# GENOME-WIDE IDENTIFICATION OF RAPIDLY EVOLVING GENES IN *XYLELLA FASTIDIOSA*: KEY ELEMENTS IN THE SYSTEMATIC IDENTIFICATION OF HOST STRAINS, AND IN THE SEARCH FOR PLANT-HOST PATHOGENICITY CANDIDATE GENES

**Project Leader:**
Leonard Nunney
Dept. of Biology
University of California
Riverside, CA 92521

**Cooperators:**

| | |
|---|---|
| Richard Stouthamer | Robert Luck |
| Dept. of Entomology | Dept. of Entomology |
| University of California | University of California |
| Riverside, CA 92521 | Riverside, CA 92521 |


**Reporting Period:** The results reported here are from work conducted from October 15, 2003 to September 31, 2004.

## ABSTRACT
We have developed a robust phylogeny of the North American isolates of *Xylella fastidiosa* based on 10 genes (9288 base pairs). This supports the recent division of *X. fastidiosa* into subspecies (piercei and multiplex in N. America), however, we found 1 additional distinct taxon. The oleander isolates form a distinct group (provisionally named sandyi) that separated from the Pierce's disease group (piercei) long before European settlement of N. America, probably substantially more than 20,000 years ago. We used the phylogenetic tree to confirm the effectiveness of multilocus sequence typing (MLST) in identifying the subspecies and (within subspecies multiplex) plant-host isolates. MLST involves sequencing at least 7 genes from pure cultures. We have also developed a simpler method that distinguishes the major groups using restriction enzymes. This method has the advantage of working on mixed cultures and requiring only 3 PCR reactions. Our sequencing has confirmed that *X. fastidiosa* is largely clonal, and that within the piercei and sandyi groups there is very little genetic variability or geographical substructure. This pattern is particularly notable given the age of these groups and suggests the action of strong natural selection favoring specific clones. Finally, we found 4 (1.6%) examples of interstrain recombination, and the clustering of 3 in each of 2 isolates suggests that recombination may drive the rapid evolution of new pathotypes.

## INTRODUCTION
We are utilizing the extraordinary power of genomic research to investigate aspects of *Xylella fastidiosa*'s evolutionary history. This history provides information essential for controlling and solving the problem of Pierce's disease. At a minimum, it provides an understanding of the origin of the Pierce's disease (PD) strain of *X. fastidiosa*, and the relationship of the PD strain to other isolates of *X. fastidiosa*. Knowing the level of variability within the PD strain provides important information regarding the nature of these bacteria. Low variability would suggest that the PD strain is subject to significant constraints that may make controlling the pathogen simpler. On the other hand, evidence of high variability and high levels of recombination would suggest that the rapid evolution of resistance to control measures could be a severe problem.

A high priority is to place the PD strain within a robust phylogeny, extending earlier work defining the interrelationships of the plant-host strains of *Xylella fastidiosa* (e.g. see Hendson *et al.* 2001). Schaad *et al.* (2004) have recently named the PD strain as subspecies piercei, based on DNA hybridization. They identified two N. American subspecies (piercei and multiplex). It is important to determine if that taxonomy is sufficient to describe all N. American isolates.

Given a robust phylogeny, genomic data can be used to develop effective methods for identifying host strains, using either simple assays (e.g. restriction enzymes) or more sophisticated methods. MLST (multiple locus sequence typing) (Maiden *et al* 1998) is a valuable technique for identifying bacterial strains. Unambiguous identification of strains is of considerable importance for understanding the epidemiology of Pierce's disease and the other plant diseases caused by this bacterium. Previously, this has been approached using a variety of DNA based methods (Banks et al. 1999; Hendson *et al*. 2001; Rodrigues et al. 2003; Meinhardt et al. 2003;); however, an effective methodology for identifying the plant-host strains, including when they are mixed together, has yet to be developed.

The bacterium *X. fastidiosa* is generally assumed to be clonal. However, virally-mediated horizontal transfer of genes must occur given the presence of unique regions of DNA in the different host strains (Van Sluys *et al*. 2003). The possibility of direct inter-strain genetic transfer is more difficult to detect, but needs to be investigated. If such transfer does occur, it could lead to the very rapid evolution of novel pathogenic forms. Studying the details of sequence evolution across many genes provides information on the past occurrence of such events and hence their future likelihood.

**OBJECTIVES**

During the last year we have focussed on the following objectives:

1. Develop a systematic multigenic method for identifying host strains of *X. fastidiosa*. Our objective is to develop a method that unambiguously identifies the known host strains, and that allows an efficient recognition of the invasion of new strains.
2. Measurement of clonal variation within host strains. Our objective is to assess within-strain genetic variability and geographical substructure at our target gene loci. From this we can infer the probable importance of plant-host adaptation.
3. Estimate the frequency of recombination. Our objective is to look for evidence of both within- and between-strain genetic transfer. Genetic transfer can dramatically increase the rate of evolution, and potentially can increase the rate at which new –more virulent- host strains arise.

