Spatial Genetic Structure of a Vector-Borne Generalist Pathogen

Helvécio D. Coletta-Filho,1,2 Leonora S. Bittleston,2 and Rodrigo P. P. Almeida2

LAC-Centro APTA Citros Sylvio Moreira, Condeopolis, Sao Paulo, Brazil, 1 and Department of Environmental Science, Policy, and Management, University of California, Berkeley, California2

Received 13 September 2010/Accepted 5 February 2011

Vector-borne generalist pathogens colonize several reservoir species and are usually dependent on polyphagous arthropods for dispersal; however, their spatial genetic structure is generally poorly understood. Using fast-evolving genetic markers (20 simple sequence repeat loci, resulting in a total of 119 alleles), we studied the genetic structure of the vector-borne plant-pathogenic bacterium Xylella fastidiosa in Napa Valley, CA, where it causes Pierce's disease when it is transmitted to grapevines from reservoir plants in adjacent riparian vegetation. Eighty-three different X. fastidiosa multilocus microsatellite genotypes were found in 93 isolates obtained from five vineyards, resulting in an index of clonal fraction closer to 0 and a Simpson's genotypic diversity index (D) closer to a maximum value of 1. Moderate values of Nei's gene diversity (HNei; average $H_{Nei} = 0.41$) were observed for most of the X. fastidiosa populations. The low Wright's index of genetic diversity among populations calculated by the FSTAT software (Wright's $F_{ST}$ index) among population pairs (0.0096 to 0.1080) indicated a weak or absent genetic structure among the five populations; a panmictic population was inferred by Bayesian analyses (with the STRUCTURE and BAPS programs). Furthermore, a Mantel test showed no significant genetic isolation by distance when both Nei ($r = -0.3459, P = 0.268$) and linearized $0 (r = -0.3106, P = 0.269$) indices were used. These results suggest that the riparian vegetation from which vectors acquire the pathogen prior to inoculation of grapevines supports a diverse population of X. fastidiosa.

This research was conducted in cooperation with the Sutter Valley Research Station, UCCE, and the North California Research and Extension Center,University of California, Davis, CA. The authors thank Claude Roumet, Bernard Lavigne, and Angélique Gaborit for their help in field work. This study was supported by the Brazilian National Council of Scientific and Technological Development (CNPq) through a grant to Helvécio D. Coletta-Filho (447555/2006-0) and supplementary funding from the Brazilian Agency for Evaluation of Graduate Education (CAPES) and the European Network of Excellence on Xylella fastidiosa (www.xylella.net).

Vector-borne generalist pathogens, which are not host specific, have different ecological dynamics than specialists that infect a single host (51). When generalist pathogens are dispersed by polyphagous vectors that visit many hosts, the number and diversity of host-pathogen interactions can be overwhelming. Previous research suggests that a larger diversity of reservoir species reduces disease risk by decreasing the probability of vectors feeding on competent hosts, a process known as the dilution hypothesis (21, 41). However, community diversity may also amplify disease risk (5). Generalist pathogens could have access to a larger number of competent reservoirs and, through the activity of polyphagous vectors, be disseminated more evenly and widely in the landscape. In addition, complex plant communities may support larger populations of polyphagous vectors than simpler communities (4). This is of relevance to agricultural systems, where pathogen spillover from vegetation adjacent to fields into crops is a hallmark of many vector-borne diseases (33).

