fastidiosa* is divided into two haplotype networks. One of the networks consists only of isolates from Georgia and Florida. The other network consists of all of the isolates from California and Texas, one isolate from Georgia (CCPM1), and one isolate from Florida (PD 92-8). The isolate from the elm, which is the only isolate from Alabama represented in this study, also forms a unique clade, supporting the ML phylogenetic analysis which placed it as a unique lineage not belonging to *X. fastidiosa* subsp. *multiplex* clade A or clade B. Elm is also the most basal isolate in *X. fastidiosa* subsp. *multiplex* clade A according to Bayesian analyses (Fig. 2).

Evidence for recombination and linkage disequilibrium. Of the nine genes used for the final MLSA-E, there is evidence for single intersubspecies recombination events each in particular alleles of the genes *acvB* and *rpfF* (Table 3). Specifically, of the 10

FIG 3 Expanded subtree for *X. fastidiososa* subsp. *fastidiososa* showing variation in isolates from grape. Branch lengths are transformed to enable viewing of branches and relationships between haplotypes. BP posterior probabilities and maximum likelihood bootstrap values are presented above and below each node, respectively. Asterisks indicate 100% support values, while dashes indicate no support on the basis of ML analysis.

TABLE 3 Statistics for sequence diversity, recombination, and linkage disequilibrium for *X. fastidiosa* unique isolate groups^a

Gene	Sequence diversity			Recombination					Linkage disequilibrium (ZZ statistic, all sites)	$\Delta \ln L$ by single-gene tree comparison
	No. of Alleles	No. of alleles (gaps excluded)	Nucleotide diversity per site (π)	No. of recombination events	No. of methods detecting recombination ^b	Size of recombination breakpoints in alignment (bp)	Major parent isolate	Recombinant isolate		
<i>acvB</i>	10	10	0.037	1	5	220-315	Orange	<i>X. fastidiosa</i> subsp. <i>multiplex</i>	0.238	−469
<i>copB</i>	18	7	0.009	0	0				0.049	−480
<i>cvaC</i>	9	9	0.033	0	0				−0.044	−1,556
<i>fimA</i>	11	11	0.035	0	0				0.143	−1,446
<i>gaa</i>	7	7	0.007	0	0				−0.018	−1,016
<i>pglA</i>	7	7	0.015	0	0				0.030	−472
<i>pilA</i>	10	10	0.084	0	0				0.056	−243
<i>rpjF</i>	6	6	0.029	1	7	4652-4951	Oleander	<i>X. fastidiosa</i> subsp. <i>fastidiosa</i>	0.223	−863
<i>xadA</i>	14	13	0.015	0	1 ^c				0.024	−270

^a $n = 29$ unique isolate groups. The $\ln L$ differences between the single-gene tree and the concatenated nine-gene tree are also presented; all single-gene trees were significantly different from the concatenated nine-gene tree on the basis of Shimodaira-Hasegawa tests. Alignment regions with gaps were excluded from calculations.

^b Out of 10 possible methods.

^c Result is from PHI test, which provides information only about overall evidence of recombination.

recombination detection methods used, 5 and 7 methods detected a recombination event in *acvB* and *rpff*, respectively. Furthermore, a larger ZZ statistic (72) indicates a decreased chance of linkage disequilibrium for *acvB* and *rpff* compared to that for the other genes (Table 3), which further supports the possibility of these recombination events. However, the regions of the alignments within the recombination breakpoints (see File S1 in the supplemental material) have no novel variability patterns compared to other areas of the alignment. Additionally, the recombinant region is shared only by all *X. fastidiosa* subsp. *multiplex*

isolates, in the case of *acvB*, and all *X. fastidiosa* subsp. *fastidiosa* isolates, in the case of *rpff* (Table 3), rather than being present in select isolates from each of these subspecies. Furthermore, the SH test comparing the nine-gene ML tree with an ML tree produced without the genes *acvB* and *rpff* showed that though the trees were significantly different, as expected, with two genes removed, one of which had a medium dN/dS (dN/dS for *acvB*, 0.26), there was a low log likelihood difference ($\Delta \ln L$, -42). This log likelihood difference is much lower than that for any of the single-gene trees ($\Delta \ln L$ range, -243 to -1,556; Table 3) and is pointedly also much lower than that for the previously mentioned tree produced with the eight lowest- dN/dS -ratio genes (only *fimA* was excluded; $\Delta \ln L$, -145).

