AEM
Journals.ASM.org

Intersubspecific Recombination in *Xylella fastidiosa* Strains Native to the United States: Infection of Novel Hosts Associated with an Unsuccessful Invasion

Leonard Nunney,[a] Donald L. Hopkins,[b] Lisa D. Morano,[c] Stephanie E. Russell,[d] Richard Stouthamer[d]

Biology Department, University of California, Riverside, California, USA[a]; University of Florida, Mid-Florida Research and Education Center, Apopka, Florida, USA[b]; Department of Natural Sciences, University of Houston—Downtown, Houston, Texas, USA[c]; Entomology Department, University of California, Riverside, California, USA[d]

The bacterial pathogen *Xylella fastidiosa* infects xylem and causes disease in many plant species in the Americas. Different subspecies of this bacterium and different genotypes within subspecies infect different plant hosts, but the genetics of host adaptation are unknown. Here we examined the hypothesis that the introduction of novel genetic variation via intersubspecific homologous recombination (IHR) facilitates host shifts. We investigated IHR in 33 *X. fastidiosa* subsp. *multiplex* isolates previously identified as recombinant based on 8 loci (7 multilocus sequence typing [MLST] loci plus 1 locus). We found significant evidence of introgression from *X. fastidiosa* subsp. *fastidiosa* in 4 of the loci and, using published data, evidence of IHR in 6 of 9 additional loci. Our data showed that IHR regions in 2 of the 4 loci were inconsistent (12 mismatches) with *X. fastidiosa* subsp. *fastidiosa* alleles found in the United States but consistent with alleles from Central America. The other two loci were consistent with alleles from both regions. We propose that the recombinant forms all originated via genomewide recombination of one *X. fastidiosa* subsp. *multiplex* ancestor with one *X. fastidiosa* subsp. *fastidiosa* donor from Central America that was introduced into the United States but subsequently disappeared. Using all of the available data, 5 plant hosts of the recombinant types were identified, 3 of which also supported non-IHR *X. fastidiosa* subsp. *multiplex*, but 2 were unique to recombinant types from blueberry (7 isolates from Georgia, 3 from Florida); and blackberry (1 each from Florida and North Carolina), strongly supporting the hypothesis that IHR facilitated a host shift to blueberry and possibly blackberry.

Understanding the mechanisms of adaptive evolution in pathogenic bacteria is central to long-term disease control. One major focus of research into adaptive bacterial evolution has been lateral (or horizontal) gene transfer (LGT), usually defined as the transfer of genes (or genetic material) across species boundaries (1–6). Until recently, discussions of LGT focused on the transfer of novel genes, as exemplified by the discovery of the plasmid-mediated transfer across species of the genes coding for penicillin resistance (7); however, with the increasing availability of genomic sequence data, it has become apparent that the transfer of homologous gene copies (i.e., novel alleles) is also widespread (8). These two kinds of exchange, the transfer of novel genes or novel alleles, are fundamentally different. The acquisition of novel genes can result in the acquisition of a completely new trait that has already been refined in other taxa by natural selection (as in the case of penicillin resistance). It can determine critical traits such as virulence, antibiotic resistance, and ecological niche (9), even though most of the material transferred appears to be evolutionarily transient (10). In contrast the acquisition of novel alleles is analogous to the effect of sexual reproduction in eukaryotes: it increases the genetic variance that natural selection can act on but does not, in itself, result in a qualitative change in the ecology of the recipient (1). Due to these fundamental differences, we favor reserving the term "LGT" for the transfer of novel genes, using the term "interspecific" or "intersubspecific homologous recombination" (IHR) for the transfer of alleles; however, both processes, if successful, lead to genetic "introgression," a term commonly used to describe the spread of genetic material across taxonomic boundaries in plants and animals and now increasingly used to describe the analogous process in bacteria (4, 11, 12).

Homologous recombination is almost ubiquitous among bacteria, although the degree to which it occurs varies widely among species (13). It involves the replacement of a stretch of DNA sequence in one individual's genome by a homologous sequence from another individual of the same (or closely related) species following any of the 3 mechanisms of DNA transfer (transformation, transduction, or conjugation). It typically involves short pieces of DNA (often <1 kb) (14). Given the prevalence of homologous recombination, it is generally assumed that it is beneficial, in some cases enabling bacteria to enhance their resistance to antibiotics (15) and avoid host defenses (16, 17) or perhaps promoting adaptation to novel environments (18). Analogy with the assumed benefits maintaining sexual recombination in metazoans strongly supports this view.

Documenting the adaptive benefit of homologous recombination in bacteria has proved difficult. This is to be expected even if the benefits are large and common. Homologous recombination typically falls off rapidly with genetic distance (19), so a well-established population will usually reflect the mixing of relatively similar alleles. This mixing can be easily detected by the lack of clonality between genes and quantified using evolutionary models (e.g., ClonalFrame) (20); however, detection of recombination breaks within genes is more problematic. The approaches currently used (e.g., LDhat [21] and PHItest [22]) have very limited

Received 28 August 2013 **Accepted** 25 November 2013
Published ahead of print 2 December 2013
Address correspondence to Leonard Nunney, Leonard.Nunney@ucr.edu.
Copyright © 2014, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.02920-13

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

power; although the introgression test (11) has improved this situation (see below). Another approach is to test loci sequenced from 2 or more taxa and use the genetic partitioning program STRUCTURE (23, 24). Alleles that cannot be confidently allocated to one or more of the taxa are likely to be mosaics generated by recombination (4, 25).

To link recombination to adaptive change, it is useful to study a system in which recombination is limited, recognizable, and likely to lead to novel adaptation. Arnold et al. (26) recently made an interesting link between the acquisition of novel adaptations in bacteria via LGT and that via hybridization in metazoans. Excellent examples of how interspecific introgression can result in adaptation to new environments in higher plants are given in the work of Rieseberg and colleagues on the effects of introgression in sunflower species (for example, see reference 27). However, it is not only metazoans that hybridize: bacterial homologous recombination can sometimes result in interspecific introgression (e.g., in *Vibrio* spp.) (28). Interspecific hybridization of this kind is likely to be relatively rare, suggesting that the ideal study system is one with a significant frequency of homologous recombination between well-defined groups within a species (such as serotypes or subspecies). This level of study appears most likely to provide valuable insights into recombination-related adaptive change in pathogens. For example, Didelot et al. (29) showed that two human-pathogenic forms of *Salmonella enterica* (Typhi and Paratyphi A) are relatively dissimilar across about 75% of their genomes but show marked convergence across the rest. They concluded that this similarity reflects adaptation to the human host, driven by homologous recombination and selection. Similarly, Sheppard et al. (25) proposed that human activity (e.g., agriculture) has probably led to an increase in recombination between *Campylobacter jejuni* and *Campylobacter coli* and may have also created novel environments that have favored the evolution of hybrids.

Another species in which homologous recombination between closely related but distinct taxa has been documented is the plant-pathogenic bacterium *Xylella fastidiosa* (11, 30–32). *X. fastidiosa* is a xylem-limited bacterium that is transmitted by xylem-feeding insects, typically leafhoppers, and is divided into four subspecies: *fastidiosa*, *sandyi*, *multiplex*, and *pauca* (33, 34). These subspecies have diverged genetically by 1 to 3%, apparently due to their geographical isolation over about the last 20,000 to 50,000 years (11, 34, 35). This isolation has now broken down, due presumably to human activity (11, 31). The cooccurrence of the previously allopatric subspecies has resulted in intersubspecific homologous recombination (IHR), recombination that can be detected relatively easily due to the preexisting genetic divergence of the subspecies (11, 30–32, 35). Consistent with these observations, recent experimental work has confirmed that *X. fastidiosa* is transformationally competent (36) and that some isolates carry a conjugative plasmid (37).

X. fastidiosa is known to infect a wide range of hosts, causing scorch and dwarfing diseases (38, 39). In citrus, it causes citrus variegated chlorosis (CVC), a disease restricted to South America, and in grapevines in the United States and Central America, it causes Pierce's disease. In the United States, it also causes disease in almond, apricot, plum, peach, alfalfa, pecan, and blueberry. However, individual *X. fastidiosa* strains are not generalists. The different subspecies infect a characteristic and largely nonoverlapping range of plant hosts, and even within subspecies, different genotypes show differences in host specificity (35). For example,