The bacterial plant pathogen Xylella fastidiosa causes diseases of economic importance in several crops, including grape, almond, citrus, and coffee, many of which are emerging diseases (19). X. fastidiosa strains have variable host plant ranges; for example, isolates belonging to the almond strains cause disease only in almond plants, while grape strains cause disease in both grape and almond plants (2). In addition, X. fastidiosa is capable of colonizing a large number of host plants asymptptomatically. The grape strain of X. fastidiosa, which causes Pierce’s disease in grapevines, is a generalist and colonizes species in at least 29 different plant families (17). X. fastidiosa is transmitted by xylem sap-feeding insects, including sharpshooter leafhoppers (Hemiptera, Cicadellidae [43]). The leafhoppers that transmit X. fastidiosa are polyphagous, with over 100 plant species reported to be laboratory or field hosts for economically important vectors (18, 39). Napa Valley in coastal Northern California is an important wine grape-growing region in the United States. Pierce’s disease of grapevine was reported in Napa Valley as early as 1887 (16). The principal leafhopper vector in the region, Graphocephala atropunctata, overwinters in the riparian vegetation adjacent to vineyards. In the spring, adults move into vineyards and inoculate grapevines with X. fastidiosa, resulting in chronic plant infections (36). There is also little, if any, evidence of vine-to-vine spread in this system (35, 36). Because grapevines become sources for vector acquisition of X. fastidiosa only after it is too late in the season to establish chronic infection (44), any vine-to-vine spread of X. fastidiosa occurs after the point at which it could be expressed as vine-to-vine dispersal of chronic disease (12). Therefore, disease epidemiology is driven by the movement of vectors from riparian vegetation into vineyards once a year. Several plant species in the riparian vegetation are natural hosts of both the pathogen and the leafhopper vector, including Himalayan and California blackberry, periwinkle, French broom, elderberry, and big leaf maple, among others (37). Furthermore, vector-mediated transmission of X. fastidiosa from riparian vegetation to adjacent vineyards is a localized phenomenon in Napa Valley (35). Thus, by analyzing the genetic structure of X. fastidiosa in its focal grapevine host, it is possible to infer dispersal patterns of this generalist pathogen in the landscape. In addition, it may allow the inference of ecological and evolutionary processes shaping pathogen populations, offering insights into pathogen evolutionary potential and improving disease management strategies (28).
Bacterial simple sequence DNA repeats (SSRs) are herita-
ble loci of contiguous short sequences of repeating nucleotides. The number of repeats in a locus is prone to mutation through slipped-strand mispairing or recombination, resulting in repeat deletion or addition. Because SSRs are hypermutable loci, they have been shown to be useful and sometimes the only tool for population genetic studies of several pathogens. Furthermore, for monomorphic species such as *Mycobacterium tuberculosis*, SSRs are helpful for typing and phylogenetic purposes (50).

Recent work showed that the *X. fastidiosa* strain causing Pierce’s disease has little allelic diversity when a multilocus sequence typing (MLST) approach based on seven housekeeping
genes is used (52); thus, hypermutable loci are needed for studies addressing the ecology of this pathogen at the population level. We analyzed the population structure of *X. fastidiosa* in Napa Valley, where it is endemic, using fast-evolving SSR makers. Contrary to expectations, given the pattern of localized Pierce’s disease spread in the region, results obtained provide evidence for extreme diversity and panmixia among isolates of this vector-borne pathogen.

**MATERIALS AND METHODS**

**Samples, culturing, and DNA extraction.** Ninety-three isolates of *X. fastidiosa* were obtained from Pierce’s disease symptomatic grapesvines from five vineyards in Napa Valley, CA, geographically isolated from each other by 1.64 to 12.16 km (see Table S1 in the supplemental material). Although two grape varieties were used in the study, that was not expected to be a problem, as both are susceptible to *X. fastidiosa*. *X. fastidiosa* was cultured from the petioles of symptomatic leaves on PWG medium (17), and selected colonies were triple cloned on PWG me-
dium prior to genotyping. A commercial kit was used for DNA extraction (DNAeasy blood and tissue kit; Qiagen, Valencia, CA), and the identity of isolates as *X. fastidiosa* was confirmed by PCR with the specific primer set RST31-RST33 (30).

**Amplification of SSR loci and fragment analysis.** Twenty SSR loci previously described by Lin et al. (25) were used to genotype isolates. These SSR loci were amplified in reactions with 4 multiplexed primer sets each (see Table S2 in the supplemental material). The multiplex amplifications were performed in a final volume of 12.5 μl containing 25 to 50 ng DNA, 6.25 μl of Taq PCR master mix kit (Qiagen, Valencia, CA), and variable primer pair concentrations (with a fluorescently labeled forward primer), as described in Table S2 in the supplement-
ary material. The following program was used for PCR amplification of the loci: a denaturation step at 95°C for 6 min, followed by 35 cycles at 95°C for 30 s, at 56°C (primer sets 1, 2, and 3) or at 57°C (primer sets 4 and 5), both for 30 s, and a 2-min extension at 72°C for all sets.