DISCUSSION

Variation in *X. fastidiosa* detected by MLSA-E. Previous studies of *X. fastidiosa* genetic variability have not detected strong genetic variation among *X. fastidiosa* strains within subspecies (76). Phylogenetic variation has been detected among *X. fastidiosa* subsp. *pauca* (2, 87) and *X. fastidiosa* subsp. *multiplex* (33, 74, 91) isolates but has been particularly limited among *X. fastidiosa* subsp. *fastidiosa* and *X. fastidiosa* subsp. *sandyi* isolates (33, 74, 76, 91), with the exception of *X. fastidiosa* subsp. *fastidiosa* isolates from Costa Rica (61). For this reason, it has been assumed that within-subspecies isolates of *X. fastidiosa* are very similar (61). However, this does not explain the infectivity differences (including symptom type and severity or population growth rate) that have been described among different isolates within *X. fastidiosa* subsp. *fastidiosa* (3, 37, 51), *X. fastidiosa* subsp. *multiplex* (3, 29), and *X. fastidiosa* subsp. *pauca* (2, 45). Recent studies (11, 14) indicate that genetic variability does exist among U.S. *X. fastidiosa* isolates infecting grape (*X. fastidiosa* subsp. *fastidiosa*). This suggests the presence of undetected genetic variation and has prompted exploration of MLSA-E as a new method for genetic analysis of *X. fastidiosa*. This MLSA-E study is the first to use variation in multiple

FIG 4 Haplotype networks for all *X. fastidiosa* isolates from the concatenated nine MLSA-E genes, coded by geographic origin. Branches (lengths not relevant) represent 95% parsimoniously plausible connections between haplotypes. Size of nodes indicates frequency, and black dots indicate missing haplotypes. The ancestral haplotype in each network is indicated by a square.

nonhousekeeping gene DNA sequences to describe differences among plant-pathogenic bacteria.

Here, *X. fastidiosa* genes possibly influenced by host/vector-bacterium interactions were selected for MLSA-E, leading to detection of sequence variability greater than that found in housekeeping genes. The life history of *X. fastidiosa* is restricted to plant host xylem vessels and the insect vector foregut. As a result, processes related to bacterial attachment to surfaces, biofilm formation, quorum sensing, degradation of host cell walls, and other adaptive life strategies are crucial for disease development (40). Genes involved in these processes are expected to reflect adaptations to this specialized environment. The genes selected for this study can be broadly classified into two categories: those used by *X. fastidiosa* for host colonization (*fimA*, *pglA*, *pilA*, *rpfF*, *xadA*) and those used for defense against environmental conditions and competing microbiota (*acvB*, *copB*, *cvaC*, *gaa*) (Table 1). Analyses of adaptive genes involved in responding to environmental changes allow detection of population variability that can be absent in neutral markers (35). These specific *X. fastidiosa* genes were initially part of a larger group of genes first identified through the literature to be important for the *X. fastidiosa* infection process and to have a high likelihood of being environmentally mediated (see literature references in Table 1). Subsequently, this subset of presumptive environmentally mediated genes was selected for MLSA-E because the genes had higher *dN/dS* ratios and so were likely under adaptive selection and would show variability that is lacking in neutral markers.

Phylogenetic results (Fig. 2) from this MLSA-E study show that the four currently proposed *X. fastidiosa* subspecies (75, 76) are recovered as reciprocally monophyletic clades. Also, the relationships between the subspecies agree with those found in recent multilocus studies using housekeeping genes (74, 76, 91). However, MLSA-E increases phylogenetic differentiation of *X. fastidiosa* isolates and identifies novel variations within the same subspecies compared to the results of schemes used in previous studies, including the recent MLST scheme (91). For example, enumeration of unique haplotypes among the 12 isolates analyzed in both this MLSA-E study and the MLST study found a larger number of haplotypes using MLSA-E ($n = 10$) than using MLST ($n = 8$), indicating that MLSA-E provides increased differentiation of even this small set of isolates. Phylogenetic results are supported by haplotype networks, which also show delineation by subspecies, with additional variability being shown within subspecies.

The *X. fastidiosa* MLST study found that little genetic variation was present within *X. fastidiosa* subsp. *fastidiosa* and *X. fastidiosa* subsp. *sandyi*, despite analyzing 86 and 21 isolates, respectively (91). While the MLSA-E study described here did not assess multiple *X. fastidiosa* subsp. *sandyi* isolates, it does identify increased variation among 30 *X. fastidiosa* subsp. *fastidiosa* isolates. MLSA-E results from *X. fastidiosa* grape isolates from different locations in the United States show genetic variability, as evidenced by phylogenetic analyses of the nine concatenated genes. Though isolates are more differentiated using ML analyses than BP analyses in this study, both analyses show variability within *X. fastidiosa* subsp. *fastidiosa*. Haplotype networks also show increased differentiation, and some geographic trends are evident. Isolates from Georgia and Florida form a network distinct from isolates from California and Texas (Fig. 4), establishing a difference among isolates from the East versus West Coasts of the United States. The excep-

tions to this are a single isolate from Georgia (CCPM1) and a single isolate from Florida (PD 92-8). Though it is difficult to know why these exceptions exist in the southeastern United States when they are genetically more similar (though not identical) to isolates from the western United States, it is possible that these isolates represent haplotypes that are adaptively intermediate between southeastern and western isolates or are a product of recombination, specifically if isolates were transported in infected plant material from a western location.