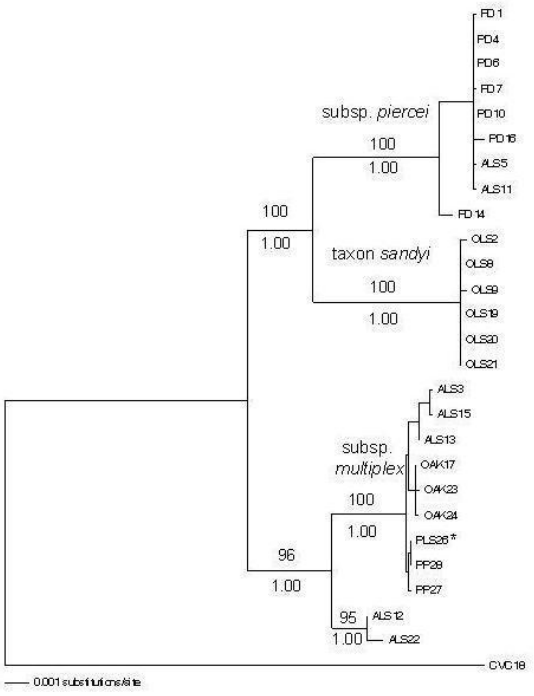


**Figure 1.** Phylogenetic relationships among 26 N. American isolates of X. fastidiosa from 6 species of host plant, using CVC (from S. America) as the outgroup. The maximum likelihood tree is based on 10 genes except PLS26, which was positioned in the tree based on the sequence of 7 genes. Isolates were from grapevine (PD), almond (ALS), oleander (OLS), oak (OAK), peach (PP), and plum (PLS).

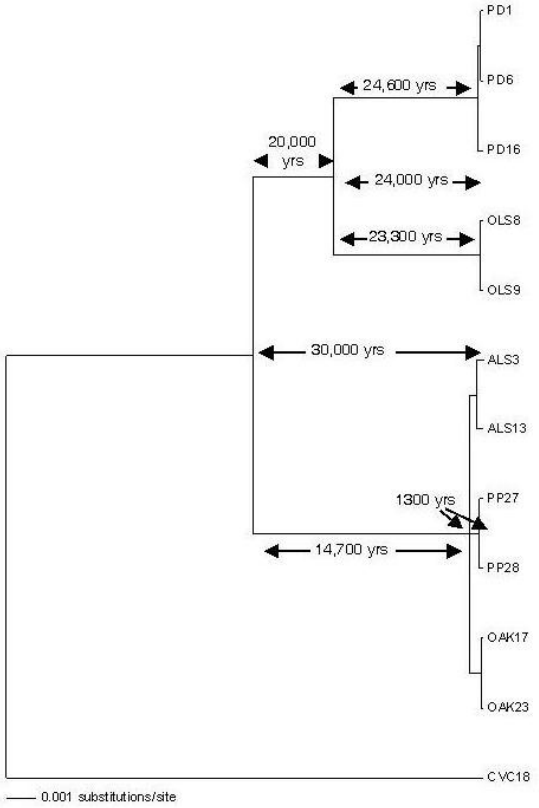


**Figure 2.** Phylogenetic estimates of the divergence times of the groups of X. fastidiosa based on the rate of synonymous substitution within each branch of the maximum likelihood tree.

**RESULTS**

***Objective 1: Develop a Systematic Multigenic Method for Identifying Host Strains of X. fastidiosa.***

To create a statistically robust phylogeny of the host-plant strains of *X. fastidiosa*, we sequenced 10 genes (9288 bp) from each of 25 isolates, and 7 genes from 1 additional isolate. The results are shown in Figure 1 using the S. American CVC strain as the outgroup. The tree shows three well-defined clades that are supported 100% by bootstrap procedures. Two of these clades correspond to the recently named subspecies piercei and multiplex (Schaad et al 2004). Subsp. piercei includes all Pierce's disease isolates. Subsp. multiplex includes a set of isolates from almond plus isolates from a range of host plants from the eastern US (oak, peach, and plum). The third clade contains only isolates from oleander. It is most closely related to subsp. piercei, but shows a high degree of differentiation from that subspecies (2.6% at synonymous sites). In addition, bacteria from these two groups cannot infect each other's major host plant (oleander vs. grapevine) and based on the lack of intermediates, we conclude that the oleander clade constitutes a third N. American subspecies that we have tentatively named sandyi (Scheunzel *et al* 2004).

