AEM Accepts, published online ahead of print on 22 December 2011
Appl. Environ. Microbiol. doi:10.1128/AEM.06679-11
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

1 **Differentiation of *Xylella fastidiosa* strains via multi-locus sequence analysis of environmentally-**
2 **mediated genes (MLSA-E) †**
3
4 Jennifer K. Parker[1], Justin C. Havird[2], and Leonardo De La Fuente[1]*
5
6 [1]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL
7
8 [2]Department of Biological Sciences, Auburn University, Auburn, AL
9
10 *Corresponding author. Mailing address: Department of Entomology and Plant Pathology, 209 Rouse
11 Life Sciences Bldg., Auburn University, Auburn, AL 36849. Phone: (334) 844-2582. Fax: (334) 844-1947.
12 Email: lzd0005@auburn.edu.
13
14 †Supplemental material for this article may be found at http://aem.asm.org/.
15
16 Running Title: MLSA-E of *Xylella fastidiosa*
17

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

18 **ABSTRACT**

19

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

40

41 Keywords: *Xylella fastidiosa*, MLSA, environment, phylogenetics, positive selection

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

42

43 **INTRODUCTION**

44

45 *Xylella fastidiosa* is a gram-negative plant pathogenic bacterium that is adapted to live inside the xylem
46 vessels of plant hosts and the foregut of leafhopper insect vectors (8). Biofilm formed inside the xylem
47 vessels is believed to play a key role in disease development by obstructing water passage from the root
48 to aerial parts of the plant (8). The bacterium infects several economically important agricultural crops
49 including grape, citrus, almond, and peach as well as a variety of landscape trees and plants (38).
50 Despite this broad host range, isolates of *X. fastidiosa* from different plant hosts are genetically very
51 similar. Comparative genomic analyses between the most divergent *X. fastidiosa* complete genome
52 strains (the Pierce's disease of grapevine type strain Temecula and the citrus variegated chlorosis type
53 strain 9a5c) have shown that 98% of their genes are shared, and the average amino acid identity is 96%
54 (87). Taxonomic analyses have grouped *X. fastidiosa* strains into several subspecies based on DNA-DNA
55 hybridization and phylogenetic studies: subsp. *fastidiosa,* subsp. *sandyi*, subsp. *multiplex*, and subsp.
56 *pauca* (77, 78). Though isolates from these subspecies are genetically too similar for further taxonomic
57 classification, studies on the biological traits of isolates from the same subspecies have indicated
58 differences in infectivity (2, 3, 29, 37, 45, 51). Because of this, efforts are ongoing to further characterize
59 the diversity seen among *X. fastidiosa* isolates within the proposed subspecies (91).

60

61 Diverse molecular techniques have been applied to genetically characterize *X. fastidiosa* isolates
62 including restriction fragment length polymorphism (RFLP) (10), random amplification of polymorphic
63 DNA (RAPD) (3, 12, 32, 67), microsatellite/simple sequence repeat (SSR) analysis (2, 35, 48, 69), and 16-
64 23S rRNA internal transcribed spacer (16S-23S ITS) sequencing (32, 54, 56, 69). Results from strain
65 typing, phylogenetic analyses, and other data comparisons in the aforementioned studies have shown

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

66 broad host-based genetic differences between strains; however, phylogenetic results on differentiation
67 of strains from within the same *X. fastidiosa* subspecies and those isolated from the same host plant
68 species are still inconclusive. More recently, a multi-locus sequence typing (MLST) scheme has been
69 implemented (2, 76, 78, 91) to provide a standardized means of interpreting phylogenetic data from
70 many isolates of *X. fastidiosa*. This MLST scheme relies on sequence analysis of 7 housekeeping genes,
71 which are required for basic cellular maintenance. Housekeeping genes are usually relatively conserved,
72 making them good candidates for evolutionary phylogenetic analyses of distantly related groups. One
73 drawback to using housekeeping genes is that they may not provide enough variability to differentiate
74 closely related isolates. The often sequenced *gyrB* gene has 97.5-100% sequence similarity among *X.*
75 *fastidiosa* isolates (71), exemplifying the limited variation found among *X. fastidiosa* housekeeping
76 genes.
77
78 Similar problems detecting genetic variability, despite differences in infectivity characteristics, have
79 frequently occurred in analyses of mammalian pathogen housekeeping genes. Researchers have
80 resorted to analyses of alternate genes with greater variability to improve genetic differentiation of
81 pathogen isolates and provide insight into recent evolutionary processes, including virulence genes
82 (*Clostridium difficile*, *Chlamydophila pecorum*, *Salmonella enterica* subsp. *enterica*, *Listeria*
83 *monocytogenes*) (44, 60, 83, 92) and other genes related to pathogen-host interactions (*S. enterica*
84 serovar Typhimurium, *Escherichia coli*, *S. enterica* subsp. *enterica*, *C. pecorum*) (28, 49, 50, 60). The
85 recent *X. fastidiosa* MLST publication also included a non-housekeeping gene related to motility (*pilU*) to
86 increase phylogenetic resolution, and it provides most of the variation among non-recombinant *X.*
87 *fastidiosa* subsp. *fastidiosa* isolates, though it is excluded from the conventional MLST scheme (91). Such
88 genes are likely subject to adaptive evolutionary pressure (positive selection), making them more
89 suitable for population genetics study than for determining deep evolutionary relationships.

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

90 Additionally, not all housekeeping genes are suitable for evolutionary phylogenies. A comparison of
91 MLST schemes for 17 human pathogens showed that, though housekeeping genes were under purifying
92 selection on average, 13 of the species analyzed had evidence of punctual adaptive evolution within one
93 or more loci (66). However, it has been argued that multi-locus sequence analysis (MLSA) of genes
94 subject to positive selection can still provide accurate predictions of short-term evolution as long as the
95 major clades identified with phylogenetic analysis of housekeeping genes are maintained in the analysis
96 of the positively selected genes (31, 50).
97
98 An alternative for discriminating between closely-related *X. fastidiosa* isolates is to perform MLSA of
99 genes influenced by environmental factors, termed environmentally-mediated genes. Environmentally-
100 mediated genes are involved in adapting to environmental changes and are usually subject to positive
101 selection pressure (35), so they should have greater sequence variability than conserved housekeeping
102 genes under stabilizing selection pressure. Therefore, sequence analysis of environmentally-mediated
103 genes should provide increased phylogenetic differentiation and be more likely to yield genetic
104 differences between recently diverged isolates, comparable to results from human bacterial pathogen
105 MLST studies using virulence and other positively selected genes (44, 50, 83, 92). However, no previous
106 studies have explored the diversity of virulence genes in a plant pathogenic bacterium. Because *X.*
107 *fastidiosa* is likely subject to strong selection pressure (78), it should have some positively selected,
108 environmentally-mediated genes with higher sequence variation. Previous studies have elucidated the
109 essential roles of many specific adaptive genes involved in the proliferation of this xylem-limited
110 pathogen in the host (e.g. adhesion, biofilm formation, motility, cell signaling, plant cell wall
111 degradation, metal homeostasis, and toxin production genes, reviewed in Varani *et al.* 2009) (88). It is
112 hypothesized that analyzing these genes in *X. fastidiosa* will increase differentiation of genetically similar
113 isolates, in particularly isolates infecting the same plant host.

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

114

115 Here, multi-locus sequence analysis of environmentally-mediated genes (MLSA-E) was applied to

116 identify *X. fastidiosa* isolate relationships. MLSA-E was used for genes related to processes important for

117 establishing *X. fastidiosa* infections and colonization of the insect vector such as surface attachment,

118 biofilm formation, virulence, and nutrient transport and utilization. These types of genes may be more

119 relevant to host-based genetic variability. Genes of interest were identified from previous studies and

120 based on dN/dS, the ratio of the rate of non-synonymous substitutions to the rate of synonymous

121 substitutions, an indication of positive selection. Target-specific PCR primers were designed from

122 available *X. fastidiosa* whole genome sequences. Target genes were PCR-amplified and sequenced from

123 *X. fastidiosa* isolates from a variety of host plants from several geographic regions in the US.

124 Phylogenetic analyses of the resulting sequence information show host-based and geographic origin-

125 based genetic relationships. Discovering the genetic relationships between *X. fastidiosa* isolates will

126 provide new insights into the epidemiology of populations of *X. fastidiosa*, which will allow for improved

127 disease management in a variety of economically important agricultural crops.

128

129 **MATERIALS AND METHODS**

130

131 **Isolation of *X. fastidiosa*.** Few *X. fastidiosa* isolates from the southeastern US, with the exception of

132 isolates from Florida, have been collected and genetically analyzed. To improve representation of *X.*

133 *fastidiosa* isolates from the southeastern US, isolates were collected from Georgia and Alabama. Leaves

134 showing *X. fastidiosa* infection symptoms of marginal leaf scorching were collected from host plants in

135 these states and stored at 4°C. Within 24 hours, leaves were processed according to a modified bacterial

136 isolation protocol (36). Briefly, each leaf was surface-sterilized with 70% ethanol (2 min), followed by

137 1.2% sodium hypochlorite (5 min), and then rinsed 3 times with sterile deionized water. The lower half

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

138 of the main vein and the petiole (with tip trimmed and discarded) were removed with a sterile blade.
139 The tissue was cut along the length of the vein, sliced into small pieces, and added to 3 ml PD2 (17)
140 liquid media. Media with leaf tissue was incubated with shaking at 28°C for 2 hrs. PD2 samples of 100 µl
141 were spread-plated on PW agar plates (16) and incubated at 28°C for up to 1 month, until presumptive
142 *X. fastidiosa* colonies appeared. Colonies were re-streaked for isolation, confirmed by PCR (58), and
143 stored in PD2 with 20% glycerol at -80°C.
144
145 ***X. fastidiosa* isolate collection and DNA extraction.** *X. fastidiosa* isolates (n = 49) were isolated as
146 discussed above (n = 8) or acquired from other laboratories (n = 41). Isolates were originally collected
147 from 33 sampling sites in the following US states: Alabama (n = 1), California (n = 13), Florida (n = 11),
148 Georgia (n = 18), and Texas (n = 6) (supplemental material, Supplemental Fig. 1, Supplemental Table 1).
149 Isolates were cultured 1-2 weeks on PW agar (16). Bacteria were then scraped from the agar plates,
150 resuspended in 200 µl molecular grade water, and DNA extracted using a modified CTAB protocol (22).
151 Sequences for isolates with complete genomes deposited in GenBank (n = 5) were also included in this
152 study (supplemental material, Supplemental Table 1, crosses). Of the combined isolates analyzed here (n
153 = 54), the majority were isolated from grape (n = 28), followed by blueberry (n = 7), almond (n = 3), oak
154 (n = 3), lupine (n = 2), sycamore (n = 2), and elderberry, elm, oleander, orange, plum, ragweed, redbud,
155 sumpweed, and sunflower (n =1 for each).
156
157 **MLSA-E gene selection.** Genes with the following characteristics were initially chosen for MLSA (Table
158 1): 1) experimental proof from previous studies regarding their role in the infection process and
159 indicating the possibility of being environmentally-mediated, and 2) high probability of sequence
160 variability based on comparisons of the 5 *X. fastidiosa* complete genome sequences deposited in
161 GenBank (Temecula, Accession # AE009442; GB514, Accession # CP002165; M23, Accession # CP001011;