Though it has been less problematic to differentiate *X. fastidiosa* isolates within *X. fastidiosa* subsp. *multiplex* than within *X. fastidiosa* subsp. *fastidiosa*, results from MLSA-E of this expanded collection of isolates have also revealed a new finding within *X. fastidiosa* subsp. *multiplex*. MLSA-E indicates the presence of two large, distinct, reciprocally monophyletic clades within *X. fastidiosa* subsp. *multiplex*, while only one of these clades was identified in the recent MLST study (91). *X. fastidiosa* subsp. *multiplex* clade A is a possible novel clade composed of *X. fastidiosa* isolates from hosts for which little to no phylogenetic analysis has been conducted, including strains from the recently identified host blueberry (7), ragweed, and (potentially) elm. In the only other study to include isolates from elm and ragweed, 16S rRNA and *gyrB* gene trees grouped these isolates in the same clade, but they were paraphyletic with an almond isolate (Dixon) and a plum isolate, respectively, and so were not differentiated from members of *X. fastidiosa* subsp. *multiplex* clade B (69). One study (76) did analyze the almond isolate analyzed here (ALS6) and found it, along with another almond isolate, to be in a clade separate from clade B. However, these two almond isolates were the only isolates in this study with the recombinant alleles used to produce this phylogeny and so could be attracted away from the main clade and toward each other due to sequence variability from recombination. It is possible that this 2-isolate, single-host clade corresponds to the 10-isolate, four-host clade A in the present study, but more data are required to determine this.

X. fastidiosa subsp. *multiplex* clade B is the previously identified clade (91) which contains the same three subclades composed of isolates from almond (including isolate M12 used in both this and the MLST study [91]), plum, and oak, respectively, in the same general topology. However, isolates from five additional host species belonging to clade B are presented in the current study, which increases complexity in the topology and indicates new relationships. Additionally, in ML analysis, clade A and clade B form a polytomy with *X. fastidiosa* subsp. *fastidiosa*, showing that their relationships are unresolved. This indicates that clade A and clade B may not even be sister clades, further strengthening the argument that they are distinctly different clades. Different haplotype networks for each of the *X. fastidiosa* subsp. *multiplex* clades support these results. Similarly, the relationship of the single *X. fastidiosa* isolate from elm analyzed in this study varies depending on whether the ML or BP method is used. Though BP analysis places this isolate as the basal member in the new *X. fastidiosa* subsp. *multiplex* clade A, ML analysis places it by itself in a paraphyletic relationship with *X. fastidiosa* subsp. *multiplex* clade A, *X. fastidiosa* subsp. *multiplex* clade B, and *X. fastidiosa* subsp. *sandyi*/*X. fastidiosa* subsp. *fastidiosa*. Haplotype network results also place the *X. fastidiosa* isolate from elm in its own network containing no other isolates. Additional *X. fastidiosa* isolates from elm and related plant hosts are needed to clarify the relationship of the elm isolate to the other isolates.

Using dN/dS to predict sequence variability. Results from this study show that dN/dS can be used as a predictor of sequence variability to identify genes that may be suitable for differentiation of isolates in closely related lineages. Environmentally mediated genes with greater numbers of alleles (greater sequence variability) can be predicted by the dN/dS ratios of the most divergent *X. fastidiosa* strains. Also, these results strongly suggest that even though genes with dN/dS ratios of <1 are not technically defined to be under positive selection (39, 88), genes with larger (though <1) dN/dS ratios may still be more likely to be under positive selection than genes with smaller dN/dS ratios for some bacteria such as *X. fastidiosa*, and they can be adequate for discriminating among closely related isolates. Genes selected for MLSA-E had significantly greater dN/dS ratios than *X. fastidiosa* housekeeping genes, though none of their ratios were >1 , and they are also assumed to be under positive selection on the basis of their functions (Table 1). Furthermore, closely related *X. fastidiosa* isolates may not be expected to have any genes with dN/dS ratios of >1 . Other studies have indicated that dN/dS is insensitive to detecting adaptive selection in genes subject to low levels of diversifying selection, necessitating the usage of other methods for estimating adaptive selection for these genes (55, 64). However, housekeeping gene and environmentally mediated gene dN/dS variability in other organisms must be determined empirically in the same manner done here for *X. fastidiosa* prior to MLSA-E. It appears appropriate, based on the methods used in the present study, to define a dN/dS threshold relevant to the specific study organism using a variety of housekeeping and nonhousekeeping genes as an alternative to the rule of a dN/dS of >1 to identify genes that are potentially under positive selection.