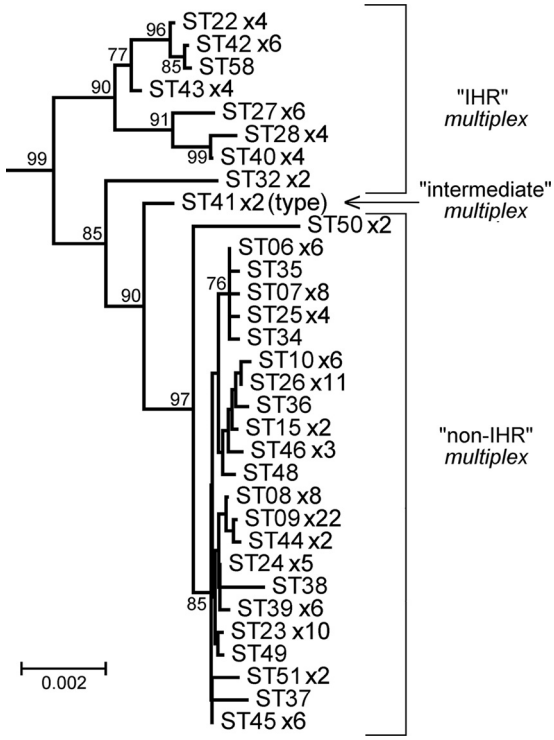

FIG 1 Distance tree showing variation within *X. fastidiosa* subsp. *multiplex* based on 31 IHR isolates (8 STs), 2 intermediate isolates (1 ST), and 110 non-IHR isolates (23 STs) using sequence data from the 7 MLST loci plus *pilU*. The IHR and intermediate forms together define the "recombinant group" of *X. fastidiosa* subsp. *multiplex* isolates. Where an ST included more than one *pilU* allele, the allele most frequently associated with the ST was used in constructing the tree. The number of isolates per ST (*N*) is shown by x*N* when *N* is >1. The tree is modified from that of Nunney et al. (35), where it was rooted by *X. fastidiosa* subspp. *fastidiosa*, *sandyi*, and *pauca*.

in the South American *X. fastidiosa* subsp. *pauca*, citrus isolates do not typically grow in coffee and vice versa (40, 41), and in *X. fastidiosa* subsp. *multiplex*, Nunney et al. (35) found associations between the genotype and host plant.

In their study of *X. fastidiosa* subsp. *multiplex*, Nunney et al. (35) used the multilocus sequence typing (MLST) protocol of Yuan et al. (31) to categorize 143 isolates. The MLST protocol is valuable for gaining insight into the evolutionary history and genetic diversity of taxa (42). MLST groups isolates into sequence types (STs), where each ST defines a unique set of alleles across the loci used (43). Based on 8 loci, 31 of these isolates (8 STs) were identified as IHR forms (since 2 or more of 8 loci sequenced showed evidence of intersubspecific introgression), and 2 isolates (1 ST) were considered "intermediate" (1 IHR allele), while the remaining 110 non-IHR isolates (23 STs) showed no evidence of introgression. The IHR and intermediate types together were considered to define the "recombinant" group of *X. fastidiosa* subsp. *multiplex* isolates (35) (Fig. 1). Most (all except ST58) were observed more than once, and 5 were found in two different U.S. states or districts (see Table 1 in reference 35).

The analysis of Nunney et al. (35) was focused on the evolution and host range of *X. fastidiosa* subsp. *multiplex*. For this purpose, it was necessary to identify and exclude isolates whose recent evolution (and potentially whose host affinity) was influenced by intersubspecific recombination. As such, once the 23 non-IHR STs

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

were identified, there was no further analysis of the remaining recombinant group STs. In particular, no evidence was presented for classifying some alleles as atypical of *X. fastidiosa* subsp. *multiplex* beyond the observation that they were never found in the non-IHR group (i.e., no genetic analysis of any of the recombinant alleles was presented).

Nunney et al. (35) did observe one intriguing pattern when they compared their results to those of Parker et al. (44). Of the 143 isolates, 13 (7 non-IHR-type and 6 IHR-type isolates) were also used in the study by Parker et al. (44), in which typing was based on a different set of 9 loci. Unexpectedly, these 13 isolates maintained the same grouping with the IHR and non-IHR types corresponding, respectively, to the clade A and clade B groupings (44). This highly statistically significant concordance (35) strongly suggested that IHR is not distributed randomly across all *X. fastidiosa* subsp. *multiplex* isolates but instead is restricted to a small subset, while the remainder is little influenced by IHR. However, Parker et al. (44) failed to find evidence of intersubspecific recombination within any of the *X. fastidiosa* subsp. *multiplex* isolates, despite applying a series of 9 tests designed to detect recombination contained within the RDP4 program (45) and the PHI program (22). This result presented a strong argument against our hypothesis that clade A members cluster because they are recombinant types carrying alleles derived from IHR (35). Here we reexamined the sequence data obtained in their study (44) by using the more sensitive introgression test (11) to determine if their tests missed evidence of IHR and, if so, whether it was confined to clade A. A second related question concerned the relationship among the recombinant IHR group members. In particular, what could be concluded about the origin of the group given the observation from 2 independent studies (35, 44) that the members appear to form a well-defined cluster of genotypes? Third, we used the sequence data to examine the hypothesis that the introgressed DNA was from *X. fastidiosa* subsp. *fastidiosa*, the subspecies that causes Pierce's disease. *X. fastidiosa* subsp. *fastidiosa* is native to Central America, and all known isolates in the United States and northern Mexico can be traced back to a single introduced genotype (32).

IHR would be of limited interest if it simply randomized the genetic differences among the subspecies but had a minimal effect on pathogenesis. For this reason, we were particularly interested in documenting any possible invasion of new plant hosts associated with IHR. The hypothesis is that IHR creates a range of novel genotypes that are far more variable than can arise from a lineage diversifying through point mutations, and this diversity facilitates adaptive evolution of a kind not possible for a clonal lineage. This kind of probabilistic evolutionary hypothesis can rarely be directly proven based on an individual case; however, it makes predictions that, if generally supported, would cause the hypothesis to be accepted. In the case of *X. fastidiosa*, compelling evidence supporting the hypothesis would be the invasion of a new native host plant that is uniquely associated with IHR. Our data support this hypothesis: in *X. fastidiosa* subsp. *multiplex*, IHR is indeed associated with the invasion of at least 2 new native plant hosts, blueberry and blackberry.

MATERIALS AND METHODS

To investigate intersubspecific homologous recombination (IHR), we analyzed 31 isolates previously identified as IHR-type and 2 isolates previously identified as intermediate-type *X. fastidiosa* subsp. *multiplex* (35), based on sequence of the 7 housekeeping loci used in the MLST scheme defined by Yuan et al. (31) plus a region of the *pilU* (cell surface) gene. Together, these 33 isolates made up the recombinant group. Details regarding the isolation (where appropriate) and typing of the 33 isolates were provided by Nunney et al. (35), and a summary of salient features is provided in Table S1 in the supplemental material in that article. All sequences used have previously been published and are available both in GenBank (see reference 35 for accession numbers) and the MLST website (http://pubmlst.org).

To detect IHR, we employed a modified version of the introgression test developed by Nunney et al. (11). In its original form, the test compares a set of target sequences, some of which may have been involved in IHR, to a set of potential donor sequences. Each variable site is classified as F, a fixed difference between the target sequences and the donor sequences, or P, a polymorphic site within the target sequences where at least one variant base is shared with the donor set. A significant shift in the ratio of F to P marks a recombination breakpoint. In the modified version of the test, the targeted introgression test, the target sequence is known *a priori* (in this case a member of the recombinant group) and is compared to two references, the donor group, D (in this case *X. fastidiosa* subsp. *fastidiosa*), and the ancestral group, A (in this case, *X. fastidiosa* subsp. *multiplex*). The minimum number of nucleotide differences between the target and the two references defines a ratio of D to A equivalent to the ratio of F to P and can be tested in the same way (see equation 1 in reference 11).

In some cases, there is no breakpoint because the whole locus appears to be an introgressed sequence (i.e., A $\gg$ D throughout the sequence). Although the signal of introgression across the entire sequenced region may be clear, it is valuable to have a statistical test that documents the strength of the signal. In this case, the null expectation (no introgression) is the ratio that reflects the pairwise differences between the donor and ancestral group versus the pairwise differences within the ancestral group (i.e., the ratio of gene diversities π_{total} versus π_{within}). We used this ratio to define the expectation of the D/A ratio for a chi-square test of complete introgression.

Gene diversity and distance trees were calculated using MEGA5 (46), and the maximum parsimony tree was created using the PARS program in Phylip (47, 48). Distance trees (using neighbor joining) and the maximum parsimony tree (using allele numbers as characters) were used rather than other methods, given the known occurrence of intersubspecific recombination in the data. ClonalFrame (20) was used to provide an independent estimate of the relative importance of recombination versus mutation in the recombinant group.

RESULTS

Characterization of the recombinant alleles. Based on 8 loci sequenced (7 MLST loci plus *pilU*), Nunney et al. (35) identified 9 sequence types (STs) belonging to the recombinant group of *X. fastidiosa* subsp. *multiplex*. These STs all showed evidence of intersubspecific homologous recombination (IHR) at one or more of the 8 loci and were characterized by 18 alleles, 10 of which were never found in non-IHR *X. fastidiosa* subsp. *multiplex* strains (35). These 10 alleles were examined for evidence of IHR (Table 1) by comparing them to the previously described non-IHR *X. fastidiosa* subsp. *multiplex* alleles (35) and to the known *X. fastidiosa* subsp. *fastidiosa* and *sandyi* alleles (31, 32). Of these 10, 4 alleles were found to be derived in their entirety from *X. fastidiosa* subsp. *fastidiosa*, and 3 were found to be chimeric for *X. fastidiosa* subsp. *multiplex* and *fastidiosa* sequences, with significant evidence of one or more recombination breakpoints. These 7 alleles encompassed 4 loci: *leuA*, *cysG*, *holC*, and *pilU*.