After PCR, we added 0.5 μl of the amplified product to 12 μl of Hi-Di formamide with 0.5 μl of GeneScan 500-LIZ size standard (both from Applied Biosystems, Foster City, CA). After DNA denaturation for 5 min at 95°C, followed by cooling in ice for 2 min, samples were submitted to capillary elec-
trophoresis in an ABI 3730d DNA sequencer (Applied Biosystems) with the LIZ500-3730 standard setting on a GeneScan module with filter set G5. The GeneScan data were analyzed with PeakScanner (version 1.0) software (Applied Biosystems) by comparing the amplions to the internal size standard (500-LIZ), consisting of 20 labeled fragments ranging from 17 to 500 bp. For a few loci, we used as a reference the length of each repeat unit previously determined (25) to establish the most accurate locus allele size (see Table S4 in the supplemental material).

**Clonal and gene diversity.** The multilocus microsatellite genotype (MLMG) for each isolate was determined by the GenoDive software (29), where samples with the same MLMG were treated as clones. Several indices of clonal diversity were estimated, including the number of observed genotypes per population (*N*<sub>o</sub>). The clonal fraction was calculated as 1 - [(number of different genotypes within the population)/(total number of isolates within the population)] (53). Simpson’s diversity index (*D*) was estimated by the equation 1 - Σ<sub>i</sub>p<sub>i</sub><sup>2</sup> (where *p*<sub>i</sub> is the relative abundance of the *i*th genotype in the population (*D* represents infinite diversity and 0 represents no diversity) (15). Evenness, or how evenly the genotypes were divided over populations, was also estimated: a pop-
ulation composed of a single genotype has an evenness value of 0, while another where all genotypes have equal frequency has a value of 1 (14). Both the mean of Nei’s genetic diversity (*H*<sub>Nei</sub>), corrected by the number of individuals (*n*) in a population (*H*<sub>Nei</sub> = (1 - Σ<sub>i</sub>p<sub>i</sub><sup>2</sup>) /n[(n - 1)], where *p*<sub>i</sub> is the frequency of allele *i* at the locus (32)), and the allelic richness within each geographical population were estimated using the software FSTAT (version 2.9.3.2) (13). Statistics for the significance between pairs of comparisons were obtained by 1,000 randomiza-
tions by FSTAT (9). A bootstrapping test using 1,000 randomizations was per-
formed between pairs of populations to determine differences in clonal diversity.

**Population structure analysis.** The null hypothesis that populations were genetically distinct was tested using several approaches. Wright’s index of genetic diversity among populations calculated by the FSTAT software (Wright’s *F*<sub>ST</sub> index) for population differentiation of haplotids was tested for each population pair by FSTAT with 1,000 randomization steps (49). The Mantel test of isolation by distance was carried out with GenoDive, assuming a linear relationship be-
tween pairwise values of *F*<sub>ST</sub> and *H*<sub>Nei</sub> or between pairwise differences of ge-
ographical distance (kilometers) between all population pairs (40). Also, an exact and nonparametric procedure for population differentiation based on hap-
loptype frequency (38) was conducted, as implemented in the Arlequin program (MCMC parameters were 100,000 steps, 10,000 steps for the burn-in period, and a significance level at 0.05). The model-based Bayesian clustering approach implemented by both the STRUCTURE (version 2.3) and BAPS (version 5.2) (8) programs was used to estimate the number of genetically homogeneous clusters or populations (*K*). In BAPS, we did five independent runs using the cluster individuals function, which was previously recommended to be the best option for estimating the number of populations (48), setting the maximum number of clusters ranging from 1 to 15. With STRUCTURE, the admixture model was used and was run three times using 10 replications of *K* ranging from 1 to 15. For these runs the following conditions were adopted: a burn-in period at 30,000 following 300,000 replicates of the MCMC; a uniform prior for alpha (initial value = 1.0, maximum value = 10.0, standard deviation [SD] = 0.25); and the allele frequencies correlated among subpopulations (prior mean = 0.01, prior SD = 0.05, λ = 1.0). In addition, the K values with the highest likelihood were determined (10). From the most appropriate *K* values obtained as well as from the number of clusters inferred by BAPS (see Results), we reran STRUCTURE (using the same assign-
mation shown above plus prior population information) to define group mem-
bership *Q*, a proportion of the membership of each predefined population in each postdefined *K* cluster.