dN/dS results also show that it is more appropriate to calculate dN/dS for the exact region of interest in a gene rather than for the entire gene, if it is to be used as a predictor of sequence variability. In this study, the Temecula/9a5c dN/dS of the entire gene was initially calculated for the MLSA-E genes and correlated to the number of alleles among the 28-isolate subset. Using a dN/dS cut-off value of 0.15, the gene *pilJ* was eliminated from further analyses because it had a dN/dS of 0.10 for the whole gene (Table 2). It was later determined that the gene dN/dS was not as closely correlated with the number of alleles as the dN/dS of the gene region of interest (Fig. 1). The dN/dS of the *pilJ* gene region used for MLSA-E was 0.20, much higher than the whole-gene dN/dS . If genes had initially been selected using the MLSA-E gene region dN/dS , *pilJ* would have been retained in the analyses. Therefore, the *pilJ* gene region analyzed in this study (see Table S2 in the supplemental material) may still be a good candidate for future MLSA-E studies.

Recombination in *X. fastidiosa*. Analysis of recombination and linkage disequilibrium in the nine final MLSA-E genes indicates evidence of intersubspecies recombination in the genes *acvB* and *rpff*. Recombinant genes can produce results inconsistent with bacterial evolutionary phylogenies (53), so it is important to consider recombination when constructing phylogenies. No single method for detecting recombination is sufficient for all circumstances (53), so multiple methods were applied in this study. Though not all methods detected recombination in *acvB* and *rpff* (Table 3), recombination was detected by the majority of methods.

However, it seems unlikely that recombination events in these genes had a significant effect on the MLSA-E results. First, *acvB*

and *rpff* have less of an effect on the phylogeny than the other MLSA-E genes. They are members of the low- dN/dS gene group and have less sequence variability and fewer alleles. The regions of the alignments within the recombination breakpoints, specifically, have limited (and not unique) variability. Also, recombination events are shared by either all *X. fastidiosa* subsp. *multiplex* isolates (*acvB*, parent isolate orange) or all *X. fastidiosa* subsp. *fastidiosa* isolates (*rpff*, parent isolate oleander). Because of this, the recombinant regions should not drastically alter the phylogenetic signal within subspecies, as isolates within these subspecies should be subjected to the same phylogenetic forces. The recombinant regions could affect the relationships between subspecies, but these effects should also be limited. This is due to the lesser variability mentioned previously and because any effects among the other nine genes should be diluted, particularly because the two intersubspecies recombination events occur between different sets of subspecies. This is supported by the SH test, which indicates that, despite removing these two genes to create an ML tree with no detectable recombination, the tree (though significantly different, as expected from removal of two genes) had a very low log likelihood difference compared to the nine-gene tree. Also, compared to the significant evidence of recombination seen in the *cysG* housekeeping gene used in the *X. fastidiosa* MLST study (91), the recombination in the MLSA-E genes is minimal. *cysG* has, by far, the greatest number of alleles ($n = 23$) of all the MLST genes found among 145 *X. fastidiosa* isolates (41). Though the Temecula/9a5c dN/dS for *cysG* is higher than that of most of the other MLST housekeeping genes ($dN/dS = 0.19$; Table 2), the number of alleles for this gene is nonetheless disproportionately large, probably due to recombination. For *acvB* and *rpff*, however, the number of alleles found in relation to the Temecula/9a5c dN/dS follows the trend line seen for all of the MLSA-E genes analyzed in this study (Fig. 1).

Recombination can play a key role in producing variation within populations (53). Previously, *X. fastidiosa* has been thought to be largely clonal, with limited evidence of recombination (74, 76). These results and a recent study (42) indicate that there may be more recombination among *X. fastidiosa* isolates than previously identified. Biologically, these recombination events in *acvB* and *rpff* may be influenced by host and/or environmental selection pressures. Since the identified recombinant regions in these genes are shared by all members of the recombinant subspecies, this may be indicative of a selective sweep through the population, where genetic diversity is eliminated at the specific allele (77). Particularly for *rpff*, responsible for the biosynthesis of diffusible signal factor (DSF), which regulates several virulence traits (8), the recombinant region may produce an adaptive advantage, causing this allele to sweep through populations of *X. fastidiosa* subsp. *fastidiosa*.