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

162 M12, Accession # CP000941; 9a5c, Accession # AE003849). Genes related to processes important for
163 establishing *X. fastidiosa* infections such as surface attachment, biofilm formation, virulence, and
164 nutrient transport/utilization were initially selected *a priori* from the literature, with preference given to
165 genes in each category with the most well-elucidated roles. A BLAST search was conducted to identify
166 the candidate genes with the greatest sequence variation among the 5 complete genome sequences.
167 Ten genes were initially selected representing several categories of potential environmental interactions
168 (Table 1, supplemental material, Supplemental Table 2). General sequence variability for each gene was
169 assessed by quantifying the number of unique haplotypes based on preliminary results using 28 *X.*
170 *fastidiosa* isolates and these 10 genes.
171
172 The ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions (dN/dS)
173 was calculated between the Pierce's disease of grapevine type strain (*X. fastidiosa* strain Temecula) and
174 the citrus variegated chlorosis type strain (*X. fastidiosa* strain 9a5c) for the 10 MLSA-E genes using
175 DnaSP v5 (47). dN/dS is an indicator of selection pressure and was calculated to determine if this ratio
176 could be used to predict additional MLSA-E genes with greater sequence variability and, if so, what
177 range of dN/dS values would be expected for high-variability sequences. Because *X. fastidiosa* isolates
178 are genetically very similar, and Temecula and 9a5c have the most dissimilar complete genome
179 sequences available, genes with a higher Temecula:9a5c dN/dS may have greater sequence variability
180 among more closely related isolates. To test this hypothesis, Temecula:9a5c dN/dS for the 10 initial
181 MLSA-E genes were plotted against the number of alleles found among the 28 *X. fastidiosa* isolates
182 initially analyzed. Results from this (see below), led to selection of three additional genes for the final
183 MLSA-E (Table 1, with asterisks) based on the following new criteria: 1) possibility of being
184 environmentally-mediated based on previous studies, and 2) Temecula:9a5c dN/dS > 0.15, an indication
185 of positive selection pressure. Finally, Temecula:9a5c dN/dS for all 13 genes regions were plotted

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

186 against the number of alleles found among the 28 initially analyzed *X. fastidiosa* isolates to determine
187 the correlation between dN/dS and sequence variability from the larger dataset. Due to lower sequence
188 variability and number of alleles as predicted by Temecula:9a5c dN/dS, 4 of the initial genes were
189 eliminated from the final MLSA-E of all 54 isolates (supplemental material, Supplemental Table 2).
190

191 **Detection of positive selection.** dN/dS were calculated for the set of housekeeping gene regions
192 described in a previous MLST study of *X. fastidiosa* (91) to determine if their dN/dS were significantly
193 different from those used in the final MLSA-E. All gene regions (of both the final MLSA-E and MLST
194 genes) were designated as either housekeeping or non-housekeeping genes based on the literature.
195 *gyrB*, an additional housekeeping gene previously used for phylogenetic analysis of *X. fastidiosa* (61, 71),
196 was also included. dN/dS from the housekeeping genes were compared with dN/dS from the non-
197 housekeeping genes using a Wilcoxon rank-sum test ($\alpha = 0.05$) in Statistix v. 8.0 (Analytical Software).
198

199 **PCR amplification and DNA sequencing.** PCR primers (Table 1,supplemental material, Supplemental
200 Table 2) were designed for the selected genes using PrimerQuest (Integrated DNA Technologies,
201 Coralville, IA) with the exception of the *malF* primers (91). Isolate DNA was amplified by PCR on an
202 S1000 Thermocycler (BioRad Laboratories, Hercules, CA) in reactions (50 µl) containing the following
203 components: 1.25 U Perfect*Taq* DNA Polymerase (5Prime, Gaithersburg, MD), 1X Perfect*Taq* PCR Buffer,
204 0.2 mM dNTPs, 0.2 µM each primer, and 1 µl DNA template. Cycling parameters for all genes were: 94°C
205 for 3 min, followed by 25 cycles of 94°C for 30s, 62°C for 30s, and 72°C for 45s, with a final step of 72°C
206 for 10 min. PCR products were verified by gel electrophoresis. DNA sequencing was conducted on an ABI
207 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA) by Lucigen Corporation (Middleton, WI).
208

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

209 **Phylogenetic analyses.** Geneious v5.4 (23) was used to trim low quality sequence end reads, create
210 consensus gene sequences, translate putative protein sequences, and align sequences (using the
211 Geneious ClustalW (43) plugin). For the *pglA* gene, the alignment was trimmed to exclude the noncoding
212 region at the beginning of the PCR amplicon and the region after the frameshift mutation that has been
213 described in *X. fastidiosa* subsp. *pauca* (87). For the *copB* gene, 1 bp was deleted from the GB514 isolate
214 sequence in GenBank (Locus Tag XFLM_05775) because it was suspected to be a sequencing error (see
215 below).
216
217 To reconstruct the evolutionary relationships of the *X. fastidiosa* isolates, phylogenies were generated
218 for each individual gene alignment and for the concatenated sequence alignment (supplemental
219 material, Supplemental File 1). Bayesian phylogenetic (BP) analyses were conducted using MrBayes v
220 3.1.2 (72) with 2 chains of 2 million generations each. Partitions were created for each codon position of
221 each gene and models of nucleotide substitution were applied to each partition as determined by
222 Modeltest 3.7 (68) according to the Akaike Information Criterion (1). Trees were sampled every 200
223 generations, and the first 5,000 trees (10%) were discarded as burnin for each chain prior to generating
224 the extended majority rule consensus tree. Maximum likelihood (ML) analyses were performed using
225 RAxML v. 7.0.3 (81) using the rapid BS algorithm (82) with 1000 bootstraps. For ML analyses, each codon
226 position was treated as a separate partition and the GTR + I + Γ model was implemented for each
227 position using the GTRGAMMAI option. Trees were initially rooted with *Xanthomonas campestris* pv.
228 *campestris* strain ATCC 33913 as the outgroup, but final trees were rooted with *X. fastidiosa* isolate 9a5c
229 (orange) as the outgroup based on initial results.
230
231 To compare topologies between the single gene ML trees and the concatenated ML tree, log likelihood
232 differences (ΔLn L) were computed and Shimodaira-Hasegawa (SH) tests (80) were performed between

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

233 each gene tree and the concatenated gene tree using the –f h option in RAxML. The nine MLSA-E genes
234 were also divided into groups of "high" dN/dS genes and groups of "low" dN/dS genes based on the
235 Temecula:9a5c MLSA-E gene region dN/dS. ML phylogenies generated from data sets with increasing
236 numbers of "high" or "low" dN/dS genes were compared to the phylogeny generated from the complete
237 nine-gene data set using SH tests to determine the minimum number of genes of each type needed to
238 obtain a tree not significantly worse than the nine-gene tree. An SH test was also used to determine if
239 genes with evidence of recombination (see methods below) have a significant effect on the nine-gene
240 ML tree. This was done by removing potentially recombinant genes from the alignment, producing a ML
241 tree with the remaining genes, and using the SH test to compare it to the original tree.
242
243 **Haplotype networks.** To visualize relationships between haplotypes, networks were created based on
244 the nine-gene alignment (supplemental material, Supplemental File 1) using TCS version 1.21 (13), which
245 uses the previously described cladogram estimator algorithm (86). The default settings were used (i.e.,
246 branches represent 95% parsimoniously plausible connections between haplotypes). Loops
247 (reticulations) in the network, which represent ambiguous and implausible connections, were broken
248 according to previous recommendations (15, 85).
249
250 **Analysis of sequence diversity, recombination, and linkage disequilibrium.** Sequence diversity of the
251 nine final single-gene alignments was described by calculating nucleotide diversity per site (π) using
252 DnaSP v5 (47). Single-gene alignments were also tested for evidence of recombination using both RDP4
253 Beta 4.11 (52) and PHI test (6). RDP tests the alignment with a suite of 9 non-parametric recombination
254 detection methods that provide the number, breakpoints, and statistical probability of recombination
255 events, while PHI test scans the alignment for overall evidence of recombination. Significance of all

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

256 recombination events was tested at *P*=0.05. Alignments were also analyzed for linkage disequilibrium
257 using the ZZ statistic (74) as implemented in DnaSP v5 (47).
258
259 ***X. fastidiosa* MLST scheme comparative analyses.** To compare the *X. fastidiosa* MLST scheme (76, 91)
260 with the MLSA-E scheme developed here, haplotype diversity was investigated using the 7 MLST
261 housekeeping genes versus the 9 MLSA-E genes. The online *X. fastidiosa* MLST database was compared
262 against the *X. fastidiosa* isolates used in this study, and 12 isolates analyzed in common were identified:
263 9a5c, ALS6, Conn Creek, EB92-3, Georgia Plum, Hopland, L95-1, L95-2, M12, M23, Preston, and
264 Temecula (supplemental material, Supplemental Table 1). Sequence data for the MLST genes were
265 downloaded from the website, concatenated alignments were created, and number of haplotypes was
266 determined.
267
268 **RESULTS**
269
270 **Gene selection for MLSA-E.** A positive correlation was observed between Temecula:9a5c dN/dS of the
271 initial 10 MLSA-E genes and the number of alleles among the 28 XF isolates initially analyzed. Based on
272 this, genes with dN/dS > 0.15 were postulated to have sufficient variability for MLSA-E because they
273 were predicted by the trendline to have a greater number of alleles and the potential to be under
274 positive selection. To select additional genes with high dN/dS (dN/dS > 0.15), the Temecula:9a5c dN/dS
275 was calculated for numerous potential environmentally-mediated genes identified in the literature (data
276 not shown). This led to selection of three additional MLSA-E genes with the highest dN/dS of the genes
277 surveyed. A significant positive correlation (R^2 = 0.66, *P* = 0.0007 for regression analysis) was found
278 between Temecula:9a5c dN/dS for the 13 target MLSA-E gene regions sequenced and the number of
279 alleles identified among the 28 *X. fastidiosa* isolates initially analyzed (Fig. 1). Based on these results,

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

280 nine genes with dN/dS > 0.15 were chosen for final MLSA-E of the 54 *X. fastidiosa* isolates. dN/dS of the
281 non-housekeeping genes in this study (mean = 0.24) were significantly higher than housekeeping genes
282 used in the MLST *X. fastidiosa* study (91) and this study (mean = 0.12) (Wilcoxon rank-sum test, two-
283 tailed $P = 0.0013$) (Table 2).
284
285 **MLSA-E gene sequences.** Sequencing results indicate specific characteristics for several genes analyzed
286 which represent novel findings. In agreement with previous reports (87), the *pglA* frameshift mutation
287 seen in *X. fastidiosa* subsp. *pauca* isolates is absent in the non-subsp. *pauca* isolates analyzed in this
288 study. There is a 1 bp insertion in the GB514 strain *copB* sequence in GenBank that likely represents a
289 sequencing error. The insertion produces a stop codon that results in truncation of the gene compared
290 to its translation from other complete genome sequences and *X. fastidiosa* isolates from grape in this
291 study (including another isolate from Texas). *copB* functions in copper homeostasis in *X. fastidiosa* (70),
292 and inactivation of the homologous *copAB* operon in *Xanthomonas axonopodis* pv. *citri* rendered the
293 bacterium incapable of growing in medium containing copper (84), so this gene is probably necessary for
294 *X. fastidiosa* survival. Also, *copB* contains a region with a 15 base pair (minisatellite) coding variable
295 number tandem repeat (VNTR). When oriented in the coding direction, a repeat of 5'-
296 ACACCCAGATGGATC-3' occurs (described here for the first time), followed by a repeat which is a
297 previously described variation of this sequence: 5'-ACACAGGGATGGATC-3' (48). The two substitutions
298 that account for the difference in the two repeats represent a nonsynonymous mutation whereby
299 glutamine (CAG) vs. glycine (GGG) is translated in the first vs. second repeat, respectively. It is also
300 notable that *xadA* contains a 21 bp indel. The deletion is present in all unique subsp. *fastidiosa* isolate
301 groups except for the haplotype groups GR_GAFL (9 Georgia grape and 2 Florida grape isolates) and
302 EB_LUPINE (1 Florida elderberry and 1 Florida lupine isolate). The insertion is present in all other isolates
303 examined here.