To begin to understand the evolution of the pathogenicity of the plant-host strains of *X. fastidiosa*, it is important that we have a good estimate of the age of these clades. In particular, since this species of bacteria appears to be restricted to the

Americas and since most of the plant hosts exhibiting disease symptoms are introduced species, we need to know if these three N. American clades pre-date European colonization. We estimated divergence dates based on the rate of synonymous substitution. Assuming that such substitutions are generally neutral and driven by genetic drift, then we have that the time of origin *T* (in years) of a given clade is $T = K/(nu)$, where *K* is the number of synonymous substitutions per site in a given branch, *u* is the mutation rate per generation, and *n* is the number of generations per year. We used $u=5.4x10^{-10}$ (the E. coli rate, see Drake *et al* 1998) and n=1000, corresponding to a long-term division rate of once every 9hrs. The generation time of *X. fastidiosa* has been estimated at between 9 and 60 hours (Wells *et al* 1987), so our assumption is conservative (reducing T). The resulting estimates are shown in Figure 2. These estimates suggest that the three clades, piercei, multiplex, and sandyi, have been distinct for at least 15,000 years, and possibly much longer.

It is notable that the estimated age of the multiplex clade is 3x less than the estimated age of the parallel piercei/sandyi group. Since they are exactly the same age, the most likely explanation is that the generation time (in nature) of members of the multiplex clade is about 3x longer (i.e. *n* is smaller in eqn 1). Note that this effect is apparent both before and after the split of piercei and sandyi, (20,000 yrs plus 24,000 yrs compared to the multiplex total of 14,700 yrs), and that the rate within the piercei and sandyi clades is extremely similar (24,600 vs. 23,300).

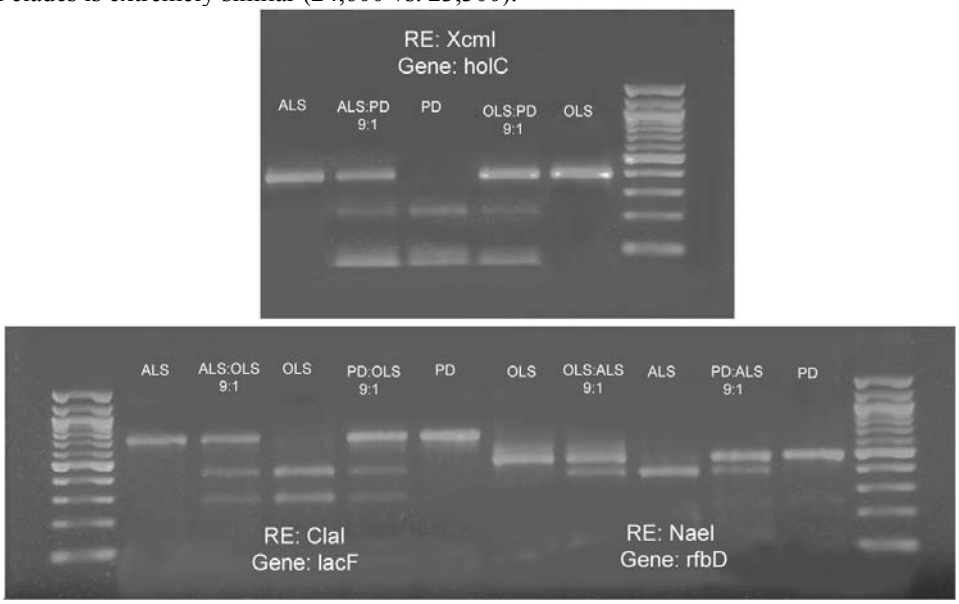


**Figure 3.** Restriction digests following amplification of single genes from pure-strain DNA, or from a 9:1 mix of the DNA of two strains.

We have shown that the MLST approach of Maiden *et al* (1998) can be used to document both the differences among the three major groups, and the differences among the plant-host isolates of subsp. multiplex (data not shown). The strength of this approach is that MLST data are unambiguous, can be held on a central database, and can be queried through the Web.

Using three of the target genes, we developed a PCR/restriction enzyme essay that separates the major groups of *X. fastidiosa*. We have shown that this method can be used to identify strains from mixtures of DNA (figure 3).

***Objective 2: Measurement of Clonal Variation Within Host Strains***

It is clear from Figure 1 that there is very little variability within the three clades. Furthermore, we found no evidence of geographical substructure. Using Kst (which measures genetic differentiation between populations relative to within populations) we found no differentiation between 2 northern California isolates of piercei (PD4,6; see fig. 1) vs. 6 southern California isolates (PD1,7,10,14, ALS5,11) (Kst = 0.00 ns), or between three northern California almond (non-piercei) isolates (ALS3,15,22) and 2 southern California isolates (ALS 12,13) (Kst = -0.26 ns). Over a longer distance, the piercei isolate from Florida (PD16) and the sandyi isolate from Texas (OLS8) showed no marked difference from the remaining isolates in their respective clades (all from California). The lack of intra-clade variability results in a phylogeny with long basal branches leading to very short terminal branches. This pattern suggests that the strains experience strong selective pressures from their host plants, eliminating all but the best-adapted clones.