The locus most strongly implicated in IHR was *cysG*, since all of the 9 recombinant-group STs were characterized at this locus by 1 of 3 *cysG* alleles (no. 6, 12, and 18) unique to the group. The involvement of IHR in the genesis of all 3 of these alleles is illustrated by their close genetic relationship to *X. fastidiosa* subsp.

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

TABLE 1 Recombination analysis of all of the alleles found in the recombinant group of *X. fastidiosa* subsp. *multiplex* that were not found in the non-IHR group

Allele	*X. fastidiosa* allele no. (bp difference)[a]		Region(s) of IHR[b]	Total length (bp)	Ratio(s)[c]	*P* value
	subsp. *multiplex*	subsp. *fastidiosa*				
holC7	3 (8)	19 (4)	**1 to (183 to 286)**	379	8:0 vs 0:4	0.004
cysG6	3 (7)	12 (7)	**1 to (23 to 46) and (435 to 481) to (516 to 524)**	600	3:0 vs 0:5 vs 4:0 (vs 0:2)[d]	0.057, 0.057 combined = 0.022[e]
cysG12	3 (14)	12 (0)	All (**1 to 600**)	600	14:0	<0.001[f]
cysG18	5 (8)	14 (3)	**(48 to 258) to 600**	600	0:3 vs 8:0	0.012
pilU1	3 (17)	1 (0)	All (**1 to 545**)	545	17:0	<0.001[f]
pilU9	3 (18)	1 (1)	All (**1 to 545**)	545	18:1	<0.001[f]
leuA4	3 (8)	9 (2)	All (**1 to 708**)	708	8:2	<0.001[f]
leuA6	3 (2)	9 (6)	None (see the text)	708	2:6	NS[g]
nuoL4	3 (3)	5 (6)	None (see the text)	557	3:6	NS
holC9	4 (2)	17 (5)	(290 to 340) to 379	379	0:5 (vs 2:0)[d]	NS (0.10)

[a] The number of base pair differences between the recombinant group allele and the most similar allele from non-IHR subsp. *multiplex* and from subsp. *fastidiosa* is shown as the designation number of the most similar allele followed by the number of base pair differences in parentheses.
[b] Regions of suspected intersubspecific homologous recombination (IHR) are identified and in boldface when statistically significant ($P < 0.05$) based on one of two analyses (see Materials and Methods): (i) recombination breakpoint analysis using the targeted introgression test, or (ii) a ratio test when the whole region appears to have been introgressed.
[c] Ratios are shown in the format "differences from the most similar *X. fastidiosa* subsp. *multiplex* allele:differences from the most similar *X. fastidiosa* subsp. *fastidiosa* allele" (see footnote *a*) expressed 5′ versus 3′ of the suspected breakpoint.
[d] A reversal of ratios at the end of a sequence sometimes lacks power but is suggestive of a breakpoint.
[e] Using Fisher's combined probability test, $\chi^2_{2k} = -2\Sigma \ln(p_i)$, where the summation i is over the k tests.
[f] Based on a ratio test comparison of the observed ratio to a 1:11 ratio, reflecting the estimated values $\pi_{within} = 0.001$ and $\pi_{total} = 0.011$ (see Materials and Methods).
[g] NS, not significant.

fastidiosa and *sandyi* alleles (see alleles shown in boxes in Fig. 2). Allele 12, apart from being found in the recombinant group, is an *X. fastidiosa* subsp. *fastidiosa* allele (32). The other two alleles were found to be chimeric: allele 18 contains a single recombinant region at the 3′ end of (at a minimum) 342 bp, while allele 6 has two short recombinant regions, one at the 5′ end of at least 23 bp and another toward the 5′ end of at least 35 bp (Table 1). The DNA sequence variation defining these patterns is shown in Table 2.

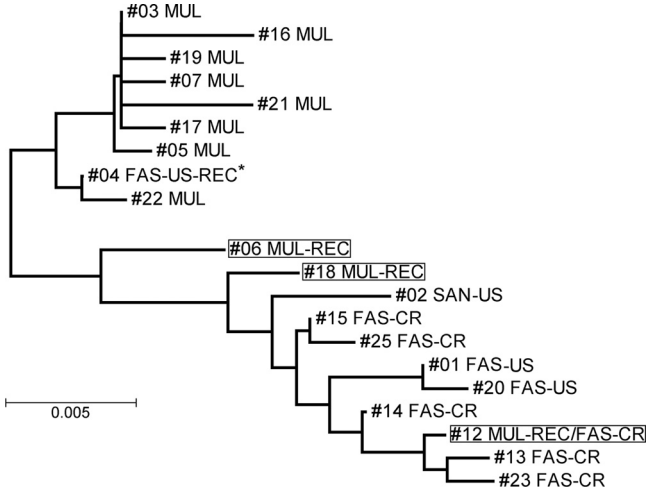

FIG 2 Distance tree based on the sequence data of the *cysG* MLST locus showing the close relationship of the 3 *X. fastidiosa* subsp. *multiplex* recombinant group alleles (boxed) to those found in the other subspecies found in the United States. The tree shows comparison of all 11 alleles identified from *X. fastidiosa* subsp. *multiplex* isolates (MUL) to all published alleles of *X. fastidiosa* subspp. *fastidiosa* (FAS) and *sandyi* (SAN). The tree is unrooted. Alleles derived from intersubspecific homologous recombination (REC) include the *X. fastidiosa* subsp. *multiplex* recombinant 12, which was identical to an *X. fastidiosa* subsp. *fastidiosa* allele from Costa Rica (CR). Allele 04 (marked with *) was found in *X. fastidiosa* subsp. *fastidiosa* in the United States (US) and previously identified as resulting from intersubspecific homologous recombination (35).

The patterns seen in the DNA sequences of the 3 *cysG* alleles are consistent with the hypothesis of a single IHR that introgressed donor allele 12 into *X. fastidiosa* subsp. *multiplex*, followed by subsequent intrasubspecific recombination reintroducing *X. fastidiosa* subsp. *multiplex* sequence to create alleles 6 and 18 (Table 2). There are no inconsistent sites, provided the 5′ recombination breakpoint in allele 18 starts (as shown) after position 71.

Introgression of *X. fastidiosa* subsp. *fastidiosa* sequence into *X. fastidiosa* subsp. *multiplex* was found in alleles at 3 other loci (*leuA*, *holC*, and *pilU*) (Table 1). In the case of *pilU*, 7 of the 9 recombinant STs carried either an allele identical to a known *X. fastidiosa* subsp. *fastidiosa* allele (allele 1) or 1 bp different from it (allele 9). Allele 1 is an allele that characterizes most U.S. isolates as well as several STs found in Costa Rica, while allele 9 is unique to the recombinant group.

The *leuA* locus has a single statistically significant recombinant allele, allele 4 (*leuA4*). It differed by 2 bp from the *X. fastidiosa* subsp. *fastidiosa* allele 9 but by 8 bp from the most similar non-IHR *X. fastidiosa* subsp. *multiplex* allele. *X. fastidiosa* subsp. *fastidiosa* allele 9 could be the donor for its entirety (Table 1), although if the recombination region started after (i.e., 3′ of) site 10 but before position 520 (Table 3), then only one site would be unexplained. That remaining site (position 550) carries a base unique to this allele and is probably a novel mutation. If the recombination breakpoint was 3′ of position 295 then *X. fastidiosa* subsp. *fastidiosa* allele 1 provides as good a match as allele 9 (Table 3).

Similarly, *holC* allele 7 was also 8 bp different from the most similar non-IHR *X. fastidiosa* subsp. *multiplex* allele, providing clear evidence that the 5′ end was derived from *X. fastidiosa* subsp. *fastidiosa* (Table 4). The pattern can be explained if *X. fastidiosa*

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

TABLE 2 Regions of IHR in the *X. fastidiosa cysG* allele unique to the recombinant group isolates of *X. fastidiosa* subsp. *multiplex*[a]

	Base at position in *cysG* MLST sequence[b]:																
	0	0	0	0	0	2	2	4	4	4	4	4	4	5	5	5	5
	1	2	2	4	7	5	6	0	3	3	8	8	9	1	2	6	8
cysG allele	2	1	2	7	1	9	6	7	4	5	2	9	7	5	5	0	8
Recombinant allele 12	A	A	C	G	A	G	T	C	C	G	T	A	T	T	G	A	G
Recombinant allele 18	G*	C	T*	A*	C	G	T	C	C	G	T	A	T	T	G	A	G
Recombinant allele 6	A	A	C	A*	C	A*	C*	C	T*	G	T	A	T	T	A*	C*	G
X. fastidiosa subsp. *multiplex* allele 3/5	G	C	T	A	C	A	C	C	T	G	C	G	C	C	A	C/A	G
Potential donors																	
X. fastidiosa subsp. *fastidiosa* allele 12 (C.R.)	A	A	C	G	A	G	T	C	C	G	T	A	T	T	G	A	G
X. fastidiosa subsp. *fastidiosa* allele 1 (U.S.)	A	C	C	G	C	G	T	G	C	G	T	A	C	T	G	A	A
X. fastidiosa subsp. *sandyi* allele 2	G	C	C	G	A	G	T	C	C	A	T	A	C	T	G	C	A

[a] The alleles are compared to the most similar *X. fastidiosa* subsp. *multiplex* alleles (which define the shaded bases), to the most similar *X. fastidiosa* subsp. *fastidiosa* allele found in the United States (U.S.), to the *X. fastidiosa* subsp. *sandyi* allele (also from the United States), and to the most similar *X. fastidiosa* subsp. *fastidiosa* allele found in Costa Rica (C.R.) but not the United States. Unshaded bases define sites different from the *X. fastidiosa* subsp. *multiplex* allele, while bases marked with asterisks are inconsistent with *X. fastidiosa* subsp. *fastidiosa* sequence. Underlined bases are unique to *X. fastidiosa* subsp. *fastidiosa* and/or *sandyi*.