**RESULTS**

**Genotypic diversity within and among populations.** Twenty SSR loci were used to type 93 *X. fastidiosa* isolates from five vineyards in Napa Valley (see Table S1 in the supplemental material). The range of repeat units varied greatly across the 20 SSR loci and was largest in the OSSL9, GSSR7, and GSSR12 loci, which produced 13, 14, and 16 different allele sizes, respectively. Except for the monomorphic loci ASSR19 and CSSR12, the other loci were polymorphic, resulting in a total of 119 alleles (see Table S2 in the supplemental material). For 11 of the 20 loci, the observed repeat length was similar to that expected on the basis of the work of Lin et al. (25). For some loci, variations of 1 or 2 bp for the unit repeat length expected were observed, but those were of no consequence to repeat unit copy number (see Table S2 in the supplemental material). The allele frequencies in the *X. fastidiosa* popu-
lations studied are shown in Table S5 in the supplemental ma-
terial. We found 83 different genotypes in the 93 isolates an-
alyzed (Table 1; see Table S4 in the supplemental material).

The index of clonal fraction, the proportion of isolates shar-
ing the same genotype, ranged from 0 (OK vineyard), where
the population was composed of different MLMGs, to 0.24 (Si
vineyard), where 16 MLMGs were found among 21 isolates.
The high number of different MLMGs resulted in a Simpson’s
D value close to the maximum value of 1 for all populations
(Table 1). Likewise, the average \(H_{\text{Nei}}\), which is based on
the total number of alleles, was high (\(H_{\text{Nei}} = 0.41\), with no dif-
fERENCE among populations, except for one (OK vineyard),
which also showed a significantly lower value for genetic di-
versity than the other four populations (\(H_{\text{Nei}} = 0.38, P \leq 0.05\);
Table 1). As a consequence of low genetic diversity, isolates
from the OK population also showed allelic richness (3.02) that
was significantly lower than that for the other populations
(Table 1). The inferred lower values of genetic diversity in that
population could be correlated to the smaller number of sam-
pled isolates. We observed a linear relationship between the
number of isolates collected from each geographical popula-
tion and the values of allelic richness (\(r^2 = 0.769\)) or the
detected MLMG (\(r^2 = 0.763\)) (see Fig. S1 in the supplemental
material). Therefore, although this population (OK) seems to
deviate from the others in these diversity indices, that may be
an artifact due to the smaller sample size for that location.

**LD.** Two approaches were used to infer the presence of
linkage disequilibrium in the sampled populations. Tests of
multilocus gametic equilibrium by the index of association (\(I_a\)
and \(r_d\)) were significantly different from 0 (\(P \leq 0.01\)) for all
populations except OK (\(I_a = 0.46\), indicating linkage disequi-
librium in four populations (SC, Si, GR, and Lu), from which the
values of \(I_a\) ranged from 0.74 to 1.50. Because the \(r_d\) index
is less sensitive to variation in the number of loci than \(I_a\),
values closer to 0 (average = 0.052) were obtained, indicating
less gametic disequilibrium. The weakness of allelic association
was verified by the low percentage of allele pairs at significant
disequilibrium in all populations (1.5 to 12.5%) except the Lu
population, which had almost 30% of pairs at disequilibrium
and, consequently, the highest index of association (see Table
S3 in the supplemental material). LD and \(I_a\) values signifi-
cantly different from 0 are indications of clonal bacterial pop-
ulations, but these values are also observed when all isolates of
a bacterial population are analyzed (45), as has been found
with the recombinogenic organism *Neisseria meningitidis* (11).
This allelic association could be a consequence of individuals
from different, randomly mating populations being incorrectly
grouped into one population (23), which in this case may be a
consequence of vector movement of *X. fastidiosa* in its natural
reservoir (i.e., riparian vegetation).

**Panmixia and lack of population structure.** Several ap-
proaches were used to infer the population structure of *X.
fastidiosa* in the five sampled vineyards. The levels of genetic
differentiation among pairwise geographical populations based
on Wright’s \(F_{\text{ST}}\) test were low and ranged from 0.0096 to
0.1080, meaning that a maximum of 10% allele frequency vari-
arion arose from population differentiation on those specific
comparisons. The \(F_{\text{ST}}\) values for most population pairs were
less than 0.06 (Table 2). The low and nonstatistically different
values between pairs of populations indicated weak or absent
genetic differentiation among all geographically tested *X. fas-
idiosa* populations. A low genetic distance was observed
among all pairs of geographical populations. We did not find
evidence of isolation by geographical distance for either \(D_{\text{Nei}}\)
(Mantel’s test, \(P = 0.268; r = -0.3459\) and linearized 0 (Man-
tel’s test, \(P = 0.269; r = -0.3106\) (Fig. 1) measures of pairwise
genetic distances, rejecting the null hypothesis of spatial struc-
turing among populations.