Specific sources of strain divergence in *X. fastidiosa*. It has previously been suggested that indels are a major source of strain divergence in *X. fastidiosa* and may be related to host specificity (21). The 21-bp indel present in the gene *xadA* may reveal such strain divergence. The insertion is present in all isolates of *X. fastidiosa* subsp. *pauca*, *X. fastidiosa* subsp. *multiplex*, and *X. fastidiosa* subsp. *sandyi* analyzed here. However, among *X. fastidiosa* subsp. *fastidiosa* isolates, the insertion is present only in isolates from Georgia and Florida (regardless of host). The region is deleted in all other *X. fastidiosa* subsp. *fastidiosa* isolates, and the larger *X. fastidiosa* subsp. *fastidiosa* haplotype network consists

only of isolates with the deletion (Fig. 4). *xadA* (Table 1) encodes an afimbrial adhesin membrane protein involved in biofilm formation and virulence in grapevines (25). As has been shown for other membrane proteins, *xadA* may be subject to positive selection due to increased exposure to the host and other environmental factors (27). Because ancestral *X. fastidiosa* strains likely lack this deletion (it is absent in the majority of isolates analyzed here), it may be an adaptive variation in non-southeastern United States *X. fastidiosa* subsp. *fastidiosa* isolates.

The two variant VNTR regions found in the gene *copB* are also of special interest. VNTR regions are able to mediate rapid phenotypic changes through changes in gene expression (30). Coding VNTRs can enhance bacterial virulence and play a role in adaptation to the environment (89). It is likely that the two kinds of coding VNTRs in this gene, which differ by only one amino acid change, represent a source of adaptive variation, particularly because *copB* is involved in resistance to copper (Table 1). Copper is often used to treat phytopathogens in agricultural crops, which triggers resistance mechanisms in these pathogens (82). Therefore, the number of each type of VNTR may play a role in responding to copper in the environment, which warrants additional exploration in future studies.

Conclusions. These analyses show that MLSA-E can be successfully applied to plant-pathogenic bacterial isolates for strain differentiation. Environmentally mediated genes provide a more variable phylogenetic signal than housekeeping genes; therefore, MLSA-E is an appropriate method for clarifying phylogenetic relationships between closely related bacterial strains with low genetic variability, such as *X. fastidiosa* strains. Additionally, this study shows that *dN/dS* may be a useful predictor of variable genes for MLSA-E or other studies and can guide careful selection of genes subject to a variety of environmental pressures. *A priori* selection of genes for MLSA-E must consider the environment where the bacterium thrives and the stresses that it encounters (e.g., xylem vessels/leafhopper foregut and strong shear forces, respectively, for *X. fastidiosa*). As complete genomes become increasingly available, it will be useful to scan genomes for regions with high *dN/dS* ratios to gain an understanding of the ranges for particular lineages and areas that may be useful for elucidating variability. This would provide a better idea of what may constitute a signal of positive selection for these specific organisms, rather than relying on finding regions with a *dN/dS* of >1 , which is difficult for closely related organisms. This MLSA-E study exposes clear genetic differences among *X. fastidiosa* subsp. *fastidiosa* isolates from the East versus West Coasts of the United States and characterizes blueberry isolates as belonging to a new clade within *X. fastidiosa* subsp. *multiplex*. Phenotypic characterization and virulence tests *in planta* of *X. fastidiosa* isolates used in this study are necessary to increase understanding of the genetic differentiation of bacterial populations found here.

ACKNOWLEDGMENTS

Funding was provided by the Auburn University College of Agriculture and the Department of Entomology and Plant Pathology. This publication made use of the *Xylella fastidiosa* MLST website (<http://pubmlst.org/X.fastidiosa/>), developed by Keith Jolley and sited at the University of Oxford (41). The development and maintenance of this site have been funded by the Wellcome Trust.

We thank the following people for sharing their *X. fastidiosa* isolate collections: Rodrigo Almeida (University of California, Berkeley, CA),

Mark Black (Texas A&M University), C. J. Chang (University of Georgia), Donald Hopkins (University of Florida), Jianchi Chen (USDA, Parlier, CA), and Harald Scherm (University of Georgia). We thank the following people for assistance collecting host plant tissue samples from which new *X. fastidiosa* isolates examined in this study were isolated: Phillip Brannen (University of Georgia), Jim Jacobi (Alabama Cooperative Extension System), and Greg Sheppard (Lumpkin County Cooperative Extension Service). Thanks go to Scott Santos (Auburn University) for advice regarding phylogenetic analyses and manuscript comments. Thanks to Stephanie Clugstone and Heather Hart for assistance with laboratory work.