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

304

305 **Haplotype diversity.** In total, 6026 bp from nine genes were sequenced or obtained from public

306 databases for 54 *X. fastidiosa* isolates. Analysis of the final concatenated nine-gene alignment for all 54

307 *X. fastidiosa* isolates (49 isolates plus data from 4 full genome sequences) revealed 29 unique

308 haplotypes (supplemental material, Supplemental Table 1). Each haplotype was represented by 1 to 11

309 isolates. The unique haplotypes demonstrate host-based differences between isolates. Grape and

310 blueberry isolates are the largest groups from a single host presented here. 10 unique haplotypes were

311 identified from 28 grape isolates, and 3 unique haplotypes were identified from 7 blueberry isolates.

312 This confirms that genetic variability is present among isolates obtained from the same host plant

313 species.

314

315 Of the 10 grape isolates collected from Georgia, all had identical MLSA-E genes except for CCPM1, which

316 was acquired from plant tissue collected in the same county on the same day as 4 of the other isolates,

317 but from a different cultivar (Petit Manseng). Using the 12 isolates shared by both MLST and MLSA-E

318 schemes, 10 unique haplotypes were identified by MLSA-E analysis, while only 8 were identified using

319 the MLST genes (data not shown).

320

321 **Phylogenetic analyses.** Phylogenetic analyses were conducted on the concatenated nine-gene

322 alignment of the 29 unique haplotypes (supplemental material, Supplemental File 1, Supplemental Table

323 1). Trees rooted using *Xanthomonas campestris* as the outgroup did not have noticeably different

324 relationships among unique *X. fastidiosa* haplotypes compared to trees rooted with the *X. fastidiosa*

325 susbp. *pauca* isolate (9a5c), so *X. fastidiosa* subsp. *pauca* was used to produce a tree with a shorter root

326 for easier viewing and to facilitate ease of use of the sequence alignments for additional analyses. Tree

327 topologies were similar for BP and ML phylogenetic analyses of the 29 unique haplotypes. Therefore,

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

328 posterior probabilities (PP) and bootstrap support (BS) values are both presented on the 50% majority
329 rule consensus Bayesian topology (Fig. 2). Support values for most clades were generally high (> 85% BS
330 and > 95% PP), and the tree is well-resolved with the exception of polyphyletic groups among the most
331 derived, closely related isolates from each lineage. However, in the ML analysis, BS values for a single
332 subsp. *multiplex* clade were low (< 50%), with subsp. *multiplex* clade A, clade B, and the isolate from elm
333 forming three clades as polytomies. In the BP analysis, the elm isolate was also most basal in the single
334 *multiplex* clade, suggesting it may represent a rogue taxa (75) with a highly variable phylogenetic
335 position.
336
337 Another difference between the BP and ML analyses occurs in terminal leaves within subsp. *fastidiosa*.
338 Within this clade, most of the variability found in the sequences analyzed here is due to
339 insertion/deletion events in the genes *copB* and *xadA*. Under most standard phylogenetic
340 implementations (including RAxML and MrBayes), gaps and missing data are treated as missing data (26)
341 (Alexandros Stamatakis, Heidelberg Institute for Theoretical Studies, personal communication). Thus,
342 gaps such as those in this study do not contribute any phylogenetic information. However, it appears
343 that RAxML is more influenced by this missing data, which can be attributed to the different strategies
344 deployed by ML versus BP to infer/integrate over trees (A. Stamatakis, personal communication). As a
345 result, whereas MrBayes produces a large polytomy for subsp. *fastidiosa* isolates, RAxML produces a
346 more defined topology with variable (55-100%) BS support values (Fig. 3).
347
348 All single-gene trees (data not shown) have significantly lower ΔLn *L* values than the concatenated gene
349 tree based on SH tests (Table 3). Additionally, the 9 MLSA-E genes were further divided into groups of
350 "high" dN/dS genes and groups of "low" dN/dS genes based on the Temecula:9a5c MLSA-E gene region
351 dN/dS (Table 2). By comparing ML trees from different numbers of the highest dN/dS ratio genes using

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

352 SH tests, it was determined that the 6 highest dN/dS ratio genes are needed to obtain a tree similar to
353 the nine-gene tree (ΔLn *L* of -1.7). However, comparison of the ML tree from the 8 lowest dN/dS ratio
354 genes (only *fimA* was excluded) with the nine-gene tree showed that it was still significantly worse
355 according to the SH test (ΔLn *L* of -145).
356
357 **Haplotype networks.** Isolates in the haplotype networks (Fig. 4) group together mostly along the lines of
358 the 5 major lineages seen in phylogenetic analyses (Fig. 2): subsp. *pauca*, subsp. *multiplex* Clade A,
359 subsp. *multiplex* Clade B, subsp. *sandyi*, and subsp. *fastidiosa*. Two additional haplotype networks were
360 also recovered. In this analysis, subsp. *fastidiosa* is divided into two haplotype networks. One of the
361 networks consists only of isolates from Georgia and Florida. The other network consists of all of the
362 isolates from California and Texas, 1 isolate from Georgia (CCPM1), and 1 isolate from Florida (PD 92-8).
363 The isolate from the elm, which is the only isolate from Alabama represented in this study, also forms a
364 unique clade, supporting the ML phylogenetic analysis which placed it as a unique lineage not belonging
365 to subsp. *multiplex* Clade A or Clade B. Elm is also the most basal isolate in subsp. *multiplex* Clade A
366 according to Bayesian analyses (Fig. 2).
367
368 **Evidence for recombination and linkage disequilibrium.** Of the nine genes used for the final MLSA-E,
369 there is evidence for single inter-subspecies recombination events each in particular alleles of the genes
370 *acvB* and *rpfF* (Table 3). Specifically, of the 10 recombination detection methods used, 5 and 7 methods
371 detected a recombination event in *acvB* and *rpfF*, respectively. Furthermore, a larger ZZ statistic (74)
372 indicates a decreased chance of linkage disequilibrium for *acvB* and *rpfF* compared to the other genes
373 (Table 3), which further supports the possibility of these recombination events. However, the regions of
374 the alignments within the recombination breakpoints (supplemental material, Supplemental File 1) have
375 no novel variability patterns compared to other areas of the alignment. Additionally, the recombinant

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

376 region is shared only by all subsp. *multiplex*, in the case of *acvB*, and all subsp. *fastidiosa*, in the case of
377 *rpfF* (Table 3), rather than being present in select isolates from each of these subspecies. Furthermore,
378 the SH test comparing the nine-gene ML tree with a ML tree produced without the genes *acvB* and *rpfF*
379 showed that, though the trees were significantly different as expected with two genes removed, one of
380 which had a medium dN/dS (*acvB* = 0.26), there was a low log likelihood difference (ΔLn *L* of -42). This
381 log likelihood difference is much lower than for any of the single gene trees (ΔLn *L* range of -243 to -
382 1556, Table 3) and is pointedly also much lower than the previously mentioned tree produced with the 8
383 lowest dN/dS ratio genes (only *fimA* excluded, ΔLn *L* of -145).
384
385 **DISCUSSION**
386
387 **Variation in *X. fastidiosa* detected by MLSA-E.** Previous studies of *X. fastidiosa* genetic variability have
388 not detected strong genetic variation among *X. fastidiosa* strains within subspecies(78). Phylogenetic
389 variation has been detected among subsp. *pauca* (2, 89) and *multiplex* (33, 76, 91) but has been
390 particularly limited among subsp. *fastidiosa* and *sandyi* (33, 76, 78, 91), with the exception of subsp.
391 *fastidiosa* from Costa Rica (63). For this reason, it has been assumed that within-subspecies isolates of *X.*
392 *fastidiosa* are very similar (63). However, this does not explain infectivity differences (including symptom
393 type and severity or population growth rate) that have been described among different isolates within *X.*
394 *fastidiosa* subsp. *fastidiosa* (3, 37, 51), subsp. *multiplex* (3, 29), and subsp. *pauca* (2, 45). Recent studies
395 (11, 14) indicate that genetic variability does exist among U.S. *X. fastidiosa* isolates infecting grape (*X.*
396 *fastidiosa* subsp. *fastidiosa*). This suggests the presence of undetected genetic variation and has
397 prompted exploration of MLSA-E as a new method for genetic analysis of *X. fastidiosa*. This MLSA-E
398 study is the first to use variation in multiple non-housekeeping gene DNA sequences to describe
399 differences among plant pathogenic bacteria.