***Objective 3: Estimate the Frequency of Recombination***

Given the low level of clade variability, the isolates exhibiting inter-strain recombination at one or more of the 10 sequenced loci can be seen quite clearly from fig. 1. They are PD14 (1 recombination), and ALS 12, 22 (recombination in 3 genes). The sites of the recombination can be seen clearly by aligning the sequences. Thus from 257 gene sequences we found 4 independent recombination events, i.e. 1.6%. It is notable that ALS 12 and ALS 22 were isolated in California from almond

trees more than 200 miles apart (Temecula and San Joaquin), but they exhibit the same 3 recombinant events. These isolates may represent the evolution of a new pathotype through recombination.

The source of the recombinant DNA could be determined by its sequence identity with the gene from a different strain. This identity suggests that these genetic transfers occurred relatively recently. Thus PD14 incorporated DNA from a multiplex ALS-type bacterium in its cysG gene.

**CONCLUSIONS.**

1. There are 3 clades of *X. fastidiosa* within N. America, corresponding to subsp. piercei and multiplex, and the newly named taxon sandyi that causes oleander leaf scorch.
2. The 3 clades originated at least 15,000 years ago. This guarantees that the clades could not have developed in response to host plants introduced by Europeans, e.g. oleander.
3. Isolates from the same clade showed very few genetic differences, and we found no evidence of geographical genetic structure within the piercei or sandyi clades. This limited variability within very old taxa suggests strong selection, possibly driven by host-plant adaptation.
4. Multi-locus sequence typing (MLST) is effective at identifying the three clades, and the plant-host strains within the multiplex group.
5. We can detect mixtures of the 3 main types of *X. fastidiosa* using 3 genes subject to restriction digests.
6. We observed 4 examples of recombination in a sample of 257 genes. Three of these recombinations were found replicated in two isolates. This highly non-random distribution is consistent with the possibility that new recombinant forms can rapidly generate novel pathotypes.

**REFERENCES**


Banks, D., R. Albibi, J. Chen, O. Lamikanra, R. Jarret, B.J. Smith. 1999. Specific detection of *Xylella fastidiosa* Pierce's disease strains. Curr. Microbiol. 39:85-88.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148: 1667-1686.

Hendson, M., A.H. Purcell, D. Chen, C. Smart, M. Guilhabert, B. Kirkpatrick. 2001. Genetic diversity of Pierce's disease strains and other pathotypes of *Xylella fastidiosa*. Appl. Environ. Microbiol. 67: 895-903.

Maiden, M.C.J., J.A. Bygraves, E. Feil, G. Morelli, J.E. Russell. 1998 Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Nat. Acad. Sci. 95: 3140-3145.

Meinhardt L.W., M.P.M.A. Ribeiro, H.D. Coletta, C.K. Dumenyo, S.M. Tsai, C.D. Bellato. 2003. Genotypic analysis of *Xylella fastidiosa* isolates from different hosts using sequences homologous to the *Xanthomonas* rpf genes. Mol. Plant Path. 4: 327-335

Rodrigues J.L.M., M.E. Silva-Stenico, J.E. Gomes, J.R.S. Lopes, S.M. Tsai. 2003. Detection and diversity assessment of *Xylella fastidiosa* in field-collected plant and insect samples by using 16S rRNA and gyrB sequences. Appl. Environ. Microbiol. 69: 4249-4255.

Schaad, N.W., E. Pastnikova, G. Lacey, M. Fatmi, C.J. Chang. 2004. *Xylella fastidiosa* supspecies: *X. fastidiosa* subsp piercei, subsp. nov., *X. fastidiosa* subsp. multiplex subsp. nov., and *X. fastidiosa* subsp. pauca subsp. nov. Syst. Appl. Microbiol. 27: 290-300.

Scheunzel, E.L., M. Scally, R. Stouthamer, L. Nunney. 2004. A multi-gene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. J. Bacteriol. (Submitted)

Van Sluys M.A., M.C. de Oliveira, C.B. Monteiro-Vitorello, C.Y. Miyaki, L.R. Furlan. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J. Bacteriol. 185: 1018-1026.

Wells, J. M., B. C. Raju, H. Y. Hung, W. G. Weisburg, L. Mandelco-Paul, and D. J. Brenner. 1987. *Xylella fastidiosa* gen. nov., sp. nov: Gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. Int. J. Syst. Bact. 37:136-143.


**FUNDING AGENCIES**


Funding for this project was provided by the University of California Pierce's Disease Grant Program.