REFERENCES

1. Akaike H. 1974. New look at statistical-model identification. *IEEE Trans. Automat. Contr.* 19:716–723.
2. Almeida RPP, et al. 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Appl. Environ. Microbiol.* 74:3690–3701.
3. Almeida RPP, Purcell AH. 2003. Biological traits of *Xylella fastidiosa* strains from grapes and almonds. *Appl. Environ. Microbiol.* 69:7447–7452.
4. Barends TR, et al. 2003. The sequence and crystal structure of the alpha-amino acid ester hydrolase from *Xanthomonas citri* define a new family of beta-lactam antibiotic acylases. *J. Biol. Chem.* 278:23076–23084.
5. Blum JK, Bommarius AS. 2010. Amino ester hydrolase from *Xanthomonas campestris* pv. *campestris*, ATCC 33913 for enzymatic synthesis of ampicillin. *J. Mol. Catal. B Enzym.* 67:21–28.
6. Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172:2665–2681.
7. Chang CJ, Donaldson R, Brannen P, Krewer G, Boland R. 2009. Bacterial leaf scorch, a new blueberry disease caused by *Xylella fastidiosa*. *HortScience* 44:413–417.
8. 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.
9. Chatterjee S, Newman KL, Lindow SE. 2008. Cell-to-cell signaling in *Xylella fastidiosa* suppresses movement and xylem vessel colonization in grape. *Mol. Plant Microbe Interact.* 21:1309–1315.
10. Chen J, Chang CJ, Jarret RL, Gawel N. 1992. Genetic variation among *Xylella fastidiosa* strains. *Phytopathology* 82:973–977.
11. Chen J, Civerolo E, Tubajika K, Livingston S, Higbee B. 2008. Hyper-variations of a protease-encoding gene, PD0218 (*pspB*), in *Xylella fastidiosa* strains causing almond leaf scorch and Pierce's disease in California. *Appl. Environ. Microbiol.* 74:3652–3657.
12. Chen J, Lamikanra O, Chang CJ, Hopkins DL. 1995. Randomly amplified polymorphic DNA analysis of *Xylella fastidiosa* Pierce's disease and oak leaf scorch pathotypes. *Appl. Environ. Microbiol.* 61:1688–1690.
13. Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9:1657–1659.
14. Coletta-Filho HD, Bittleston LS, Almeida RP. 2011. Spatial genetic structure of a vector-borne generalist pathogen. *Appl. Environ. Microbiol.* 77:2596–2601.
15. Crandall KA, Templeton AR, Sing CF. 1994. Intraspecific phylogenetics: problems and solutions, p 273–297. In Scotland RW, Siebert DJ, Williams DM (ed), *Models in phylogeny reconstruction*, 1st ed. Oxford University Press/Clarendon Press, Oxford, United Kingdom.
16. Davis MJ, French WJ, Schaad NW. 1981. Axenic culture of the bacteria associated with phony disease of peach and plum leaf scald. *Curr. Microbiol.* 6:309–314.
17. Davis MJ, Purcell AH, Thomson SV. 1980. Isolation media for the Pierce's disease bacterium. *Phytopathology* 70:425–429.
18. De La Fuente L, Burr TJ, Hoch HC. 2007. Mutations in type I and type IV pilus biosynthetic genes affect twitching motility rates in *Xylella fastidiosa*. *J. Bacteriol.* 189:7507–7510.
19. De La Fuente L, et al. 2007. Assessing adhesion forces of type I and type IV pili of *Xylella fastidiosa* bacteria by use of a microfluidic flow chamber. *Appl. Environ. Microbiol.* 73:2690–2696.
20. de Souza AA, et al. 2003. Analysis of gene expression in two growth states of *Xylella fastidiosa* and its relationship with pathogenicity. *Mol. Plant Microbe Interact.* 16:867–875.
21. Doddapaneni H, Yao JQ, 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.