[b] The Costa Rica *X. fastidiosa* subsp. *fastidiosa* allele 12 is identical to recombinant allele 12. The italic regions show potential regions of secondary recombination involving the reintroduction of *X. fastidiosa* subsp. *multiplex* sequence into the *cysG12* IHR background (see Table 1 for the breakpoint analysis).

subsp. *fastidiosa* allele 19 is the donor of the 5′ region ending somewhere between positions 183 and 286, since it leaves no inconsistent bases (Table 4).

The loci *leuA* and *holC* each had an additional allele (no. 6 and 9, respectively) that were unique to the recombinant group, as was an allele at another locus, *nuoL4*. Although these last 3 alleles did not show statistically significant evidence of introgression (Table 1), they all showed a grouping of 2 or 3 nucleotide changes that were not found in non-IHR *X. fastidiosa* subsp. *multiplex* but were present in *X. fastidiosa* subsp. *fastidiosa*. Of these 3, the strongest case for IHR is *holC9* ($P = 0.10$) (Table 1), where a region of possible IHR can be seen at the 3′ end of the sequence (Table 4).

Origins of the recombinant group. Based on the recombination analysis, it is possible to reconstruct some features of the ancestral *X. fastidiosa* subsp. *fastidiosa* strain that contributed to the variation seen in the recombinant *X. fastidiosa* subsp. *multiplex*. Of particular interest is whether the most parsimonious reconstruction involves more than one donor genotype of *X. fastidiosa* subsp. *fastidiosa*. There is enough information from the 4 of the 8 loci that showed significant IHR to address this question, using the introgression patterns summarized in Table 1.

To examine the possibility the recombinant group was derived primarily from a single *X. fastidiosa* subsp. *multiplex* recipient genotype and a single *X. fastidiosa* subsp. *fastidiosa* donor genotype, we defined both the most likely *X. fastidiosa* subsp. *multiplex* recipient genotype and the most likely *X. fastidiosa* subsp. *fastidiosa* donor genotype based on the alleles found in the recombinant group (35). These alleles and their ST associations are summarized in Fig. 3. For the recipient genotype, the simplest hypothesis is to assume that the non-IHR *X. fastidiosa* subsp. *multiplex* alleles present in the recombinant group are ancestral. These are *leuA3* (STs 41 and 43), *petC3* (all STs), *malF5* (all STs), *cysG?* (no data), *holC4* (all STs except 27, 28, and 40), *nuoL3* (all STs except 28 and 32), *gltT3* or *-7* (all STs), and *pilU3* (STs 22 and 41). These data define a single allele at 6 of the 8 loci examined, with a 7th locus (*gltT*) defined by two possibilities and only *cysG* undefined.

Similarly, we can define the most likely donor genotype at 4 loci: *leuA9*, *cysG12*, *holC19*, and *pilU1* (Tables 2 to 4). *nuoL4*, although unique to the recombinant group, was not established as a recombinant allele, so no ancestral donor can be proposed, and the remaining 3 loci (*petC*, *malF*, and *gltT*) are also undefined since the members of the recombinant group carry no *X. fastidiosa*

TABLE 3 Regions of IHR in the *leuA* allele unique to the recombinant group isolates of *X. fastidiosa* subsp. *multiplex*[a]

	Base at position in *leuA* MLST sequence:													
	0	1	2	2	3	5	5	5	5	5	5	5	6	7
	1	0	2	9	0	2	2	2	4	5	6	7	3	0
leuA allele	0	0	0	5	1	0	2	9	8	0	8	1	1	3
Recombinant allele 4	C*	A	C	C	G	C	G	C	G	C*	T	C	T	A
Unique allele 6	C*	A	C	C	G	C	G	C	A*	T	G*	T	C*	C*
X. fastidiosa subsp. *multiplex* allele 3	C	A	C	C	G	T	A	C	A	T	G	T	C	C
Potential donors														
X. fastidiosa subsp. *fastidiosa* allele 9 (C.R.)	T	A	C	C	G	C	G	C	G	T	T	C	T	A
X. fastidiosa subsp. *fastidiosa* allele 1 (U.S.)	T	A	C	T	G	C	G	C	G	T	T	C	T	A
X. fastidiosa subsp. *sandyi* allele 2	T	G	T	C	A	C	G	T	G	T	T	C	T	A

[a] The alleles are compared to the most similar *X. fastidiosa* subsp. *multiplex* alleles (which define the shaded bases), to the most similar *X. fastidiosa* subsp. *fastidiosa* allele found in the United States (U.S.), to the *X. fastidiosa* subsp. *sandyi* allele (also from the United States), and to the most similar *X. fastidiosa* subsp. *fastidiosa* allele found in Costa Rica (C.R.) but not the United States. Unshaded bases define sites different from the *X. fastidiosa* subsp. *multiplex* allele, while bases marked with asterisks are inconsistent with *X. fastidiosa* subsp. *fastidiosa* sequence. Underlined bases are unique to *X. fastidiosa* subsp. *fastidiosa* and/or *sandyi*.

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

TABLE 4 Regions of IHR in the *X. fastidiosa holC* allele unique to the recombinant group isolates of *X. fastidiosa* subsp. *multiplex*[a]

	Base at position in *holC* MLST sequence:															
	0	0	0	0	0	0	0	0	0	1	1	1	2	3	3	3
	0	2	5	7	7	8	8	8	8	1	7	8	8	1	2	7
holC allele	9	7	6	7	8	3	6	7	9	0	6	2	7	7	3	2
Recombinant allele 7	T	G	G	T	A	A	T	A	G	G	A	T	C*	C*	C*	A
Unique allele 9	T	G	A	C	C	G	C	G*	G	A	A	C*	C*	T	T	G
X. fastidiosa subsp. *multiplex* 3	T	G	A	C	C	G	C	G	G	A	A	C	C	C	C	A
Potential donors																
X. fastidiosa subsp. *fastidiosa* allele 19 (C.R.)	T	G	G	T	A	A	T	A	G	G	A	T	T	T	T	G
X. fastidiosa subsp. *fastidiosa* allele 1 (U.S.)	G	G	A	C	C	G	C	A	A	G	A	T	T	T	T	G
X. fastidiosa subsp. *sandyi* allele 2	T	A	G	T	A	A	T	A	G	G	G	C	T	T	C	G

[a] The alleles are compared to the most similar *X. fastidiosa* subsp. *multiplex* alleles (which define the shaded bases), to the most similar *X. fastidiosa* subsp. *fastidiosa* allele found in the United States (U.S.), to the *X. fastidiosa* subsp. *sandyi* allele (also from the United States), and to the most similar *X. fastidiosa* subsp. *fastidiosa* allele found in Costa Rica (C.R.) but not the United States. Unshaded bases define sites different from the *X. fastidiosa* subsp. *multiplex* allele, while bases marked with asterisks are inconsistent with *X. fastidiosa* subsp. *fastidiosa* sequence. Underlined bases are unique to *X. fastidiosa* subsp. *fastidiosa* and/or *sandyi*.

subsp. *fastidiosa* sequence at these loci. There is ambiguity at the *pilU* locus, since the recombinant allele 9 also occurs; however, only allele 1 has been found in *X. fastidiosa* subsp. *fastidiosa*, so this allele was assumed to be ancestral, with allele 9 subsequently arising by mutation.