We used Bayesian clustering analysis (7, 20) to infer the number of genetically homogeneous clusters or subpopulations
(K) in the sampled isolates. In other words, isolates were as-
signed to different genetic clusters, which themselves were
determined to be cohesive units on the basis of the data set.
Both STRUCTURE (20) and BAPS (7) can work well to infer
the number of clusters when genetic groups are not well dif-
fereNced, i.e., have low values of \(F_{\text{ST}}\) (23). The number of
subpopulations (K clusters) observed was larger than the five
geographical populations. Using STRUCTURE, the modal

### Table 1. Genotypic and genetic diversity of *Xylella fastidiosa*
isolates from five populations in Napa Valley, CA

<table>
<thead>
<tr>
<th>Geographical population</th>
<th>No. of genotypes</th>
<th>Clonal fraction&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Evenness</th>
<th>Simpson’s D</th>
<th>(H_{\text{Nei}})</th>
<th>Allelic richness</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>20</td>
<td>0.05</td>
<td>0.95&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.99&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.55&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>OK</td>
<td>12</td>
<td>0.00</td>
<td>1.00&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.38&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.02&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Si</td>
<td>16</td>
<td>0.42</td>
<td>0.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.34&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>17</td>
<td>0.11</td>
<td>0.92&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.99&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.39&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>3.25&lt;sup&gt;ab&lt;/sup&gt;</td>
</tr>
<tr>
<td>Lu</td>
<td>18</td>
<td>0.10</td>
<td>0.92&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.99&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.26&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Overall</td>
<td>83</td>
<td>0.11</td>
<td>0.89&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.99&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.41&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> See Materials and Methods for details on how indices were estimated.

### Table 2. Measures of pair-wise population differentiation based on the \(F_{\text{ST}}\) statistic and Nei’s genetic distance (\(D_{\text{Nei}}\))

<table>
<thead>
<tr>
<th>Geographical population</th>
<th>(F_{\text{ST}}) and (D_{\text{Nei}}) values&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>0.058 (NS)&lt;sup&gt;c&lt;/sup&gt; 0.047 (NS) 0.028 (NS) 0.062 (NS)</td>
</tr>
<tr>
<td>OK</td>
<td>0.0457 (NS)&lt;sup&gt;c&lt;/sup&gt; 0.110 (NS) 0.028 (NS) 0.068 (NS)</td>
</tr>
<tr>
<td>Si</td>
<td>0.0336 (NS) 0.1080 (NS) 0.061 (NS) 0.070 (&lt;i&gt;NS&lt;/i&gt;)</td>
</tr>
<tr>
<td>GR</td>
<td>0.0896 (NS) 0.0538 (NS) 0.0255 (NS) 0.053 (NS)</td>
</tr>
<tr>
<td>Lu</td>
<td>0.0502 (NS) 0.0581 (NS) 0.0607 (NS) 0.0423 (NS)</td>
</tr>
</tbody>
</table>

<sup>a</sup> \(F_{\text{ST}}\) was calculated by the FSTAT software, data for which are presented
below the diagonal. \(D\), maximum genetic distance between pairs of populations, data for which are presented
above the diagonal (32).

<sup>b</sup> Nonsignificant (NS) differences between pairs of populations were estimated
by 1,000 randomization tests.