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

400

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

417

418 Phylogenetic results (Fig. 2) from this MLSA-E study show that the four currently proposed *X. fastidiosa*
419 subspecies (77, 78) are recovered as reciprocally monophyletic clades. Also, the relationships between
420 the subspecies agree with recent multi-locus studies using housekeeping genes (76, 78, 91). However,
421 MLSA-E increases phylogenetic differentiation of *X. fastidiosa* isolates and identifies novel variation
422 within the same subspecies compared to previous studies, including the recent MLST scheme (91). For
423 example, enumeration of unique haplotypes among the 12 isolates analyzed in both this MLSA-E and the

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

424 MLST study found a larger number of haplotypes using MLSA-E (n = 10) versus MLST (n = 8), indicating
425 that MLSA-E provides increased differentiation of even this small set of isolates. Phylogenetic results are
426 supported by haplotype networks, which also show delineation by subspecies, with additional variability
427 within subspecies.
428
429 The *X. fastidiosa* MLST study found that little genetic variation was present within subsp. *fastidiosa* and
430 subsp. *sandyi*, despite analyzing 86 and 21 isolates, respectively (91). While the MLSA-E study here did
431 not assess multiple subsp. *sandyi* isolates, it does identify increased variation in subsp. *fastidiosa* among
432 30 isolates. MLSA-E results from *X. fastidiosa* grape isolates from different locations in the US show
433 genetic variability as evidenced by phylogenetic analyses of the nine concatenated genes. Though
434 isolates are more differentiated using ML analyses than BP analyses in this study, both analyses show
435 variability within subsp. *fastidiosa*. Haplotype networks also show increased differentiation, and some
436 geographic trends are evident. Isolates from Georgia and Florida form a distinct network from isolates
437 from California and Texas (Fig. 4), establishing a difference among isolates from the East vs. West coasts
438 of the US. The exceptions to this are a single isolate from Georgia (CCPM1) and a single isolate from
439 Florida (PD 92-8). Though it is difficult to know why these exceptions exist in the southeastern US when
440 they are genetically more similar (though not identical) to isolates from the western US, it is possible
441 that these isolates represent haplotypes that are adaptively intermediate between southeastern and
442 western isolates or are a product of recombination, specifically if isolates were transported in infected
443 plant material from a western location.
444
445 Though it has been less problematic to differentiate *X. fastidiosa* isolates within subsp. *multiplex* than
446 within subsp. *fastidiosa*, results from MLSA-E of this expanded collection of isolates have also revealed a
447 new finding within subsp. *multiplex*. MLSA-E indicates the presence of two large, distinct, reciprocally

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

448 monophyletic clades within subsp. *multiplex*, while only one of these clades was identified in the recent
449 MLST study (91). Subspecies *multiplex* Clade A is a possible novel clade composed of *X. fastidiosa*
450 isolates from hosts for which little to no phylogenetic analysis has been conducted, including strains
451 from the recently identified host blueberry (7), ragweed, and (potentially) elm. In the only other study to
452 include isolates from elm and ragweed, 16S and *gyrB* trees grouped these isolates in the same clade, but
453 they were paraphyletic with an almond isolate (Dixon) and a plum isolate, respectively, so were not
454 differentiated from members of subsp. *multiplex* Clade B (71). One study (78) did analyze the almond
455 isolate analyzed here (ALS6) and found it, along with another almond isolate, to be in a separate clade
456 from Clade B. However, these two almond isolates were the only isolates in this study with recombinant
457 alleles used to produce this phylogeny, and so could be attracted away from the main clade and towards
458 each other due to sequence variability from recombination. It is possible that this two-isolate, single
459 host clade corresponds to the 10-isolate, four-host Clade A in the present study, but more data is
460 required to determine this.
461
462 Subspecies *multiplex* Clade B is the previously identified clade (91) which contains the same three
463 subclades composed of isolates from almond (including isolate M12 used in both this and the MLST
464 study (91)), plum, and oak, respectively, in the same general topology. However, isolates from five
465 additional host species belonging to Clade B are presented in the current study, which increases
466 complexity in the topology and indicates new relationships. Additionally, in ML analysis, Clade A and
467 Clade B form a polytomy with *X. fastidiosa* subsp. *fastidiosa*, showing that their relationships are
468 unresolved. This indicates that Clade A and Clade B may not even be sister clades, further strengthening
469 the argument that are distinctly different clades. Different haplotype networks for each of the subsp.
470 *multiplex* clades support these results. Similarly, the relationship of the single *X. fastidiosa* isolate from
471 elm analyzed in this study varies depending on whether ML or BP methods are used. Though BP analysis

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

472 places this isolate as the basal member in the new subsp. *multiplex* Clade A, ML analysis places it by
473 itself in a paraphyletic relationship with subsp. *multiplex* Clade A, subsp. *multiplex* Clade B, and subsps.
474 *sandyi*/*fastidiosa*. Haplotype network results also place the *X. fastidiosa* isolate from elm in its own
475 network containing no other isolates. Additional *X. fastidiosa* isolates from elm and related plant hosts
476 are needed to clarify the relationship of the elm isolate to the other isolates.
477
478 **Using dN/dS to predict sequence variability.** Results from this study show that dN/dS can be used as a
479 predictor of sequence variability to identify genes that may be suitable for isolate differentiation in
480 closely related lineages. Environmentally-mediated genes with greater numbers of alleles (greater
481 sequence variability) can be predicted by dN/dS of the most divergent *X. fastidiosa* strains. Also, these
482 results strongly suggest that even though genes with dN/dS < 1 are not technically defined as under
483 positive selection (39, 90), genes with larger (though < 1) dN/dS may still be more likely to be under
484 positive selection than genes with smaller dN/dS for some bacteria such as *X. fastidiosa*, and they can be
485 adequate for discriminating among closely-related isolates. Genes selected for MLSA-E had significantly
486 greater dN/dS than *X. fastidiosa* housekeeping genes, though none of their ratios were > 1, and they are
487 also assumed to be under positive selection based on their functions (Table 1). Furthermore, closely-
488 related *X. fastidiosa* isolates may not be expected to have any genes with dN/dS > 1. Other studies have
489 indicated that dN/dS are insensitive to detecting adaptive selection in genes subject to low levels of
490 diversifying selection, necessitating the usage of other methods for estimating adaptive selection for
491 these genes (55, 66). However, housekeeping gene and environmentally-mediated gene dN/dS
492 variability in other organisms must be determined empirically in the same manner as done here for *X.*
493 *fastidiosa* prior to MLSA-E. It appears appropriate, based on the methods used in the present study, to
494 define a dN/dS threshold relevant to the specific study organism using a variety of housekeeping and

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

495 non-housekeeping genes as an alternative to the dN/dS > 1 rule to identify genes that are potentially

496 under positive selection.

497

498 dN/dS results also show that it is more appropriate to calculate dN/dS for the exact region of interest in

499 a gene, rather than for the entire gene, if it is to be used as a predictor of sequence variability. In this

500 study, Temecula:9a5c dN/dS of the entire gene was initially calculated for the MLSA-E genes and

501 correlated to the number of alleles among the 28 isolate subset. Using a dN/dS cutoff value of 0.15, the

502 gene *pilJ* was eliminated from further analyses because it had a dN/dS of 0.10 for the whole gene (Table

503 2). It was later determined that the gene dN/dS was not as closely correlated with the number of alleles

504 as dN/dS of the gene region of interest (Fig. 1). dN/dS of the *pilJ* gene region used for MLSA-E was 0.20,

505 much higher than the whole gene dN/dS. If genes had been selected initially using the MLSA-E gene

506 region dN/dS, *pilJ* would have been retained in the analyses. Therefore, the *pilJ* gene region analyzed in

507 this study (supplemental material, Supplemental Table 2) may still be a good candidate for future MLSA-

508 E studies.

509

510 **Recombination in *X. fastidiosa*.** Analysis of recombination and linkage disequilibrium in the nine final

511 MLSA-E genes indicates evidence of inter-subspecies recombination in the genes *acvB* and *rpfF*.

512 Recombinant genes can produce results inconsistent with bacterial evolutionary phylogenies (53), so it is

513 important to consider recombination when constructing phylogenies. No single method for detecting

514 recombination is sufficient for all circumstances (53), so multiple methods were applied in this study.

515 Though not all methods detected recombination in *acvB* and *rpfF* (Table 3), recombination was detected

516 by the majority of methods.

517

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

518 However, it seems unlikely that recombination events in these genes had a significant effect on the
519 MLSA-E results. First, *acvB* and *rpfF* have less of an effect on the phylogeny than the other MLSA-E
520 genes. They are members of the low dN/dS gene group and have less sequence variability and fewer
521 alleles. The regions of the alignments within the recombination breakpoints, specifically, have limited
522 (and not unique) variability. Also, recombination events are shared by either all subsp. *multiplex* (*acvB*,
523 parent isolate orange) or all subsp. *fastidiosa* (*rpfF*, parent isolate oleander). Because of this, the
524 recombinant regions should not drastically alter the phylogenetic signal within subspecies, as isolates
525 within these subspecies should be subjected to the same phylogenetic forces. The recombinant regions
526 could affect the relationships between subspecies, but these effects should also be limited. This is due
527 to the lesser variability mentioned previously and because any effects should be diluted among the
528 other nine genes, particularly because the two inter-subspecies recombination events occur between
529 different sets of subspecies. This is supported by the SH test, which indicates that, despite removing
530 these two genes to create a ML tree with no detectable recombination, the tree (though significantly
531 different as expected from removing two genes) had a very low log likelihood difference compared to
532 the nine-gene tree. Also, compared to the significant evidence of recombination seen in the *cysG*
533 housekeeping gene used in the *X. fastidiosa* MLST study (91), the recombination in the MLSA-E genes is
534 minimal. *cysG* has, by far, the greatest number of alleles (n=23) of all the MLST genes found among 145
535 *X. fastidiosa* isolates (41). Though the Temecula:9a5c dN/dS for *cysG* is higher than most of the other
536 MLST housekeeping genes (dN/dS = 0.19, Table 2), the number of alleles for this gene is nonetheless
537 disproportionately large, probably due to recombination. For *acvB* and *rpfF*, however, the number of
538 alleles found in relation to the Temecula:9a5c dN/dS follows the trendline seen for all of the MLSA-E
539 genes analyzed in this study (Fig. 1).
540

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

541 Recombination can play a key role in producing variation within populations (53). Previously, *X.*
542 *fastidiosa* has been thought to be largely clonal, with limited evidence of recombination (76, 78). These
543 results and a recent study (42) indicate that there may be more recombination among *X. fastidiosa*
544 isolates than previously identified. Biologically, these recombination events in *acvB* and *rpfF* may be
545 influenced by host and/or environmental selection pressures. Since the identified recombinant regions
546 in these genes are shared by all members of the recombinant subspecies, this may be indicative of a
547 selective sweep through the population, where genetic diversity is eliminated at the specific allele (79).
548 Particularly for *rpfF*, responsible for the biosynthesis of diffusible signal factor (DSF) which regulates
549 several virulence traits (8), the recombinant region may produce an adaptive advantage, causing this
550 allele to sweep through populations of *X. fastidiosa* subsp. *fastidiosa*.
551
552 **Specific sources of strain divergence in *X. fastidiosa*.** It has previously been suggested that indels are a
553 major source of strain divergence in *X. fastidiosa* and may be related to host specificity (21). The 21 bp
554 indel present in the gene *xadA* may reveal such strain divergence. The insertion is present in all isolates
555 analyzed here from subsp. *pauca*, subsp. *multiplex*, and subsp. *sandyi*. However, among subsp.
556 *fastidiosa* isolates, the insertion is only present in isolates from Georgia and Florida (regardless of host).
557 The region is deleted in all other subsp. *fastidiosa* isolates, and the larger subsp. *fastidiosa* haplotype
558 network consists only of isolates with the deletion (Fig. 4). *xadA* (Table 1) encodes an afimbrial adhesin
559 membrane protein involved in biofilm formation and virulence in grapevines (25). As has been shown for
560 other membrane proteins, *xadA* may be subject to positive selection due to increased exposure to the
561 host and other environmental factors (27). Because ancestral *X. fastidiosa* strains likely lack this deletion
562 (it is absent in the majority of isolates analyzed here), it may be an adaptive variation in non-
563 southeastern US *X. fastidiosa* subsp. *fastidiosa* isolates.
564