Evaluation of the plausibility of a single initial IHR event is complicated by the possibility of subsequent intrasubspecific recombination both within the recombinant group and between the recombinant group and the dominant non-IHR *X. fastidiosa* subsp. *multiplex* strains. Plausible sets of recombination events were determined by creating a tree using maximum parsimony

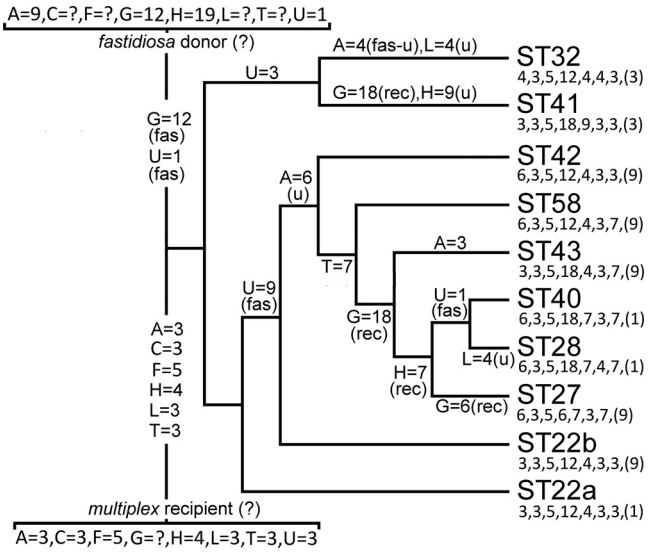

FIG 3 A maximum parsimony tree of the recombinant group STs showing a possible evolutionary trajectory from the *X. fastidiosa* subsp. *fastidiosa* donor that maximizes the fit to all of the IHR regions (Table 1) and from the potential ancestral *X. fastidiosa* subsp. *multiplex* recipient (ST45). The gene names A, C, F, G, H, L, T, and U refer to the MLST loci *leuA*, *petC*, *malF*, *cysG*, *holC*, *nuoL*, and *gltT* and the non-MLST *pilU*. The numbers under each ST name are its defining alleles (plus *pilU* added parenthetically). Within the tree, allele numbers following a gene name show substitutions with added annotation defining the allele type: "rec" indicates evidence of an internal recombination break, "fas" labels alleles wholly derived from *X. fastidiosa* subsp. *fastidiosa*; and "u" denotes alleles unique to the recombinant group. Nonannotated alleles were found in non-IHR *X. fastidiosa* subsp. *multiplex*.

applied to the 10 8-locus genotypes (the 9 STs, with ST22 divided into "a," with *pilU1*, and "b," with *pilU9*). Using allele numbers as characters, there were 2 equally parsimonious trees, each with 14 steps. They differed only in the precise positioning of 22a (which remained 1 step away from 22b in both); however, assuming a basal introgression of *pilU1*, only the tree shown in Fig. 3 remained. The hypothetical donor and recipient genotypes were added to root the tree, with the tree dictating *gltT3* (rather than *gltT7*) in the ancestral recipient genotype.

The most parsimonious tree (Fig. 3) showed that the pattern of introgression was more complex than could result from a single IHR. There are four main events that illustrate this complexity. First, based on this tree, the grouping of STs 27, 28, and 40 is defined by the introgression of *holC7*, a recombinant allele introduced into the tree far from the basal recombination event. Second, although the mutation of *pilU1* (introduced in the basal recombination) could explain the appearance of *pilU9*, a second introgression of *pilU1* would be necessary to account for its appearance in STs 28 and 40. Third, a number of events are necessary to account for the evolution of the *cysG* locus. While *cysG12*, an *X. fastidiosa* subsp. *fastidiosa* allele introduced in the initial recombination event, could give rise to *cysG18* by the introgression of *X. fastidiosa* subsp. *multiplex* sequence (Table 2), this allele appears in two places in the tree, necessitating a lateral transfer within the recombinant group.

Despite this complexity, the hypothesis of a single primary IHR event creating the founder of the recombinant group is strongly supported by the pattern seen at the *cysG* locus. As noted above, all members of the recombinant group share one of 3 alleles that appear to be derived from a single introgression of donor allele 12.

Analysis of the *X. fastidiosa* subsp. *fastidiosa* donor. The proposed *X. fastidiosa* subsp. *fastidiosa* donor is defined at 4 of the 8 loci: *leuA9*, *cysG12*, *holC19*, and *pilU1*. Of these 4 alleles, only *pilU1* was found in an extensive genetic survey of 86 isolates of *X. fastidiosa* subsp. *fastidiosa* within the United States and northern Mexico (31). The results of this survey, combined with similar genetic data from Costa Rica, led to the conclusion that all isolates of *X. fastidiosa* subsp. *fastidiosa* found in North America were derived from a single ancestral strain introduced from Central America (32). Consistent with this hypothesis was the observation that, in the North American isolates, no allele at the 7 MLST loci or

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

the *pilU* locus was more than 1 bp different from the most common allele.

Given this background, we can examine the hypothesis that the proposed ancestral donor is consistent with the *X. fastidiosa* subsp. *fastidiosa* strains currently found in the United States. That allele *pilU1* is found in the United States is consistent with this view. Similarly, at *leuA* there is no inconsistency with U.S. allele 1 if the recombination breakpoint in the recombinant allele 4 was after (3′) position 295 (discounting the derived base in allele 4 at position 550 [Table 3]). If the breakpoint is before that point, then Costa Rica allele 9 provides a better fit of only 1 bp, a minor difference. In marked contrast, the alleles *cysG12* (found in its entirety in the recombinant group) and *holC19* (matching the 5′ end of recombinant allele 7) have only been found in Costa Rica, not in the United States (28), and differ markedly from the U.S. alleles. In particular, within the IHR regions, the U.S. *X. fastidiosa* subsp. *fastidiosa* alleles *cysG1* and *holC1* are 5 and 7 bp different, respectively, from the recombinant group sequence, while the Costa Rica alleles precisely match the donor sequence (Tables 2 and 4). These large differences require us to reject the hypothesis that the primary *X. fastidiosa* subsp. *fastidiosa* donor was derived from the introduced genotype that was the ancestor of all of the North American *X. fastidiosa* subsp. *fastidiosa* isolates that have been typed.

Estimating recombination rates in the recombinant group of *X. fastidiosa* subsp. *multiplex*. The prevalence of recombination over mutation in the evolution of the recombinant group was supported by a ClonalFrame analysis: the estimated ratio of recombination events to mutation (ρ/θ) was 19,310, with a 95% confidence lower bound of 45.3. Addition of the potential ancestor and/or potential donor genotypes (Fig. 3) to the analysis maintained high estimates of the lower bound of ρ/θ (54.7 with the ancestor only, 8.2 with the donor only, and 13.0 with both). These lower bounds are high for a largely clonal organism, and they illustrate the pervasive involvement of recombination in the genesis of the recombinant group. They contrast markedly with the mean estimate of $\rho/\theta = 0.02$ with an upper bound of 0.06 obtained for the non-IHR group (36).

Recombination in clade A. As noted earlier, Nunney et al. (35) proposed that the two clades A and B in the *X. fastidiosa* subsp. *multiplex* tree of Parker et al. (44) corresponded to isolates that had been subject to IHR (clade A) versus those largely free of IHR (clade B). However, this hypothesis is undermined by the absence of evidence of intersubspecific introgression from *X. fastidiosa* subsp. *fastidiosa* to *X. fastidiosa* subsp. *multiplex* in the analysis of clade A by Parker et al. (44). However, in our reanalysis of these data using the targeted introgression test (see Materials and Methods), we found that out of the 9 loci, 6 (*copB*, *cvaC*, *fimA*, *pslA*, *rpfF*, and *xadA*) showed statistically significant evidence of IHR in alleles found in the clade A isolates (Table 5). Furthermore, in this reanalysis, all of the 6 clade A genotypes (10 isolates) carried recombinant alleles in at least 4 of these loci, while none of the 9 clade B genotypes (11 isolates) showed any similar evidence of IHR.

As noted previously, the 6 recombinant group isolates common to both studies were all classified into group A (35). They encompassed 4 of the genotypes identified by Parker et al. (44): Almond2 (ALS0022, ST27), the Blueberry1 group (BB0385, BB0387, and BB0493, all ST43), the Blueberry2 group (BB0488, ST42), and the Ragweed group (AT0166, ST42). They showed significant evidence of IHR at 4 or 5 of the 9 loci (Table 5), so that in total they showed significant evidence of IHR in 6 or 7 out of the 17 genes examined.

Plant hosts of the recombinant group. Given the identification of a discrete recombinant subgroup within *X. fastidiosa* subsp. *multiplex*, we investigated the possibility that it corresponded with a shift in the plant hosts that were infected. Among the 33 isolates defining 9 STs, there were 5 plant hosts represented more than once: ST27 on almond (*Prunus dulcis*), ST27 and ST40 on purple leaf plum (*Prunus cerasifera*), ST22, ST28, ST42, and ST58 on giant ragweed (*Ambrosia trifida*), ST32 on blackberry (*Rubus fruticosus*), and ST42 and ST43 on blueberry (*Vaccinium corymbosum* and *V. corymbosum* × *Vaccinium angustifolium* hybrid), to which we can add American elm (*Ulmus americana*) by including the 4 additional clade A isolates from Parker et al. (44). Closer examination of *X. fastidiosa* subsp. *multiplex* isolates from these host plants, using the 143 isolates subjected to MLST (31), showed that almond and purple leaf plum isolates (all from California) were recombinant types only about one-quarter of the time (almond, 25% [3/12]; purple leaf plum, 28% [4/14]). The ratios in American elm from Washington, DC (and Alabama [44]), and western ragweed from Texas were substantially higher (67% [2/3] and 78% [7/9], respectively), but it was only in blackberry (2 isolates), and blueberry (7 isolates, plus 3 more from reference 44) that 100% of the isolates were recombinant. While the sample size for blackberry was very small, the isolates were geographically separate (Florida and North Carolina) and defined a single sequence type (ST32) that was found on no other host. ST32 differed from all other STs in the recombinant group, except ST41, in carrying the nonrecombinant *pilU3* allele (35). Blueberry isolates were better represented and again were isolated from two different states, Florida and Georgia. The 7 isolates that we typed defined two STs: ST43, which was found in both Florida (3 isolates) and Georgia (1 isolate) and was unique to these blueberry isolates, and ST42, which was isolated in Georgia on blueberry but also in Texas on western ragweed (plus a single isolate on western soapberry).