22. Doyle J, Doyle JL. 1987. Genomic plant DNA preparation from fresh tissue—CTAB method. *Phytochem. Bull.* 19:11–15.
23. Drummond AJ, et al. 2011. Geneious, version 5.4. Geneious, Auckland, New Zealand.
24. Feil H, Feil WS, Detter JC, Purcell AH, Lindow SE. 2003. Site-directed disruption of the *fimA* and *fimF* fimbrial genes of *Xylella fastidiosa*. *Phytopathology* 93:675–682.
25. Feil H, Feil WS, Lindow SE. 2007. Contribution of fimbrial and afimbrial adhesins of *Xylella fastidiosa* to attachment to surfaces and virulence to grape. *Phytopathology* 97:318–324.
26. Felsenstein J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, MA.
27. Fitzpatrick DA, McInerney JO. 2005. Evidence of positive Darwinian selection in *Omp85*, a highly conserved bacterial outer membrane protein essential for cell viability. *J. Mol. Evol.* 60:268–273.
28. Foley SL, et al. 2006. Comparison of subtyping methods for differentiating *Salmonella enterica* serovar Typhimurium isolates obtained from food animal sources. *J. Clin. Microbiol.* 44:3569–3577.
29. Francis M, Civerolo EL, Bruening G. 2008. Improved bioassay of *Xylella fastidiosa* using *Nicotiana tabacum* cultivar SR1. *Plant Dis.* 92:14–20.
30. Gemayel R, Vences MD, Legendre M, Verstrepen KJ. 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* 44:445–477.
31. Gomez-Diaz E. 2009. Linking questions to practices in the study of microbial pathogens: sampling bias and typing methods. *Infect. Genet. Evol.* 9:1418–1423.
32. Hendson M, et al. 2001. Genetic diversity of Pierce's disease strains and other pathotypes of *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 67:895–903.
33. Hernandez-Martinez R, de la Cerda KA, Costa HS, Cooksey DA, Wong FP. 2007. Phylogenetic relationships of *Xylella fastidiosa* strains isolated from landscape ornamentals in southern California. *Phytopathology* 97:857–864.
34. Hernandez-Martinez R, Dumenyo KC, Cooksey DA. 2006. Site-directed mutagenesis of *acvB* gene in a Pierce's disease strain of *Xylella fastidiosa*. *Phytopathology* 96:S47–S48.
35. Hoffmann AA, Willi Y. 2008. Detecting genetic responses to environmental change. *Nat. Rev. Genet.* 9:421–432.
36. Hopkins D. 2001. Gram-negative bacteria: *Xylella fastidiosa*, p 201–213. In Schaad NW, Jones JB, Chun W (ed), Laboratory guide for identification of plant pathogenic bacteria, 3rd ed. APS Press, St. Paul, MN.
37. Hopkins DL. 1985. Physiological and pathological characteristics of virulent and avirulent strains of the bacterium that causes Pierce's disease of grapevine. *Phytopathology* 75:713–717.
38. Hopkins DL, Purcell AH. 2002. *Xylella fastidiosa*: cause of Pierce's disease of grapevine and other emergent diseases. *Plant Dis.* 86:1056–1066.
39. Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170.
40. Janse JD, Obradovic A. 2010. *Xylella fastidiosa*: its biology, diagnosis, control and risks. *J. Plant Pathol.* 92:S35–S48.
41. Jolley KA, Chan MS, Maiden MC. 2004. mlstdbNet-distributed multilocus sequence typing (MLST) databases. *BMC Bioinformatics* 5:86.
42. Kung SH, Almeida RP. 2011. Natural competence and recombination in the plant pathogen *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 77:5278–5284.
43. Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
44. Lemee L, et al. 2005. Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*. *Microbiology* 151:3171–3180.
45. Li WB, et al. 2002. Citrus and coffee strains of *Xylella fastidiosa* induce Pierce's disease in grapevine. *Plant Dis.* 86:1206–1210.
46. Li Y, et al. 2007. Type I and type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell-cell aggregation. *Microbiology* 153:719–726.
47. Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
48. Lin H, et al. 2005. Multilocus simple sequence repeat markers for differentiating strains and evaluating genetic diversity of *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 71:4888–4892.
49. Liu FY, et al. 2011. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl. Environ. Microbiol.* 77:1946–1956.
50. Liu K, Knabel SJ, Dudley EG. 2009. *rhs* genes are potential markers for multilocus sequence typing of *Escherichia coli* O157:H7 strains. *Appl. Environ. Microbiol.* 75:5853–5862.
51. Lopes JRS, Daugherty MP, Almeida RPP. 2010. Strain origin drives virulence and persistence of *Xylella fastidiosa* in alfalfa. *Plant Pathol.* 59:963–971.
52. Martin DP, et al. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463.
53. Martin DP, Lemey P, Posada D. 2011. Analysing recombination in nucleotide sequences. *Mol. Ecol. Resour.* 11:943–955.
54. Martinati JC, Pacheco FTH, de Miranda VFO, Tsai SM. 2005. Phylogenetic relationships of *Xylella fastidiosa* strains based on 16S-23S rDNA sequences. *Curr. Microbiol.* 50:190–195.
55. McClellan DA, et al. 2005. Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Mol. Biol. Evol.* 22:437–455.
56. Mehta A, Rosato YB. 2001. Phylogenetic relationships of *Xylella fastidiosa* strains from different hosts, based on 16S rDNA and 16S-23S intergenic spacer sequences. *Int. J. Syst. Evol. Microbiol.* 51:311–318.
57. Meng Y, et al. 2005. Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. *J. Bacteriol.* 187:5560–5567.
58. Minsavage GV, Thompson CM, Hopkins DL, Leite RMVBC, Stall RE. 1994. Development of a polymerase chain-reaction protocol for detection of *Xylella fastidiosa* in plant tissue. *Phytopathology* 84:456–461.
59. Morano LD, et al. 2008. Initial genetic analysis of *Xylella fastidiosa* in Texas. *Curr. Microbiol.* 56:346–351.
60. Newman KL, Almeida RPP, Purcell AH, Lindow SE. 2004. Cell-cell signaling controls *Xylella fastidiosa* interactions with both insects and plants. *Proc. Natl. Acad. Sci. U. S. A.* 101:1737–1742.
61. Nunney L, et al. 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.
62. Pashalidis S, et al. 2005. Whole-genome expression profiling of *Xylella fastidiosa* in response to growth on glucose. *OMICS* 9:77–90.
63. Perez-Donoso AG, et al. 2010. Cell wall-degrading enzymes enlarge the pore size of intervessel pit membranes in healthy and *Xylella fastidiosa*-infected grapevines. *Plant Physiol.* 152:1748–1759.
64. Perez-Losada M, et al. 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect. Genet. Evol.* 6:97–112.
65. Pooler MR, Hartung JS. 1995. Genetic relationships among strains of *Xylella fastidiosa* from RAPD-PCR data. *Curr. Microbiol.* 31:134–137.
66. Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
67. Randall JJ, et al. 2009. Genetic analysis of a novel *Xylella fastidiosa* subspecies found in the southwestern United States. *Appl. Environ. Microbiol.* 75:5631–5638.
68. Rodrigues CM, et al. 2008. Copper resistance of biofilm cells of the plant pathogen *Xylella fastidiosa*. *Appl. Microbiol. Biotechnol.* 77:1145–1157.
69. Rodrigues JLM, Silva-Stenico ME, Gomes JE, Lopes JRS, Tsai SM. 2003. Detection and diversity assessment of *Xylella fastidiosa* in field-collected plant and insect samples by using 16S rRNA and *gyrB* sequences. *Appl. Environ. Microbiol.* 69:4249–4255.
70. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
71. Roper MC, Greve LC, Warren JG, Lavavitch JM, Kirkpatrick BC. 2007. *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Mol. Plant Microbe Interact.* 20:411–419.
72. Rozas J, Gullaud M, Blandin G, Aguade M. 2001. DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* 158:1147–1155.
73. Sanderson MJ, Shaffer HB. 2002. Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* 33:49–72.
74. 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.
75. Schaad NW, Postnikova E, Lacy G, Fatmi M, Chang CJ. 2004. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. [correction] *fastidiosa* [correc-