**FIG. 1.** Nonsignificant regression between genetic and geographical
distances (kilometers) among pair-wise populations of *Xylella fas-
idiosa* from Napa Valley. Circles and dashed line (\(P = 0.268\)), \(F_{\text{ST}}\)
values; triangles and solid line (\(P = 0.269\)), \(D_{\text{Nei}}\) of pair-wise popu-
lations against the geographical distances.
values of the change in $K$ ($\Delta K$) (10) were at $K$ equal to 10 (run 1) and $K$ equal to 12 (runs 2 and 3) (Fig. 2A). However, the likelihood of each $K$, averaged for the three runs, varied little for $K$ values from 1 to 7, while the likelihoods were lower and more variable for $K$ values from 8 to 14. We interpret these results to be an inability of the algorithm to assign a clear $K$ value to the data set due to the high diversity observed among isolates. With BAPS, the number of clusters was identified at a value to the data set due to the high diversity observed among isolates. With BAPS, the number of clusters was identified at a value ranging from 90.9% to 99.6% (Fig. 2B). The output of replicated analyses had little variation in the range of likelihood values ranging from 90.9% to 99.6% (Fig. 2B). The output of replicated analyses had little variation in the range of likelihood values (maximum difference among values, 3.27), indicating that the analyses conducted in BAPS were statistically stable. Also, in BAPS the most genetically homogeneous individuals were distributed across the five geographical populations, independent from their own source localization, emphasizing the lack of structure for $X. fastidiosa$ in Napa Valley (Fig. 3A). Using STRUCTURE, the estimated membership coefficient for each individual in each cluster ($Q$) was tested assuming $K$ values of 10 and 12 (the best numbers of subpopulations or clusters provided by Bayesian analysis). For $K$ values of 10 and 12, we found that there was no population structure for $X. fastidiosa$ in Napa Valley as a consequence of the low posterior probability found for the individual in its original population (Fig. 3B). For only one population (Lu), a few isolates with a high probability of belonging to their source population were found; i.e., evidence of geographical structuring or isolation from other genotypes was detected. To our knowledge, no biological or environmental factor could be associated with this observation.

**DISCUSSION**

$X. fastidiosa$ grape isolates from vineyards in Napa Valley were found to have high levels of population genetic and genotypic diversity. In addition, little clonality was observed in the data set, suggesting that in this endemic pathosystem, $X. fastidiosa$ is very diverse in its reservoir (i.e., riparian vegetation). Alternatively, the markers used may evolve very quickly in this pathogen. The results are supported by the similar but unequal values of evenness of genotypes and richness of alleles across all populations. Other pathogens such as *Bacillus anthracis* (26) and *Neisseria meningitidis* (42) are considered to have low and high SSR diversities, respectively, and the values for the $X. fastidiosa$ isolates surveyed in this study were at the high end of that range. Variation in the number of allele sizes was observed in the data set, and their range of variation was consistent with the stepwise mutational model, where the mutations involve a single-repeat change (the insertion or deletion of one repeat), which has been found in *Escherichia coli* (47) and *Mycobacterium tuberculosis* (50). SSRs have been intensively used as neutral molecular markers in analyses of intra-species diversity of bacteria because of the high rate of polymorphism. However, functional effects have also been attributed to SSR evolution in coding regions within bacterial genomes (24, 46).

No evidence of population structure was found at the geographical scale sampled in this study; genetically similar isolates were not grouped geographically but apparently were randomly distributed among the five geographical populations in Napa Valley vineyards. Statistical inference based on genotypic clustering suggested the existence of at least 10 different

---

**FIG. 2.** Highest posterior probability distributions for $K$ present in the five *Xylella fastidiosa* geographical populations. (A) Results of three independent runs by STRUCTURE, indicating that $K$ is equal to 10 or 12, on the basis of the $\Delta K$ values (10). The second $y$ axis (filled diamonds) indicates the average likelihood values $[L(K)]$ for various values of $K$. (B) A value of $K$ of 10 was obtained with five independent replicate runs by BAPS.

**FIG. 3.** Inferred population structure of *Xylella fastidiosa* from five vineyards in Napa Valley, CA. Each isolate is represented by a single vertical line, and the black line separates the five geographical populations. (A) BAPS clustering of individuals. Each isolate has the color corresponding to the cluster where it was placed. Shared colors denote individuals resembling each other genetically as much as possible. (B) STRUCTURE clustering (admixture), where each isolate is represented by a single vertical line, which is partitioned into $K$ colored segments that represent the individual’s estimated membership fractions in $K$ clusters. The length of the color segment shows the strain’s estimated $Q$ in that cluster. Individuals with multiple colors have admixed genotypes from the previously defined multiple populations. The figure shown for $K$ equal to 10 and $K$ equal to 12 is based on the highest probability run at that $K$. 