DISCUSSION

Analysis of the recombinant group of *X. fastidiosa* subsp. *multiplex* showed three important results. First, intersubspecific recombination (IHR) was shown to have occurred in 50% of 8 loci scattered throughout the genome that were chosen independently of the data (7 housekeeping loci for MLST and one nonhousekeeping cell surface locus for comparison) (30). Second, it was shown that the donor of the introgressed sequence was *X. fastidiosa* subsp. *fastidiosa*, a subspecies introduced from Central America into the United States as a single strain (32). However, the introgressed sequence at two of the loci did not come from any of the *X. fastidiosa* subsp. *fastidiosa* genotypes that have been found in the United States. This result suggests that another introduction of *X. fastidiosa* subsp. *fastidiosa* must have occurred, an introduction that resulted in successful IHR, after which the donor genotype seems to have disappeared. This involvement of an unexpected *X. fastidiosa* subsp. *fastidiosa* strain supports the hypothesis that the members of the recombinant group share a single ancestral IHR event. Third, the hypothesis that IHR has facilitated a shift to new hosts is strongly supported by the example of blueberry, where 10 isolates have been typed (7 from the present study, plus 3 additional isolates from another study [44]) and potentially supported

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

TABLE 5 Characteristics of the alleles found in the clade A *X. fastidiosa* subsp. *multiplex* isolates of Parker et al. (44)[a]

Locus	Subgroup	*X. fastidiosa* allele subgroup (bp difference)[b]		Region of IHR	Maximum length (bp)	Ratio(s)	*P* value(s)
		subsp. *multiplex*	subsp. *fastidiosa*				
acvB	All	All (0)	All (45)	None		0:45	
copB	All	All (4)	All (2)	**(1 to 789)**[c]	789	4:2	<0.001[d]
cvaC	Rag, Elm	Oak1 (15)	All (0)	**1 to 285**	285	15:0	<0.001[d]
	BB1, -2, and -3	Plum (8)	All (9)	**(135 to 13) to (220 to 222)**	285	1:5 vs 7:0 vs 0:4	0.010, 0.015, combined = 0.0014
	Alm2	Oak1 (0)	All (15)	None	285	0:15	
fimA	All	Oak1 (4)	Ca05 (4)	**(268 to 378) to 506**	506	0:4 vs 4:0	0.029
gaa	All	Oak1 (0)	All (13)	None	1,064	0:13	
pglA	All (−Elm)	Lupine (8)	All (2)	**1 to 497**	497	8:2	<0.001[d]
	Elm	Lupine (6)	All (4)	1 to (374 to 395)	497	6:1 vs 0:3	0.067
pilA	BB1,-2, and -3; Rag	Plum (7)	All (49)	None	353	7:49	
	Alm2	Lupine (7)	All (48)	None	353	7:48	
	Elm	Oak1 (0)	All (54)	None	353	0:54	
rpfF	All (−Rag)	All (0)	All (36)	None	777	0:36	
	Rag	All (4)	All (32)	**1 to (94 to 158)**	777	4:1 vs 0:31	0.0002
xadA	BB1 and -3, Alm2	Redbud (9)	Gafl (14)	**(834 to 901) to 1060**	1,060	2:12 vs 7:2	0.0083
	BB2	Redbud (0)	Gafl (22)	None	1,060	0:22	
	Rag	Lupine (2)	Gafl (21)	None	1,060	2:21	
	Elm	Redbud (1)	Gafl (22)	None	1,060	1:21	

[a] Regions of suspected intersubspecific homologous recombination (IHR) in the alleles found in clade A based on our analysis are identified and in boldface. All differences were statistically significant (see *P* values). Parker et al. (44) divided clade A into 6 subgroups: Ragweed (Rag), Elm, Blueberry1 (BB1), BB2, BB3, and Almond2 (Alm2). Only Elm and BB3 are not represented in the present study (see the text).

[b] The number of base pair differences between each clade A allele and the most similar allele from their clade B *X. fastidiosa* subsp. *multiplex* isolates and from the *X. fastidiosa* subsp. *fastidiosa* isolates is shown in parentheses. For additional details, see Table 1.

[c] Excludes the minisatellite region at positions 69 to 356 (based on Elm).

[d] Based on a comparison of the observed ratio to a 1:11 ratio (see Table 1).

by the example of blackberry (based on 2 isolates, both defining an ST not found in any other host).

A link between the shift to a novel plant host and homologous recombination has not been previously identified. Of course, the direct causation of this link can never be proved without knowledge of the genetic changes driving this shift. It can always be argued that the link is fortuitous and that one or more point mutations in the nonrecombined *X. fastidiosa* subsp. *multiplex* genome are causal in the host shift. Arguing against this possibility are 2 additional pieces of information. First, both blueberry and blackberry are native to the United States, so if only a simple genetic change was required to infect these species, why did the native nonrecombinant *X. fastidiosa* subsp. *multiplex* apparently never acquire these changes? Second, a similar but even more extensive mixing of the genomes of *X. fastidiosa* subspp. *fastidiosa* and *multiplex* is found in the only form of *X. fastidiosa* that infects another U.S. native plant, mulberry (49; L. Nunney, E. L. Schuenzel, M. Scally, R. E. Bromley, and R. Stouthamer, unpublished data). Furthermore, in other bacterial species, it has been demonstrated that recombination can drive rapid evolution, both in the laboratory (18, 50) and, in the case of *Helicobacter pylori*, in mice (51). Similarly, McCarthy et al. (52) concluded that lineages of *Campylobacter jejuni* in chickens versus cattle and sheep were able to shift host type, because rapid adaptation was facilitated by recombination with the resident host population.

In the study by Nunney et al. (35), it was shown that the recombinant genotypes formed a well-defined group (Fig. 1), demonstrating that intersubspecific homologous recombination was not randomly distributed across the *X. fastidiosa* subsp. *multiplex* isolates. This work was based on a survey of 143 *X. fastidiosa* subsp. *multiplex* isolates using just 8 loci. There were 33 isolates that showed some evidence of IHR in at least 1 locus: all but 2 showed statistically significant evidence in at least 2 loci, while the remaining 110 showed no such evidence (35). The generality of this discrete group of recombinant forms was supported by our analysis presented here of the sequence data from 9 more loci sequenced by Parker et al. (44). These loci divided isolates into 2 groups (clades A and B) that appeared to correspond to the recombinant and non-IHR groups, respectively (35), even though Parker et al. (44) found no evidence of IHR. Upon reanalysis, we found statistically significant IHR in 6 of the 9 loci in the clade A data but no evidence of IHR in the clade B data. Clade A included 6 isolates that we had typed in the present study, and each of these showed evidence of IHR in 4 or 5 of the additional 9 loci. Thus, in two independent samplings that together examined 17 loci, there was clear evidence of substantial genomewide IHR in the recombinant group iso-

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

lates, amounting to 50% (4/8) of the genes showing IHR across the MLST locis plus the *pilU* locus (Table 1). The average was higher (67% [6/9]) when based on the loci sequenced by Parker et al. (44) (Table 5); however, this was probably biased upwards by the manner in which the loci were chosen (as rapidly evolving variable loci).

None of the IHR events in 6 of the 9 loci identified using the targeted introgression test, or in the case of complete introgression, a chi-square test, were detected by Parker et al. (44) using PHI (22) and the 9 tests implemented in RDP (45). This failure of the standard tests of recombination to detect IHR was previously noted by Nunney et al. (11), motivating the development of their introgression test.

We examined the hypothesis that the recombinant group STs were derived from a single IHR event involving a *X. fastidiosa* subsp. *multiplex* recipient and an *X. fastidiosa* subsp. *fastidiosa* donor. The distribution of allelic differences among the recombinant STs was consistent with them all being derived from a single initial event, but a small number of other intersubspecific and intrasubspecific recombination events would also be needed (Fig. 3). More importantly, the genotypes seen in the recombinant group can be accounted for entirely, or very nearly so, based on a single *X. fastidiosa* subsp. *fastidiosa* donor genotype. For example, the substantial variation in *cysG* (alleles 12, 18, and 6) can all be accounted for by an ancestral introgression of *X. fastidiosa* subsp. *fastidiosa* allele 12 followed by subsequent intrasubspecific recombination of *X. fastidiosa* subsp. *multiplex* sequence to form the other two alleles (Table 2). In contrast, variation at *pilU* could be accounted for by a second donor contributing the (so far undiscovered) *X. fastidiosa* subsp. *fastidiosa pilU9* allele, but it could also have arisen by a single mutation in *pilU1* unique to the recombinant group. A possible single *X. fastidiosa* subsp. *multiplex* recipient genotype was also identified (Fig. 3). This genotype is consistent with a known ST: setting (the undefined) *cysG* to allele 3 (*cysG3*) makes the recipient identical to ST45, which was sampled from the states of California, Kentucky, and Texas (35). Elsewhere, we consider a slightly different hypothesis regarding the origin of the recombinant group in which the donor and recipient subspecies are reversed—i.e., that it was derived from a single IHR event, but involving an *X. fastidiosa* subsp. *multiplex* donor and an *X. fastidiosa* subsp. *fastidiosa* recipient; however, apart from the role reversal, the conclusions are unaltered (Nunney et al., unpublished data).