- tion] subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. Syst. Appl. Microbiol. 27:290–300.
76. 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.
 77. Shapiro BJ, David LA, Friedman J, Alm EJ. 2009. Looking for Darwin's footprints in the microbial world. Trends Microbiol. 17:196–204.
 78. Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.
 79. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
 80. Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758–771.
 81. Tankouo-Sandjong B, et al. 2007. MLST-v, multilocus sequence typing based on virulence genes, for molecular typing of *Salmonella enterica* subsp. *enterica* serovars. J. Microbiol. Methods 69:23–36.
 82. Teixeira EC, de Oliveira JCF, Novo MTM, Bertolini MC. 2008. The copper resistance operon *copAB* from *Xanthomonas axonopodis* pathovar *citri*: gene inactivation results in copper sensitivity. Microbiology 154:402–412.
 83. Templeton AR. 2006. Genetic drift in large populations and coalescence, p 118–168. In Templeton AR (ed), Population genetics and microevolutionary theory. Wiley-Liss, Hoboken, NJ.
 84. Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619–633.
 85. Van Sluys MA, et al. 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.
 86. Varani AM, et al. 2009. Common genes and genomic breaks: a detailed case study of the *Xylella fastidiosa* genome backbone and evolutionary insights, p 113–133. In Jackson RW (ed), Plant pathogenic bacteria: genomics and molecular biology. Caister Academic, Norfolk, United Kingdom.
 87. Wickert E, Machado MA, Lemos EGM. 2007. Evaluation of the genetic diversity of *Xylella fastidiosa* strains from citrus and coffee hosts by single-nucleotide polymorphism markers. Phytopathology 97:1543–1549.
 88. Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15:496–503.
 89. Yousef Mohamad K, Rekiki A, Myers G, Bavoil PM, Rodolakis A. 2008. Identification and characterisation of coding tandem repeat variants in *incA* gene of *Chlamydophila pecorum*. Vet. Res. 39:56.
 90. Yousef Mohamad K, et al. 2008. Preliminary phylogenetic identification of virulent *Chlamydophila pecorum* strains. Infect. Genet. Evol. 8:764–771.
 91. Yuan X, et al. 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. Phytopathology 100:601–611.
 92. Zhang W, Jayarao BM, Knabel SJ. 2004. Multi-virulence-locus sequence typing of *Listeria monocytogenes*. Appl. Environ. Microbiol. 70:913–920.