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

565 The two variant VNTR regions found in the gene *copB* are also of special interest. VNTR regions are able
566 to mediate rapid phenotypic changes through changes in gene expression (30). Coding VNTRs can
567 enhance bacterial virulence and play a role in adaptation to the environment (59). It is likely that the
568 two kinds of coding VNTR in this gene, which differ by only one amino acid change, represent a source
569 of adaptive variation, particularly because *copB* is involved in resistance to copper (Table 1). Copper is
570 often used to treat phytopathogens in agricultural crops, which triggers resistance mechanisms in these
571 pathogens (84). Therefore, the number of each type of VNTR may play a role in responding to copper in
572 the environment, which warrants additional exploration in future studies.
573
574 **Conclusions.** These analyses show that MLSA-E can be successfully applied to plant pathogenic bacterial
575 isolates for strain differentiation. Environmentally-mediated genes provide a more variable phylogenetic
576 signal than housekeeping genes; therefore, MLSA-E is an appropriate method for clarifying phylogenetic
577 relationships between closely related bacterial strains with low genetic variability, such as *X. fastidiosa*
578 strains. Additionally, this study shows that dN/dS may be a useful predictor of variable genes for MLSA-E
579 or other studies and can guide careful selection of genes subject to a variety of environmental
580 pressures. *A priori* selection of genes for MLSA-E must consider the environment where the bacterium
581 thrives and the stresses it encounters (e.g. xylem vessels/leafhopper foregut and strong shear forces,
582 respectively, for *X. fastidiosa*). As complete genomes become increasingly available, it will be useful to
583 scan genomes for regions of high dN/dS to gain an understanding of the ranges for particular lineages
584 and areas that may be useful for elucidating variability. This would provide a better idea of what may
585 constitute a signal of positive selection for these specific organisms, rather than relying on finding
586 regions of dN/dS > 1, which is difficult for closely-related organisms. This MLSA-E study exposes clear
587 genetic differences among *X. fastidiosa* subsp. *fastidiosa* isolates from the East vs. West coasts of the US
588 and characterizes blueberry isolates as belonging to a new clade within *X. fastidiosa* subsp. *multiplex*.

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

589 Phenotypic characterization and virulence tests *in planta* of *X. fastidiosa* isolates used in this study are
590 necessary to increase understanding of the genetic differentiation of bacterial populations found here.
591
592 **ACKNOWLEDGEMENTS**
593
594 Funding was provided by the Auburn University College of Agriculture and the Department of
595 Entomology and Plant Pathology. We would like to thank the following people for sharing their *X.*
596 *fastidiosa* isolate collections: Rodrigo Almeida (University of California, Berkeley), Mark Black (Texas
597 A&M University), C.J. Chang (University of Georgia), Donald Hopkins (University of Florida), Jianchi Chen
598 (USDA, Parlier, CA), and Harald Scherm (University of Georgia). We would like to thank the following
599 people for assistance collecting host plant tissue samples from which new *X. fastidiosa* isolates
600 examined in this study were isolated: Phillip Brannen (University of Georgia), Jim Jacobi (Alabama
601 Cooperative Extension System), and Greg Sheppard (Lumpkin County Cooperative Extension Service).
602 Thanks to Scott Santos (Auburn University) for advice regarding phylogenetic analyses and manuscript
603 comments. Thanks to Stephanie Clugstone and Heather Hart for assistance with laboratory work. This
604 publication made use of the *Xylella fastidiosa* MLST website (http://pubmlst.org/*X. fastidiosa*/)
605 developed by Keith Jolley and sited at the University of Oxford (Jolley et al. 2004, *BMC Bioinformatics*,
606 **5**:86). The development and maintenance of this site has been funded by the Wellcome Trust.
607
608

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

609 REFERENCES

610 1. **Akaike, H.** 1974. New look at statistical-model identification. IEEE Trans. Automat. Contr. **AC-**
611 **19:**716-723.
612 2. **Almeida, R. P. P., F. E. Nascimento, J. Chau, S. S. Prado, C. W. Tsai, S. A. Lopes, and J. R. S.**
613 **Lopes.** 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus
614 and coffee in Brazil. Appl. Environ. Microbiol. **74:**3690-3701.
615 3. **Almeida, R. P. P., and A. H. Purcell.** 2003. Biological traits of *Xylella fastidiosa* strains from
616 grapes and almonds. Appl. Environ. Microbiol. **69:**7447-7452.
617 4. **Barends, T. R., J. J. Polderman-Tijmes, P. A. Jekel, C. M. Hensgens, E. J. de Vries, D. B. Janssen,**
618 **and B. W. Dijkstra.** 2003. The sequence and crystal structure of the alpha-amino acid ester
619 hydrolase from *Xanthomonas citri* define a new family of beta-lactam antibiotic acylases. J. Biol.
620 Chem. **278:**23076-84.
621 5. **Blum, J. K., and A. S. Bommarius.** 2010. Amino ester hydrolase from *Xanthomonas campestris*
622 pv. *campestris*, ATCC 33913 for enzymatic synthesis of ampicillin. J. Mol. Catal. B Enzym. **67:**21-
623 28.
624 6. **Bruen, T. C., H. Philippe, and D. Bryant.** 2006. A simple and robust statistical test for detecting
625 the presence of recombination. Genetics **172:**2665-81.
626 7. **Chang, C. J., R. Donaldson, P. Brannen, G. Krewer, and R. Boland.** 2009. Bacterial leaf scorch, a
627 new blueberry disease caused by *Xylella fastidiosa*. Hortscience **44:**413-417.
628 8. **Chatterjee, S., R. P. P. Almeida, and S. Lindow.** 2008. Living in two worlds: The plant and insect
629 lifestyles of Xylella fastidiosa. Annual Review of Phytopathology **46:**243-271.
630 9. **Chatterjee, S., K. L. Newman, and S. E. Lindow.** 2008. Cell-to-cell signaling in *Xylella fastidiosa*
631 suppresses movement and xylem vessel colonization in grape. Mol. Plant Microbe Interact.
632 **21:**1309-1315.
633 10. **Chen, J., C. J. Chang, R. L. Jarret, and N. Gawel.** 1992. Genetic variation among *Xylella fastidiosa*
634 strains. Phytopathology **82:**973-977.
635 11. **Chen, J., E. Civerolo, K. Tubajika, S. Livingston, and B. Higbee.** 2008. Hypervariations of a
636 protease-encoding gene, PD0218 (*pspB*), in *Xylella fastidiosa* strains causing almond leaf scorch
637 and Pierce's disease in California. Appl. Environ. Microbiol. **74:**3652-3657.
638 12. **Chen, J., O. Lamikanra, C. J. Chang, and D. L. Hopkins.** 1995. Randomly amplified polymorphic
639 DNA analysis of *Xylella fastidiosa* Pierces-Disease and Oak Leaf Scorch pathotypes. Appl.
640 Environ. Microbiol. **61:**1688-1690.
641 13. **Clement, M., D. Posada, and K. A. Crandall.** 2000. TCS: a computer program to estimate gene
642 genealogies. Mol. Ecol. **9:**1657-9.
643 14. **Coletta-Filho, H. D., L. S. Bittleston, and R. P. Almeida.** 2011. Spatial genetic structure of a
644 vector-borne generalist pathogen. Appl. Environ. Microbiol. **77:**2596-2601.
645 15. **Crandall, K. A., A. R. Templeton, and C. F. Sing.** 1994. Intraspecific phylogenetics: problems and
646 solutions, p. 273-297. *In* R. W. Scotland, D. J. Siebert, and D. M. Williams (ed.), Models in
647 phylogeny reconstruction, 1st ed. Published for the Systematics Association by Clarendon Press;
648 Oxford University Press, Oxford, UK.
649 16. **Davis, M. J., W. J. French, and N. W. Schaad.** 1981. Axenic culture of the bacteria associated
650 with phony disease of peach and plum leaf scald. Curr. Microbiol. **6:**309-314.
651 17. **Davis, M. J., A. H. Purcell, and S. V. Thomson.** 1980. Isolation media for the Pierce's Disease
652 bacterium. Phytopathology **70:**425-429.
653 18. **De La Fuente, L., T. J. Burr, and H. C. Hoch.** 2007. Mutations in type I and type IV pilus
654 biosynthetic genes affect twitching motility rates in Xylella fastidiosa. Journal of Bacteriology
655 **189:**7507-7510.

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

656 19. **De La Fuente, L., E. Montanes, Y. Z. Meng, Y. X. Li, T. J. Burr, H. C. Hoch, and M. M. Wu.** 2007.
657 Assessing adhesion forces of type I and type IV pili of Xylella fastidiosa bacteria by use of a
658 microfluidic flow chamber. Applied and Environmental Microbiology **73:**2690-2696.
659 20. **de Souza, A. A., M. A. Takita, H. D. Coletta-Filho, C. Caldana, G. H. Goldman, G. M. Yanai, N. H.**
660 **Muto, R. C. de Oliveira, L. R. Nunes, and M. A. Machado.** 2003. Analysis of gene expression in
661 two growth states of *Xylella fastidiosa* and its relationship with pathogenicity. Mol. Plant.
662 Microbe. Interact. **16:**867-75.
663 21. **Doddapaneni, H., J. Q. Yao, H. Lin, M. A. Walker, and E. L. Civerolo.** 2006. Analysis of the
664 genome-wide variations among multiple strains of the plant pathogenic bacterium *Xylella*
665 *fastidiosa*. BMC Genomics **7:**225.
666 22. **Doyle, J., and J. L. Doyle.** 1987. Genomic plant DNA preparation from fresh tissue - CTAB
667 method. Phytochem. Bull. **19:**11-15.
668 23. **Drummond, A. J., B. Ashton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S.**
669 **Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, and A. Wilson.** 2011. Geneious
670 v5.4, Available from http://www.genious.com/.
671 24. **Feil, H., W. S. Feil, J. C. Detter, A. H. Purcell, and S. E. Lindow.** 2003. Site-directed disruption of
672 the *fimA* and *fimF* fimbrial genes of *Xylella fastidiosa*. Phytopathology **93:**675-682.
673 25. **Feil, H., W. S. Feil, and S. E. Lindow.** 2007. Contribution of fimbrial and afimbrial adhesins of
674 *Xylella fastidiosa* to attachment to surfaces and virulence to grape. Phytopathology **97:**318-24.
675 26. **Felsenstein, J.** 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Mass.
676 27. **Fitzpatrick, D. A., and J. O. McInerney.** 2005. Evidence of positive Darwinian selection in
677 Omp85, a highly conserved bacterial outer membrane protein essential for cell viability. J. Mol.
678 Evol. **60:**268-273.
679 28. **Foley, S. L., D. G. White, P. F. McDermott, R. D. Walker, B. Rhodes, P. J. Fedorka-Cray, S.**
680 **Simjee, and S. H. Zhao.** 2006. Comparison of subtyping methods for differentiating *Salmonella*
681 *enterica* serovar Typhimurium isolates obtained from food animal sources. J. Clin. Microbiol.
682 **44:**3569-3577.
683 29. **Francis, M., E. L. Civerolo, and G. Bruening.** 2008. Improved bioassay of *Xylella fastidiosa* using
684 *Nicotiana tabacum* cultivar SR1. Plant Dis. **92:**14-20.
685 30. **Gemayel, R., M. D. Vinces, M. Legendre, and K. J. Verstrepen.** 2010. Variable tandem repeats
686 accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. **44:**445-477.
687 31. **Gomez-Diaz, E.** 2009. Linking questions to practices in the study of microbial pathogens:
688 Sampling bias and typing methods. Infect. Genet. Evol. **9:**1418-1423.
689 32. **Hendson, M., A. H. Purcell, D. Chen, C. Smart, M. Guilhabert, and B. Kirkpatrick.** 2001. Genetic
690 diversity of Pierce's disease strains and other pathotypes of *Xylella fastidiosa*. Appl. Environ.
691 Microbiol. **67:**895-903.
692 33. **Hernandez-Martinez, R., K. A. de la Cerda, H. S. Costa, D. A. Cooksey, and F. P. Wong.** 2007.
693 Phylogenetic relationships of Xylella fastidiosa strains isolated from landscape ornamentals in
694 southern California. Phytopathology **97:**857-864.
695 34. **Hernandez-Martinez, R., K. C. Dumenyo, and D. A. Cooksey.** 2006. Site-directed mutagenesis of
696 *acvB* gene in a Pierce's disease strain of *Xylella fastidiosa*. Phytopathology **96:**S47-S48.
697 35. **Hoffmann, A. A., and Y. Willi.** 2008. Detecting genetic responses to environmental change. Nat.
698 Rev. Genet. **9:**421-432.
699 36. **Hopkins, D.** 2001. Gram-negative bacteria: *Xylella fastidiosa*, p. 201-213. *In* N. W. Schaad, J. B.
700 Jones, and W. Chun (ed.), Laboratory guide for identification of plant pathogenic bacteria, Third
701 ed. APS Press, St. Paul, Minnesota.
702 37. **Hopkins, D. L.** 1985. Physiological and pathological characteristics of virulent and avirulent
703 strains of the bacterium that causes Pierces Disease of grapevine. Phytopathology **75:**713-717.