The ancestral reconstruction allows us to consider the second question posed earlier: is the donor consistent with the *X. fastidiosa* subsp. *fastidiosa* genotypes found in the United States? The answer is very clearly "no." The original donor (or donors) carried *cysG12* and *holC19* (or another *holC* allele with an identical 183-bp 5′ end). These alleles are both found in isolates from Central America, but no *X. fastidiosa* subsp. *fastidiosa* isolate found in the United States comes close to matching this criterion: the most similar U.S. ST has a 12-bp mismatch. There has been extensive sampling of *X. fastidiosa* subsp. *fastidiosa* within the United States, based on 85 isolates sampled across the United States (California, Florida, Georgia, Kentucky, North Carolina, and Texas) from 15 different host plants (31). There is very little variation within *X. fastidiosa* subsp. *fastidiosa* isolates from the United States, consistent with the hypothesis that all *X. fastidiosa* subsp. *fastidiosa* isolates currently found in the United States are derived from a single strain introduced from Central America (32). Based on these data, we conclude that the *X. fastidiosa* subsp. *fastidiosa* donor (carrying *cysG12* and *holC19*) was introduced into the United States from Central America and recombined with a native *X. fastidiosa* subsp. *multiplex* genotype similar to ST45; however, this donor lineage of *X. fastidiosa* subsp. *fastidiosa* was ultimately unsuccessful and died out. We can never conclusively prove the absence of this genotype from North America. However, *X. fastidiosa* has been extensively sampled from many plant species throughout the United States, and no isolates of *X. fastidiosa* subsp. *fastidiosa* have been found that carry alleles similar to the inferred donor alleles *cysG12* and *holC19* (31); indeed all *X. fastidiosa* subsp. *fastidiosa* isolates so far found in the United States are consistent with the introduction into the United States of just a single genotype (32).

The transient presence of the donor genotype is consistent with a single large-scale introgression event founding the recombinant group. This raises the possibility that conjugation might have been involved; however, if this was the case, the genomic DNA was broken into pieces prior to homologous recombination, since the data show short regions of recombination. The data from the MLST loci plus *pilU* show 7 significant recombination events (Table 1), and 3 of them included at least one recombination breakpoint. Since these loci range in length from roughly 400 to 700 bp, this result would be consistent with an average recombination length of no more than a few kb, similar to the 2.6-kb average length observed by Nunney et al. (32) in a comparison of two *X. fastidiosa* subsp. *fastidiosa* genomes (see also reference 53 for an updated four-genome comparison with a mean size of 1.9 kb). Similarly, the regions identified from the data of Parker et al. (44) showed the same pattern, with a high proportion of recombination breakpoints identified within the sequenced loci (Table 5). In this context, it is important to note that Rogers and Stenger (37) have found a conjugative plasmid in *X. fastidiosa*. Furthermore, a high rate of transformation has been demonstrated in the lab (36), and it has been shown that this process can result in efficient recombination with only a few hundred bases of homologous sequence (53). Both conjugation and transformation may have been involved in the evolution of the recombinant group, since the data raise the possibility of both large-scale intersubspecific and smaller-scale intrasubspecific recombination (Fig. 3).

The results support the general conclusion that successful recombination is a rare but important event, a possibility emphasized by Wiedenbeck and Cohan (5) in their review of bacterial adaptation to new environments. However, given the high rates of recombination observed experimentally in *X. fastidiosa* (36), this rarity is somewhat surprising, perhaps suggesting that in *X. fastidiosa* the majority of intersubspecific recombination events fail due to their negative fitness consequences. Fitness loss due to recombination is consistent with the high level of plant host specificity observed among the genotypes of *X. fastidiosa* (35). On the other hand, it is clear that recombination can create combinations that are beneficial to the species, enabling it to invade new plant hosts. Specifically, the successful invasion of blueberry and blackberry appears to have resulted from large-scale recombination between two subspecies, a pattern that appears to be repeated in the invasion of mulberry (49; Nunney et al., unpublished data). Furthermore, Nunney et al. (11) suggested that introgression into *X. fastidiosa* subsp. *pauca* in South America from a donor (presumed to be introduced *X. fastidiosa* subsp. *multiplex*) may have enabled *X. fastidiosa* subsp. *pauca* to infect citrus, causing the economically devastating disease of citrus variegated chlorosis

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

(CVC). This would help explain why CVC did not appear in Brazil until the 1980s (54), despite the presence of the native pathogen and vectors ever since citrus was introduced several hundred years ago. These observations raise an important concern: that mixing of genetically divergent forms of the same species can result in recombinant forms capable of invading new niches—in this case, new plant hosts. Thus, the presence of a pathogen in an area should not lead to the assumption that further introductions will cause no further harm; in fact, as a result of recombination, further introductions may result in a qualitative worsening of the problem.

ACKNOWLEDGMENTS

This study was supported by a USDA-CSREES-NRI grant 2007-55605-17834 to L.N., R.S., D.L.H., and L.D.M.

We gratefully acknowledge Mark Black, Donald Cooksey, Rufina Hernandez-Martinez, Erin Schuenzel, Norm Schaad, Frank Wong, John Hartung, Sandy Purcell, and Rodrigo Almeida for providing the strains used in this study and two anonymous reviewers for helpful comments. Bernadette Amsden, Laramy Enders, Elizabeth Mah, Jennie Sotelo, Josh Wang, and Xiaoli Yuan are thanked for sample collection, DNA extraction, and/or sequencing genes of many of the isolates used in this study. We thank the three reviewers for valuable comments that enabled us to significantly improve the manuscript.