Downloaded from http://aem.asm.org on October 15, 2017 by guest
subpopulations with extensive admixture, a characteristic that could have contributed to the significant LD patterns observed for most *X. fastidiosa* populations, as previously discussed (23, 31). In addition, it should be noted that evidence for so many subpopulations with such high levels of admixture might also mean that there is no real population structure. The lack of genetic structure for *X. fastidiosa* in Napa Valley contrasts with the structure for another *X. fastidiosa*-caused disease, citrus variegated chlorosis in Brazil, for which a clear geographical structure was observed (6). The two systems have important ecological distinctions that support the genetic data. Pierce’s disease of grapevines has been endemic in Napa Valley for over 100 years, whereas citrus variegated chlorosis emerged in 1987, became epidemic, and is still increasing in incidence (19). Emerging plant pathogens affecting monocultures are expected to expand in a spatially structured fashion, as there are no dispersal barriers and all hosts are similarly susceptible. The spread of *X. fastidiosa* in Napa Valley vineyards shows a different disease progression, as it occurs from a habitat adjacent to the focal host grape, rather than through vine-to-vine spread within vineyards (19). In our study, the vineyards were located within 14 kilometers of each other and shared riparian corridors, possibly allowing the mixing of strains over time due to the movement of vectors in that habitat. The opposite pattern occurs in citrus, where the pathogen spreads by infection waves that originate from infected trees within an orchard, as well as from surrounding orchards that may serve as reservoirs (22). Furthermore, although both citrus and grape strains of *X. fastidiosa* are generalists, the grape strain seems to be capable of colonizing a larger number of host plants of epidemiological importance (37) than the citrus strain (27).

Generalist pathogens characteristically have high genetic diversity and abundant opportunities for cross-host infection (51). In addition to the high level of genetic diversity observed, we found similar values for allelic richness and evenness in all populations. The analyzed populations had a high level of genotypic diversity and little clonality. These results indicate that no bottlenecks have recently occurred in this population. Because the riparian vegetation is the main source of *X. fastidiosa* to vines (35), we suggest that genetically different strains of *X. fastidiosa* are endemically established in natural hosts throughout the riparian vegetation bordering the vineyards. Naturally infected leafhopper vectors could move the pathogen along the riparian vegetation as well as into the vineyards, creating the observed lack of population structure. Additionally, a test of population differentiation, which tests the null hypothesis of no population differentiation (38), was used and supported the findings of other analyses, with highly nonsignificant results being obtained for all the population pair-wise comparisons (*P* = 1.0). These results support the concept that high pathogen diversity in focal hosts may occur if generalist pathogens are capable of exploiting multiple reservoir species. It is possible that different reservoir species preferentially harbor certain genotypes of *X. fastidiosa*, as is the case in Lyme disease, where some *Borrelia burgdorferi* genotypes are predominantly found in different mammal reservoirs (3). Furthermore, leafhopper vectors in Napa Valley likely migrate in the riparian corridors where the disease occurs, facilitating dissemination of strains across the landscape, as these insects are long-lived and *X. fastidiosa* is retained in adults for life (43). Surveying for *X. fastidiosa* genotypes in reservoirs in the riparian vegetation will be necessary to address these questions; however, there is a very low rate of success in isolating *X. fastidiosa* even from known infected symptomless plants (17, 37).

Pathogens with a large effective population size (of high genetic diversity) are expected to adapt faster under natural selection and, consequently, represent a higher risk of circumventing the effects of host resistance genes (28). The high genotypic diversity of the *X. fastidiosa* isolates analyzed in this study indicates that a large population size with multiple haplotypes spread over the vineyards by the influx of infected vectors from alternative reservoirs to the focal host. On the basis of this particular ecological condition plus the endemic presence of *X. fastidiosa* in the region, this pathogen could be capable of rapidly evolving to adapt to new opportunities for exploiting novel or changed resources. These opportunities would include naturally resistant varieties and genetically modified plant material.

ACKNOWLEDGMENTS

We thank Mate Lopes for assistance with *X. fastidiosa* culturing, John Chau for help with multiplexing of primer sets, and Sandy Purcell for helpful discussions. We also thank our lab colleagues for constructive suggestions and discussions.

This work was funded by the United States Department of Agriculture and the California Department of Food and Agriculture. R.P.P.A. acknowledges CAPES/Brazil for a postdoctoral fellowship, project number 3663-08-6, which made this collaborative project possible.

REFERENCES