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

704 38. **Hopkins, D. L., and A. H. Purcell.** 2002. Xylella fastidiosa: Cause of Pierce's disease of grapevine
705 and other emergent diseases. Plant Disease **86:**1056-1066.
706 39. **Hughes, A. L., and M. Nei.** 1988. Pattern of nucleotide substitution at major histocompatibility
707 complex class I loci reveals overdominant selection. Nature **335:**167-70.
708 40. **Janse, J. D., and A. Obradovic.** 2010. *Xylella fastidiosa*: its biology, diagnosis, control and risks. J.
709 Plant Pathol. **92:**S35-S48.
710 41. **Jolley, K. A., M. S. Chan, and M. C. Maiden.** 2004. mlstdbNet - distributed multi-locus sequence
711 typing (MLST) databases. BMC Bioinformatics **5:**86.
712 42. **Kung, S. H., and R. P. Almeida.** 2011. Natural competence and recombination in the plant
713 pathogen *Xylella fastidiosa*. Appl. Environ. Microbiol. **77:**5278-5284.
714 43. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F.**
715 **Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.**
716 2007. Clustal W and clustal X version 2.0. Bioinformatics **23:**2947-2948.
717 44. **Lemee, L., I. Bourgeois, E. Ruffin, A. Collignon, J. F. Lemeland, and J. L. Pons.** 2005. Multilocus
718 sequence analysis and comparative evolution of virulence-associated genes and housekeeping
719 genes of *Clostridium difficile*. Microbiology **151:**3171-3180.
720 45. **Li, W. B., C. H. Zhou, W. D. Pria, D. C. Teixeira, V. S. Miranda, E. O. Pereira, A. J. Ayres, and J. S.**
721 **Hartung.** 2002. Citrus and coffee strains of *Xylella fastidiosa* induce Pierce's disease in
722 grapevine. Plant Dis. **86:**1206-1210.
723 46. **Li, Y., G. Hao, C. D. Galvani, Y. Meng, L. De La Fuente, H. C. Hoch, and T. J. Burr.** 2007. Type I
724 and type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell-cell
725 aggregation. Microbiology **153:**719-26.
726 47. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of DNA
727 polymorphism data. Bioinformatics **25:**1451-1452.
728 48. **Lin, H., E. L. Civerolo, R. Hu, S. Barros, M. Francis, and M. A. Walker.** 2005. Multilocus simple
729 sequence repeat markers for differentiating strains and evaluating genetic diversity of *Xylella*
730 *fastidiosa*. Appl. Environ. Microbiol. **71:**4888-4892.
731 49. **Liu, F. Y., R. Barrangou, P. Gerner-Smidt, E. M. Ribot, S. J. Knabel, and E. G. Dudley.** 2011.
732 Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR)
733 multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica*
734 subsp. *enterica*. Appl. Environ. Microbiol. **77:**1946-1956.
735 50. **Liu, K., S. J. Knabel, and E. G. Dudley.** 2009. *rhs* genes are potential markers for multilocus
736 sequence typing of *Escherichia coli* O157:H7 strains. Appl. Environ. Microbiol. **75:**5853-5862.
737 51. **Lopes, J. R. S., M. P. Daugherty, and R. P. P. Almeida.** 2010. Strain origin drives virulence and
738 persistence of Xylella fastidiosa in alfalfa. Plant Pathol. **59:**963-971.
739 52. **Martin, D. P., P. Lemey, M. Lott, V. Moulton, D. Posada, and P. Lefeuvre.** 2010. RDP3: a flexible
740 and fast computer program for analyzing recombination. Bioinformatics **26:**2462-2463.
741 53. **Martin, D. P., P. Lemey, and D. Posada.** 2011. Analysing recombination in nucleotide sequences.
742 Mol. Ecol. Resour. Article first published online: 19 May 2011. doi: 10.1111/j.1755-
743 0998.2011.03026.x.
744 54. **Martinati, J. C., F. T. H. Pacheco, V. F. O. de Miranda, and S. M. Tsai.** 2005. Phylogenetic
745 relationships of *Xylella fastidiosa* strains based on 16S-23S rDNA sequences. Curr. Microbiol.
746 **50:**190-195.
747 55. **McClellan, D. A., E. J. Palfreyman, M. J. Smith, J. L. Moss, R. G. Christensen, and J. K. Sailsbery.**
748 2005. Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl
749 cytochrome b proteins (vol 22, pg 437, 2005). Mol. Biol. Evol. **22:**1157-1157.

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

750 56. **Mehta, A., and Y. B. Rosato.** 2001. Phylogenetic relationships of *Xylella fastidiosa* strains from
751 different hosts, based on 16S rDNA and 16S-23S intergenic spacer sequences. Int. J. Syst. Evol.
752 Microbiol. **51:**311-318.
753 57. **Meng, Y., Y. Li, C. D. Galvani, G. Hao, J. N. Turner, T. J. Burr, and H. C. Hoch.** 2005. Upstream
754 migration of *Xylella fastidiosa* via pilus-driven twitching motility. J. Bacteriol. **187:**5560-7.
755 58. **Minsavage, G. V., C. M. Thompson, D. L. Hopkins, R. M. V. B. C. Leite, and R. E. Stall.** 1994.
756 Development of a polymerase chain-reaction protocol for detection of *Xylella fastidiosa* in plant
757 tissue. Phytopathology **84:**456-461.
758 59. **Mohamad, K. Y., A. Rekiki, G. Myers, P. M. Bavoil, and A. Rodolakis.** 2008. Identification and
759 characterisation of coding tandem repeat variants in *incA* gene of *Chlamydophila pecorum*. Vet.
760 Res. **39**.
761 60. **Mohamad, K. Y., S. M. Roche, G. Myers, P. M. Bavoil, K. Laroucau, S. Magnino, S. Laurent, D.
762 Rasschaert, and A. Rodolakis.** 2008. Preliminary phylogenetic identification of virulent
763 Chlamydophila pecorum strains. Infect. Genet. Evol. **8:**764-771.
764 61. **Morano, L. D., B. R. Bextine, D. A. Garcia, S. V. Maddox, S. Gunawan, N. J. Vitovsky, and M. C.
765 Black.** 2008. Initial genetic analysis of *Xylella fastidiosa* in Texas. Curr. Microbiol. **56:**346-351.
766 62. **Newman, K. L., R. P. P. Almeida, A. H. Purcell, and S. E. Lindow.** 2004. Cell-cell signaling
767 controls *Xylella fastidiosa* interactions with both insects and plants. Proc. Natl. Acad. Sci. U.S.A.
768 **101:**1737-1742.
769 63. **Nunney, L., X. L. Yuan, R. Bromley, J. Hartung, M. Montero-Astua, L. Moreira, B. Ortiz, and R.
770 Stouthamer.** 2010. Population genomic analysis of a bacterial plant pathogen: novel insight into
771 the origin of Pierce's Disease of grapevine in the US. PLoS One **5:**e15488.
772 64. **Pashalidis, S., L. M. Moreira, P. A. Zaini, J. C. Campanharo, L. M. Alves, L. P. Ciapina, R. Z.
773 Vencio, E. G. Lemos, A. M. Da Silva, and A. C. Da Silva.** 2005. Whole-genome expression
774 profiling of *Xylella fastidiosa* in response to growth on glucose. OMICS **9:**77-90.
775 65. **Perez-Donoso, A. G., Q. Sun, M. C. Roper, L. C. Greve, B. Kirkpatrick, and J. M. Labavitch.** 2010.
776 Cell wall-degrading enzymes enlarge the pore size of intervessel pit membranes in healthy and
777 *Xylella fastidiosa*-infected grapevines. Plant Physiol. **152:**1748-59.
778 66. **Perez-Losada, M., E. B. Browne, A. Madsen, T. Wirth, R. P. Viscidi, and K. A. Crandall.** 2006.
779 Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST)
780 data. Infect. Genet. Evol. **6:**97-112.
781 67. **Pooler, M. R., and J. S. Hartung.** 1995. Genetic relationships among strains of *Xylella fastidiosa*
782 from RAPD-PCR data. Curr. Microbiol. **31:**134-137.
783 68. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution.
784 Bioinformatics **14:**817-818.
785 69. **Randall, J. J., N. P. Goldberg, J. D. Kemp, M. Radionenko, J. M. French, M. W. Olsen, and S. F.
786 Hanson.** 2009. Genetic analysis of a novel *Xylella fastidiosa* subspecies found in the
787 Southwestern United States. Phytopathology **99:**S107-S107.
788 70. **Rodrigues, C. M., M. A. Takita, H. D. Coletta-Filho, J. C. Olivato, R. Caserta, M. A. Machado, and
789 A. A. de Souza.** 2008. Copper resistance of biofilm cells of the plant pathogen *Xylella fastidiosa*.
790 Appl. Microbiol. Biotechnol. **77:**1145-1157.
791 71. **Rodrigues, J. L. M., M. E. Silva-Stenico, J. E. Gomes, J. R. S. Lopes, and S. M. Tsai.** 2003.
792 Detection and diversity assessment of *Xylella fastidiosa* in field-collected plant and insect
793 samples by using 16S rRNA and gyrB sequences. Appl. Environ. Microbiol. **69:**4249-4255.
794 72. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under
795 mixed models. Bioinformatics **19:**1572-1574.