REFERENCES

1. **Ochman H, Lawrence JG, Groisman EA.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299–304. http://dx.doi.org/10.1038/35012500.
2. **Koonin EV, Makarova KS, Aravind L.** 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. **55:**709–742. http://dx.doi.org/10.1146/annurev.micro.55.1.709.
3. **Zhaxybayeva O, Doolittle WF.** 2011. Lateral gene transfer. Curr. Biol. **21:**R242–R246. http://dx.doi.org/10.1016/j.cub.2011.01.045.
4. **Sheppard SK, McCarthy ND, Jolley KA, Maiden MCJ.** 2011. Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles. Microbiology **157:**1066–1074. http://dx.doi.org/10.1099/mic.0.045153-0.
5. **Wiedenbeck J, Cohan FM.** 2011. Origins of bacterial diversity through horizontal gene transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. **35:**957–976. http://dx.doi.org/10.1111/j.1574-6976.2011.00292.x.
6. **Baltrus DA.** 2013. Exploring the costs of horizontal gene transfer. Trends Ecol. Evol. **28:**489–495. http://dx.doi.org/10.1016/j.tree.2013.04.002.
7. **Datta N, Kontomichalou P.** 1965. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. Nature **208:**239. http://dx.doi.org/10.1038/208239a0.
8. **Didelot X, Maiden MCJ.** 2010. Impact of recombination on bacterial evolution. Trends Microbiol. **18:**315–322. http://dx.doi.org/10.1016/j.tim.2010.04.002.
9. **Wren BW.** 2000. Microbial genome analysis: insights into virulence, host adaptation and evolution. Nat. Rev. Genet. **1:**30–39. http://dx.doi.org/10.1038/35049551.
10. **Marri PR, Hao W, Golding GB.** 2007. The role of laterally transferred genes in adaptive evolution. BMC Evol. Biol. **7**(Suppl 1)**:**S8. http://dx.doi.org/10.1186/1471-2148-7-S1-S8.
11. **Nunney L, Yuan X, Bromley RE, Stouthamer R.** 2012. Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. Appl. Environ. Microbiol. **78:**4702–4714. http://dx.doi.org/10.1128/AEM.01126-12.
12. **Sheppard SK, Didelot X, Jolley KA, Darling AE, Pascoe B, Meric G, Kelly DJ, Cody A, Colles FM, Strachan NJC, Ogden ID, Forbes K, French NP, Carter P, Miller WG, McCarthy ND, Owen R, Litrup E, Egholm M, Affourtit JP, Bentley SD, Parkhill J, Maiden MCJ, Falush D.** 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol. Ecol. **22:**1051–1064. http://dx.doi.org/10.1111/mec.12162.
13. **Vos M, Didelot X.** 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. **3:**199–208. http://dx.doi.org/10.1038/ismej.2008.93.
14. **Smith JM, Dowson CG, Spratt BG.** 1991. Localized sex in bacteria. Nature **349:**29–31. http://dx.doi.org/10.1038/349029a0.
15. **Coffey TJ, Daniels M, McDougal LK, Dowson CG, Tenover FC, Spratt BG.** 1995. Genetic analysis of clinical isolates of *Streptococcus pneumoniae* with high-level resistance to expanded-spectrum cephalosporins. Antimicrob. Agents Chemother. **39:**1306–1313. http://dx.doi.org/10.1128/AAC.39.6.1306.
16. **Halter R, Pohlner J, Meyer TF.** 1989. Mosaic-like organization of IgA protease genes in *Neisseria gonorrhoeae* generated by horizontal genetic exchange in vivo. EMBO J. **8:**2737–2744.
17. **Feil EJ, Spratt BG.** 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. **55:**561–590. http://dx.doi.org/10.1146/annurev.micro.55.1.561.
18. **Baltrus DA, Guillemin K, Phillips PC.** 2008. Natural transformation increases the rate of adaptation in the human pathogen *Helicobacter pylori*. Evolution **62:**39–49. http://dx.doi.org/10.1111/j.1558-5646.2007.00271.x.
19. **Fall S, Mercier A, Bertolla F, Calteau A, Gueguen L, Perriere G, Vogel TM, Simonet P.** 2007. Horizontal gene transfer regulation in bacteria as a "spandrel" of DNA repair mechanisms. PLoS One **2:**e1065. http://dx.doi.org/10.1371/journal.pone.0001065.
20. **Didelot X, Falush B.** 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics **175:**1251–1266. http://dx.doi.org/10.1534/genetics.106.063305.
21. **McVean GAT, Awadalla P, Fearnhead P.** 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:**1231–1241. http://www.genetics.org/content/160/3/1231.long.
22. **Bruen TC, Philippe H, Bryant D.** 2006. A simple robust statistical test for detecting the presence of recombination. Genetics **172:**2665–2681. http://dx.doi.org/10.1534/genetics.105.048975.
23. **Pritchard JK, Stephens M, Donnelly P.** 2000. Inference of population structure using multilocus genotype data. Genetics **155:**945–959. http://www.genetics.org/content/155/2/945.long.
24. **Falush D, Stephens M, Pritchard JK.** 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164:**1567–1587. http://www.genetics.org/content/164/4/1567.long.
25. **Sheppard SK, McCarthy ND, Falush D, Maiden MCJ.** 2008. Convergence of *Campylobacter* species: implications for bacterial evolution. Science **320:**237–239. http://dx.doi.org/10.1126/science.1155532.
26. **Arnold ML, Sapir Y, Martin NH.** 2008. Genetic exchange and the origin of adaptations: prokaryotes to primates. Philos. Trans. R. Soc. Lond. B Biol. Sci. **363:**2813–2820. http://dx.doi.org/10.1098/rstb.2008.0021.
27. **Whitney KD, Randell RA, Rieseberg LH.** 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytologist. **187:**230–239. http://dx.doi.org/10.1111/j.1469-8137.2010.03234.x.
28. **Keymer DP, Boehm AB.** 2011. Recombination shapes the structure of an environmental *Vibrio cholerae* population. Appl. Environ. Microbiol. **77:**537–544. http://dx.doi.org/10.1128/AEM.02062-10.
29. **Didelot X, Achtman M, Parkhill J, Thomson NR, Falush D.** 2007. Bimodal pattern of relatedness between the *Salmonella paratyphi* and *typhi* genomes: convergence or divergence by homologous recombination? Genome Res. **17:**61–68. http://dx.doi.org/10.1101/gr.5512906.
30. **Scally M, Schuenzel EL, Stouthamer R, Nunney L.** 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. Appl. Environ. Microbiol. **71:**8491–8499. http://dx.doi.org/10.1128/AEM.71.12.8491-8499.2005.
31. **Yuan XL, Morano LB, Bromley RE, Spring-Pearson S, Stouthamer R, Nunney L.** 2010. MLST analysis of *Xylella fastidiosa* isolates causing Pierce's disease of grape and oleander leaf scorch across the United States. Phytopathology **100:**601–611. http://dx.doi.org/10.1094/PHYTO-100-6-0601.
32. **Nunney L, Yuan X, Bromley RE, Hartung J, Montero-Astua M, Moreira L, Ortiz B, Stouthamer R.** 2010. Population genomic analysis of a bacterial plant pathogen: novel insight into the origin of Pierce's disease of grapevine in the US. PLoS One **5:**e15488. http://dx.doi.org/10.1371/journal.pone.0015488.
33. **Schaad NW, Postnikova E, Lacy G, Fatmi M, Chang CJ.** 2004. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. *piercei*, subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov.

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest

Syst. Appl. Microbiol. **27:**290–300. http://dx.doi.org/10.1078/0723-2020-00263.
34. **Schuenzel EL, Scally M, Stouthamer R, Nunney L.** 2005. A multigene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. Appl. Environ. Microbiol. **71:**3832–3839. http://dx.doi.org/10.1128/AEM.71.7.3832-3839.2005.
35. **Nunney L, Vickermann DB, Bromley RE, Russell SA, Hartman JR, Morano LD, Stouthamer R.** 2013. Recent radiation and host plant specialization in *Xylella fastidiosa* native to the United States. Appl. Environ. Microbiol. **79:**2189–2200. http://dx.doi.org/10.1128/AEM.03208-12.
36. **Kung SH, Almeida RPP.** 2011. Natural competence and recombination in the plant pathogen *Xylella fastidiosa*. Appl. Environ. Microbiol. **77:**5278–5284. http://dx.doi.org/10.1128/AEM.00730-11.
37. **Rogers EE, Stenger DC.** 2012. A conjugative 38 kb plasmid is present in multiple subspecies of *Xylella fastidiosa*. PLoS One **7:**e52131. http://dx.doi.org/10.1371/journal.pone.0052131.
38. **Purcell AH, Hopkins DL.** 1996. Fastidious xylem-limited bacterial plant pathogens. Annu. Rev. Phytopathol. **34:**131–151. http://dx.doi.org/10.1146/annurev.phyto.34.1.131.
39. **Sherald JL.** 2007. Bacterial leaf scorch of landscape trees: what we know and what we do not know. Arboric. Urban For. **33:**376–385. http://joa.isa-arbor.com/request.asp?JournalID=1&ArticleID=3015&Type=2.
40. **Lopes SA, Marcussi S, Torres SCZ, Souza V, Fagan C, França SC, Fernandes NG, Lopes JRS.** 2003. Weeds as alternative hosts of the citrus, coffee, and plum strains of *Xylella fastidiosa* in Brazil. Plant Dis. **87:**544–549. http://dx.doi.org/10.1094/PDIS.2003.87.5.544.
41. **Almeida RPP, Nascimento FE, Chau J, Prado SS, Tsai CW, Lopes SA, Lopes JRS.** 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. Appl. Environ. Microbiol. **74:**3690–3701. http://dx.doi.org/10.1128/AEM.02388-07.
42. **Nunney L, Elfekih S, Stouthamer R.** 2012. The importance of multilocus sequence typing: cautionary tales from the bacterium *Xylella fastidiosa*. Phytopathology **102:**456–460. http://dx.doi.org/10.1094/PHYTO-10-11-0298.
43. **Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou JJ, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. **95:**3140–3145. http://dx.doi.org/10.1073/pnas.95.6.3140.
44. **Parker JK, Havird JC, De La Fuente L.** 2012. Differentiation of *Xylella fastidiosa* strains via multilocus analysis of environmentally mediated genes (MLSA-E). Appl. Environ. Microbiol. **78:**1385–1396. http://dx.doi.org/10.1128/AEM.06679-11.
45. **Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P.** 2010. RDP3: a flexible and fast computer program for analyzing recombination Bioinformatics **26:**2462–2463. http://dx.doi.org/10.1093/bioinformatics/btq467.
46. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
47. **Felsenstein J.** 1989. PHYLIP: phylogeny inference package (version 3.2). Cladistics **5:**164–166.
48. **Felsenstein J.** 2005. PHYLIP (phylogeny inference package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA.
49. **Nunney L.** 2011. Homologous recombination and the invasion of a new plant host by the pathogenic bacterium, *Xylella fastidiosa*. Phytopathology **101:**S130. http://www.apsnet.org/meetings/Documents/2011_Meeting_Abstracts/a11ma762.htm.
50. **Cooper TF.** 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. PLoS Biol. **225:**1899–1905. http://dx.doi.org/10.1371/journal.pbio.0050225.
51. **Robinson K, Loughlin MF, Potter R, Jenks PJ.** 2005. Host adaptation and immune modulation are mediated by homologous recombination in *Helicobacter pylori*. J. Infect. Dis. **191:**579–587. http://dx.doi.org/10.1086/427657.
52. **McCarthy ND, Colles FM, Dingle KE, Bagnall MC, Manning G, Maiden MCJ, Falush D.** 2007. Host-associated genetic import in *Campylobacter jejuni*. Emerg. Infect. Dis. **13:**267–272. http://dx.doi.org/10.3201/eid1302.060620.
53. **Kung SH, Retchless AC, Kwan JY, Almeida RPP.** 2013. Effects of DNA size on transformation and recombination efficiencies in *Xylella fastidiosa*. Appl. Environ. Microbiol. **79:**1712–1717. http://dx.doi.org/10.1128/AEM.03525-12.
54. **Chang CJ, Garnier RM, Zreik L, Rossetti V, Bove JM.** 1993. Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. Curr. Microbiol. **27:**137–142. http://dx.doi.org/10.1007/BF01576010.

Downloaded from http://aem.asm.org/ on October 24, 2020 by guest