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

796 73. **Roper, M. C., L. C. Greve, J. G. Warren, J. M. Labavitch, and B. C. Kirkpatrick.** 2007. *Xylella*
797 *fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera*
798 grapevines. Mol. Plant Microbe Interact. **20:**411-9.
799 74. **Rozas, J., M. Gullaud, G. Blandin, and M. Aguade.** 2001. DNA variation at the rp49 gene region
800 of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. Genetics
801 **158:**1147-55.
802 75. **Sanderson, M. J., and H. B. Shaffer.** 2002. Troubleshooting molecular phylogenetic analyses.
803 Annu. Rev. Ecol. Syst. **33:**49-72.
804 76. **Scally, M., E. L. Schuenzel, R. Stouthamer, and L. Nunney.** 2005. Multilocus sequence type
805 system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and
806 point mutation to clonal diversity. Appl. Environ. Microbiol. **71:**8491-8499.
807 77. **Schaad, N. W., E. Postnikova, G. Lacy, M. Fatmi, and C. J. Chang.** 2004. *Xylella fastidiosa*
808 subspecies: *X. fastidiosa* subsp. [correction] *fastidiosa* [correction] subsp. nov., *X. fastidiosa*
809 subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. Syst. Appl. Microbiol.
810 **27:**290-300.
811 78. **Schuenzel, E. L., M. Scally, R. Stouthamer, and L. Nunney.** 2005. A multigene phylogenetic study
812 of clonal diversity and divergence in North American strains of the plant pathogen *Xylella*
813 *fastidiosa*. Appl. Environ. Microbiol. **71:**3832-9.
814 79. **Shapiro, B. J., L. A. David, J. Friedman, and E. J. Alm.** 2009. Looking for Darwin's footprints in
815 the microbial world. Trends. Microbiol. **17:**196-204.
816 80. **Shimodaira, H., and M. Hasegawa.** 1999. Multiple comparisons of log-likelihoods with
817 applications to phylogenetic inference. Molecular Biology and Evolution **16:**1114-1116.
818 81. **Stamatakis, A.** 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with
819 thousands of taxa and mixed models. Bioinformatics **22:**2688-2690.
820 82. **Stamatakis, A., P. Hoover, and J. Rougemont.** 2008. A rapid bootstrap algorithm for the RAxML
821 web servers. Systematic Biology **57:**758-771.
822 83. **Tankouo-Sandjong, B., A. Sessitsch, E. Liebana, C. Kornschober, F. Allerberger, H. Hachler, and**
823 **L. Bodrossy.** 2007. MLST-v, multilocus sequence typing based on virulence genes, for molecular
824 typing of *Salmonella enterica* subsp. *enterica* serovars. J. Microbiol. Methods **69:**23-36.
825 84. **Teixeira, E. C., J. C. F. de Oliveira, M. T. M. Novo, and M. C. Bertolini.** 2008. The copper
826 resistance operon copAB from *Xanthomonas axonopodis* pathovar *citri*: gene inactivation results
827 in copper sensitivity. Microbiology **154:**402-412.
828 85. **Templeton, A. R.** 2006. Genetic drift in large populations and coalescence, p. 118-168. *In* A. R.
829 Templeton (ed.), Population genetics and microevolutionary theory. Wiley-Liss, Hoboken, N.J.
830 86. **Templeton, A. R., K. A. Crandall, and C. F. Sing.** 1992. A cladistic analysis of phenotypic
831 associations with haplotypes inferred from restriction endonuclease mapping and DNA
832 sequence data. III. Cladogram estimation. Genetics **132:**619-33.
833 87. **Van Sluys, M. A., M. C. de Oliveira, C. B. Monteiro-Vitorello, C. Y. Miyaki, L. R. Furlan, L. E.**
834 **Camargo, A. C. da Silva, D. H. Moon, M. A. Takita, E. G. Lemos, M. A. Machado, M. I. Ferro, F.**
835 **R. da Silva, M. H. Goldman, G. H. Goldman, M. V. Lemos, H. El-Dorry, S. M. Tsai, H. Carrer, D.**
836 **M. Carraro, R. C. de Oliveira, L. R. Nunes, W. J. Siqueira, L. L. Coutinho, E. T. Kimura, E. S. Ferro,**
837 **R. Harakava, E. E. Kuramae, C. L. Marino, E. Giglioti, I. L. Abreu, L. M. Alves, A. M. do Amaral,**
838 **G. S. Baia, S. R. Blanco, M. S. Brito, F. S. Cannavan, A. V. Celestino, A. F. da Cunha, R. C. Fenille,**
839 **J. A. Ferro, E. F. Formighieri, L. T. Kishi, S. G. Leoni, A. R. Oliveira, V. E. Rosa, Jr., F. T. Sassaki, J.**
840 **A. Sena, A. A. de Souza, D. Truffi, F. Tsukumo, G. M. Yanai, L. G. Zaros, E. L. Civerolo, A. J.**
841 **Simpson, N. F. Almeida, Jr., J. C. Setubal, and J. P. Kitajima.** 2003. Comparative analyses of the
842 complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella*
843 *fastidiosa*. J. Bacteriol. **185:**1018-26.

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

844 88. **Varani, A. M., W. C. Lima, L. M. Moreira, M. C. de Oliveira, R. de Souza, E. Civerolo, A. T. R. de**
845 **Vasconcelos, and M. A. Van Sluys.** 2009. Common genes and genomic breaks: a detailed case
846 study of the *Xylella fastidiosa* genome backbone and evolutionary insights, p. x, 330 p. *In* R. W.
847 Jackson (ed.), Plant pathogenic bacteria : genomics and molecular biology. Caister Academic,
848 Norfolk.
849 89. **Wickert, E., M. A. Machado, and E. G. M. Lemos.** 2007. Evaluation of the genetic diversity of
850 Xylella fastidiosa strains from citrus and coffee hosts by single-nucleotide polymorphism
851 markers. Phytopathology **97:**1543-1549.
852 90. **Yang, Z., and J. P. Bielawski.** 2000. Statistical methods for detecting molecular adaptation.
853 Trends Ecol. Evol. **15:**496-503.
854 91. **Yuan, X., L. Morano, R. Bromley, S. Spring-Pearson, R. Stouthamer, and L. Nunney.** 2010.
855 Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's Disease and Oleander Leaf
856 Scorch in the United States. Phytopathology **100:**601-611.
857 92. **Zhang, W., B. M. Jayarao, and S. J. Knabel.** 2004. Multi-virulence-locus sequence typing of
858 *Listeria monocytogenes*. Appl. Environ. Microbiol. **70:**913-920.
859
860
861

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

862 FIGURES

863

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

869

870 Figure 2. Phylogeny of the 29 unique haplotypes found among all *X. fastidiosa* isolates analyzed (n = 54)
871 as inferred by Bayesian phylogenetic (BP) analysis of the 9 concatenated gene sequences. Nodes with <
872 50% support were collapsed. BP posterior probabilities and maximum likelihood bootstrap values are
873 presented above and below each node, respectively. Asterisks indicate 100% support values while
874 dashes indicate relationships not supported based on ML analysis. See Fig. 3 for the expanded *X.*
875 *fastidiosa* subsp. *fastidiosa* subtree.

876

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

882

883 Figure 4. Haplotype networks for all *X. fastidiosa* isolates from the concatenated 9 MLSA-E genes, color-
884 coded by geographic origin. Branches (lengths not relevant) represent 95% parsimoniously plausible

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

885 connections between haplotypes. Size of nodes indicates frequency and black dots indicate missing

886 haplotypes. The ancestral haplotype in each network is indicated by a square.

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

887 TABLES

888

889 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)	References
acvB	PD_1902	900	Virulence	Virulence protein: suggested to regulate pathogenicity and disease symptoms	ACAGTA TCGCCG TCGAAG TGATGA CATGCA TAC**R**GC GATG**Y**T TCCGAT	743	(34)
copB	PD_0101	930	Resistance	Copper resistance protein B precursor: encodes a copper-resistance cation-translocating ATPase induced by copper exposure	ATGAAC ACCCGT ACCTGG TTCGTA ATTTAG TCTCCA CCATGA GCCGCA	607-862	(70, 84)
cvaC*	PD_0215	309	Toxin	Colicin V precursor: encodes a bacteriocin precursor proposed to be a defense mechanism	TGCGTG AATT**R**A CATTGA CCG CCTAGT CTGCGG CTTAAG CAGATT	330	(20, 64)
fimA	PD_0062	555	Attachment	Fimbrial subunit precursor: component of Type I pili important for biofilm formation and aggregation	CCCAGT GCGTCG TTATCG ATTATT GGT TTT**Y**G**Y** ACTCTC AAGCAT CGCATC	557	(18-20, 24, 25, 46)
gaa	PD_0315	1992	Toxin	Glutaryl-7-ACA acylase precursor: encodes a member of the newly described family of β-lactam antibiotic acylases	TGAGAG CTGC**Y**G ATGTTC CAATGA ACAGCT TCTGGC AAGAAC AAGCAC	1129	(4, 5)
pglA*	PD_1485	1635	Cell Wall Degradation	Polygalacturonase precursor: needed for degrading host plant cell walls to allow colonization	TAGTGC TGGCCT AACGAT GT**Y**GGT CCGTAT CAGCAA CCACAT GGAAGT	828/829	(65, 73)
pilA	PD_1924	447	Motility	Fimbrial protein: major structural protein of Type IV pili used for twitching motility and upstream migration	ATCGC**K** CTGCC**Y** ATGTAC CAAA CAGCAT TGATCG T**R**TTGC TGTRTG	405	(57)
rpfF	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 TCGGAG TGATTT ATGTCC GCTGTA CATCCC ATTCCT	825	(8, 9, 62)
xadA*	PD_0731	2994	Attachment	Outer membrane afimbrial adhesin: may contribute to initial cell binding to surfaces	TGGGAG GTCAAA G**Y**ACTG CCATCA GCATTG GCAGCA ACACTC GAATCA	1087/1108	(25)

890 * Genes selected based on dN/dS

891

892

893

894

895

896

897

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print

898 Table 2. Temecula:9a5c dN/dS for housekeeping genes, non-housekeeping genes, and the final MLSA-E gene regions used in this study. Unless noted,

899 housekeeping genes are from the MLST study (91). Non-housekeeping genes are from this study. dN/dS of the non-housekeeping genes were significantly higher

900 than dN/dS of the housekeeping genes (Wilcoxon rank-sum test, two-tailed $P = 0.0013$).

Housekeeping Genes		**Non-Housekeeping Genes**		**Final MLSA-E Gene Regions**	
Gene	dN/dS	Gene	dN/dS	Gene	dN/dS
cysG	0.19	***acvB***	0.29	***acvB***	0.26
gltT	0.09	***copB***	0.28	***copB***	0.26
gyrB*	0.06	***cvaC***	0.34	***cvaC***	0.35
holC	0.11	***exbD1***	0.11	-	-
leuA	0.07	***fimA***	0.40	***fimA***	0.37
malF	0.10	***gaa***	0.23	***gaa***	0.22
nuoL	0.10	***pglA***	0.21	***pglA***	0.21
petC	0.28	***pilA***	0.24	***pilA***	0.30
pyrE**	0.08	***pilJ***	0.10	-	-
		rpfF	0.19	***rpfF***	0.19
		xadA	0.30	***xadA***	0.36
Mean	0.12	**Mean**	0.24	**Mean**	0.28
Median	0.10	**Median**	0.24	**Median**	0.26

901 * *gyrB* used in other *X. fastidiosa* phylogenetic analyses (61, 71).

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

903

904

AEM Accepts published online ahead of print

Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

905 Table 3. Statistics for sequence diversity, recombination, and linkage disequilibrium for *X. fastidiosa* unique isolate groups (n=29). Ln *L* differences between the
906 single gene tree and the concatenated nine-gene tree are also presented; all single gene trees were significantly different than the concatenated nine-gene tree
907 based on Shimodaira-Hasegawa tests. Alignment regions with gaps were excluded from calculations.

Gene	**Sequence Diversity**			**Recombination**					**Linkage Disequilibrium**	**Single Gene Tree Comparison**
	# Alleles	# Alleles (gaps excluded)	Nucleotide Diversity per Site (π)	Recombination Events	Methods Detecting Recombination*	Recombination Breakpoints in Alignment (bp)	Major Parent Isolate	Recombinant Isolates	ZZ Statistic (All Sites)	Log Likelihood Difference
acvB	10	10	0.037	1	5	220 – 315	Orange	subsp. *multiplex*	0.238	-469
copB	18	7	0.009	0	0	-	-	-	0.049	-480
cvaC	9	9	0.033	0	0	-	-	-	-0.044	-1556
fimA	11	11	0.035	0	0	-	-	-	0.143	-1446
gaa	7	7	0.007	0	0	-	-	-	-0.018	-1016
pglA	7	7	0.015	0	0	-	-	-	0.030	-472
pilA	10	10	0.084	0	0	-	-	-	0.056	-243
rpfF	6	6	0.029	1	7	4652 – 4951	Oleander	subsp. *fastidiosa*	0.223	-863
xadA	14	13	0.015	0	1†	-	-	-	0.024	-270

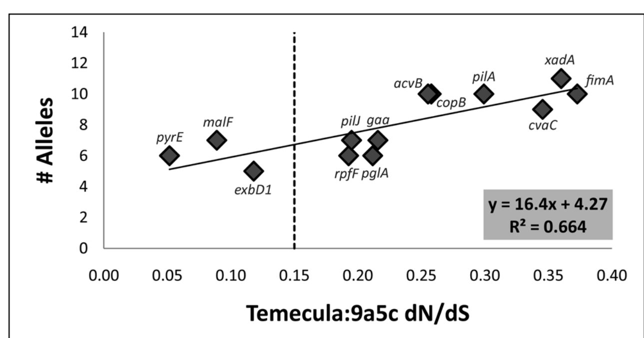
908 * Out of 10 possible methods.

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

AEM Accepts published online ahead of print

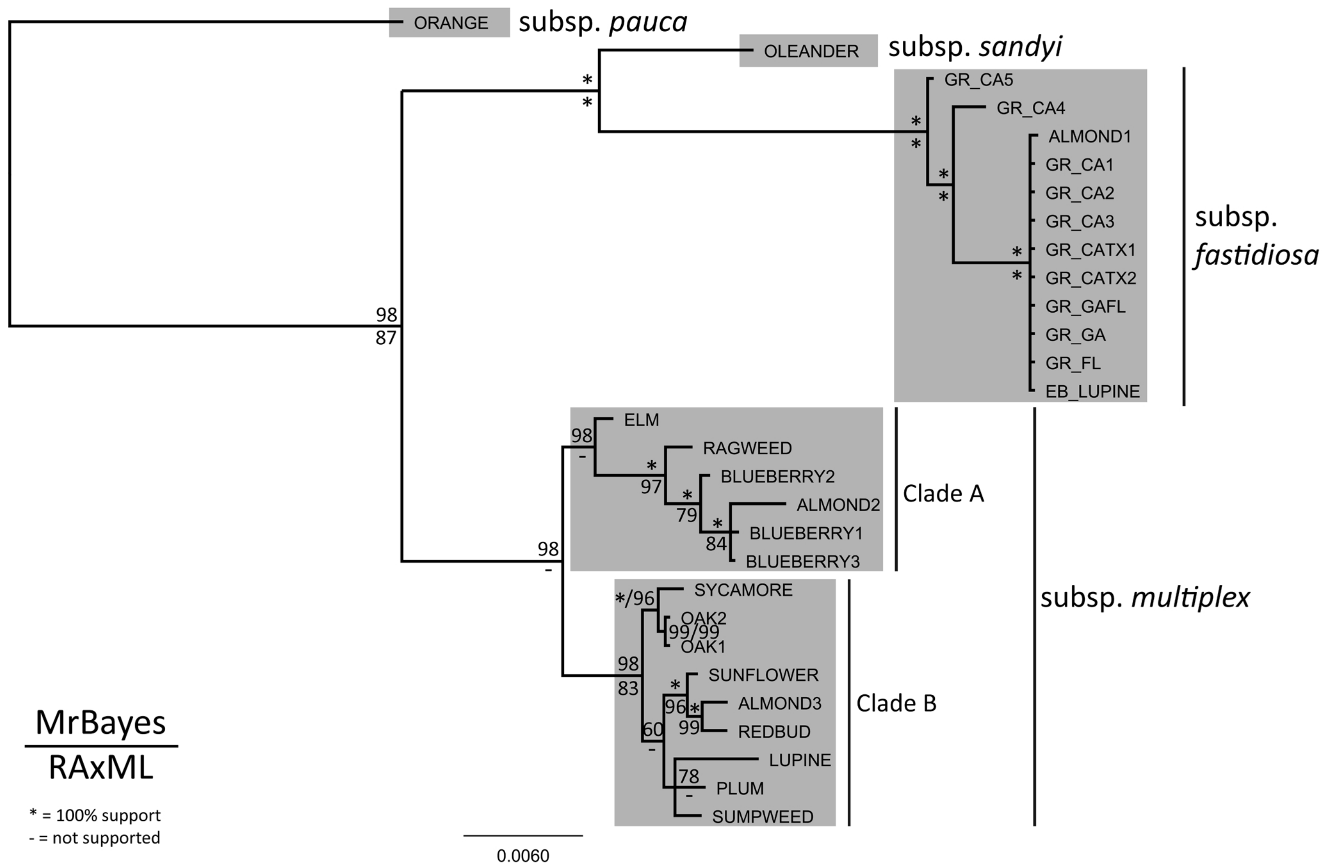
Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print



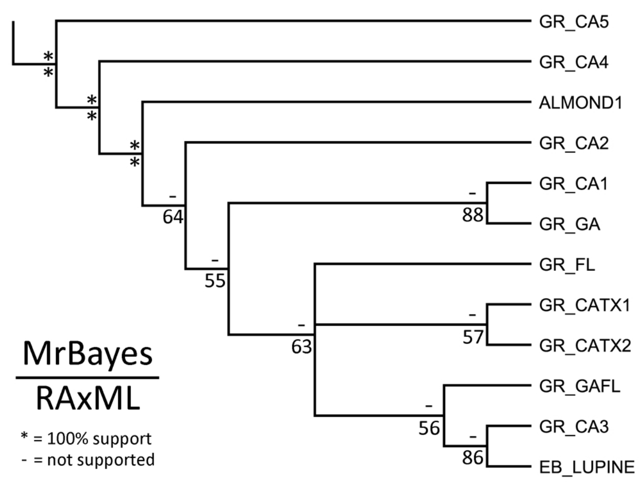
Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

AEM Accepts published online ahead of print



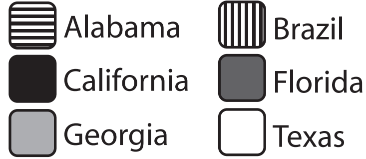
Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

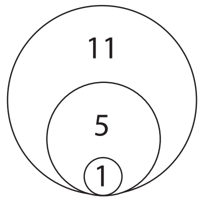
AEM Accepts published online ahead of print

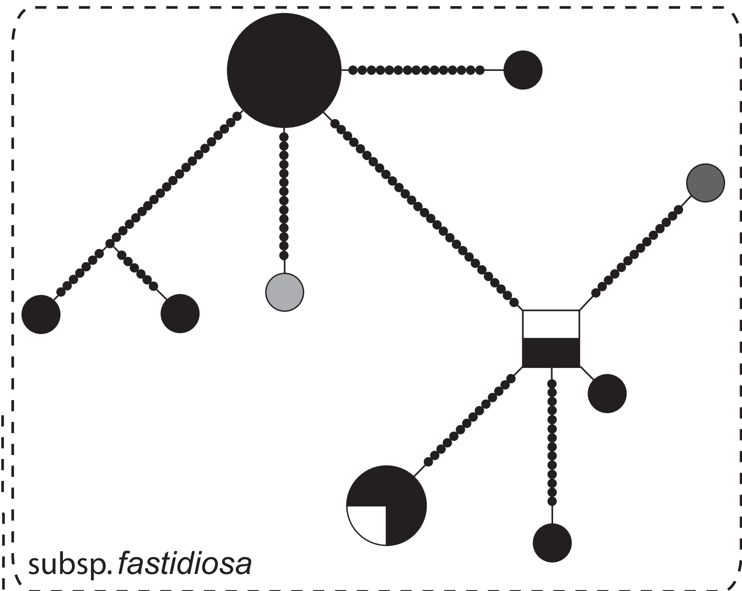


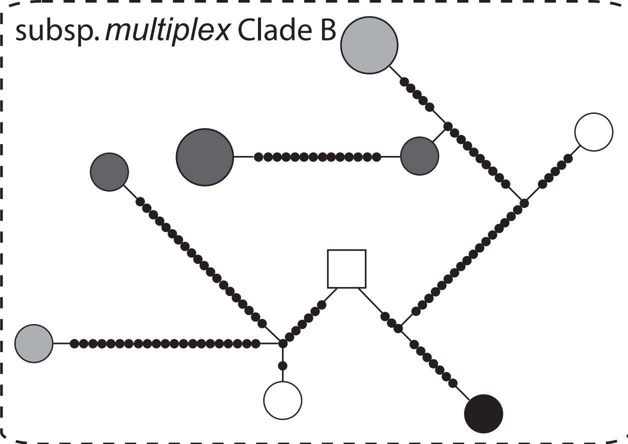
Downloaded from http://aem.asm.org/ on April 21, 2019 by guest

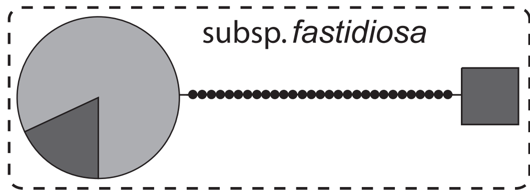
AEM Accepts published online ahead of print

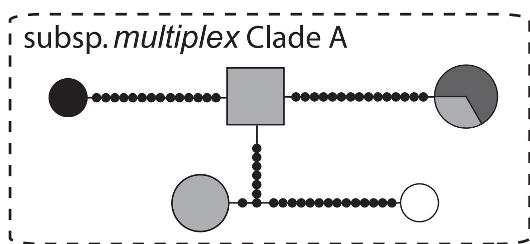












Downloaded from http://aem.asm.org/ on April 21, 2019 by guest