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
Department of Biology
University of California
Riverside, CA 92521

Cooperators:
Richard Stouthamer
Robert Luck
Department of Entomology
University of California
Riverside, CA 92521

Reporting Period: The results reported here are from work conducted from November 1, 2002 to October 15, 2003.

INTRODUCTION

Genomic research provides an extraordinarily powerful new tool for solving applied problems. It can be extremely effective at resolving evolutionary relationships and can be used to extend earlier work defining the interrelationships of the plant-host strains of *Xylella fastidiosa* (see Hendson *et al.* 2001). Similarly, it can provide effective methods for identifying these host strains. Unambiguous identification is of considerable importance for understanding the epidemiology of Pierce's disease and the other plant diseases caused by this bacterium. 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 availability of sequenced genomes allows us to analyze the evolutionary history of not just one or two genes, but of all of the genes that make up this bacterium. In particular, evolutionary genomic techniques developed to detect the action of natural selection (Yang 1998) provide a new approach to identifying genes important in plant-host specificity.

The bacterium *X. fastidiosa* is generally assumed to be clonal. However, horizontal transfer of genes must occur given the presence of unique regions of DNA in the different host strains (Van Sluys *et al*. 2003). Such transfers are assumed to be virally mediated. The possibility of direct inter-strain genetic transfer is more difficult to detect, but needs to be investigated for two important reasons. First, bacteria from different strains are expected to mix at high density within the insect vector, and second, 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

1. The identification of the rapidly evolving genes in the *Xylella fastidiosa* genome. This is the first step towards achieving our four primary objectives. These are:
2. 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.
3. Identify plant-host specificity candidate genes. We will use our database of rapidly evolving proteins to test for evidence of strong natural selection and for links between the rapid genetic divergence of host strains and specific biochemical functions.
4. Measurement of clonal variation within host strains. Our objective is to assess within-strain genetic variability at rapidly evolving gene loci and to use these results to assess the evidence that all members of a given host strain share common ancestry.
5. 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.

RESULTS

Objectives 1 and 3

We completed our gene-specific database of the four plant-host *Xylella fastidiosa* genomes, PD (Pierce's disease), OLS (Oleander Leaf Scorch), ALS (Almond Leaf Scorch), and CVC (Citrus Variegated Chlorosis). This database identifies all genes that occur in all four genomes, and enables us to compare the evolutionary characteristics of these genes. The genome-wide analysis is ongoing; however, we identified 11 genes to use in the detailed investigation of plant-host strain relationships, based on their significant shifts in Ka/Ks ratio within the phylogeny, and their distribution throughout the

genome (Table 1). Specifically, we measured shifts in this ratio of non-synonymous (Ka) to synonymous (Ks) substitutions between the OLS and ALS branches of the phylogeny. A significant shift indicates a change in the nature of the selection acting on the gene, and may indicate that these genes are important in plant-host adaptation.

Table 1. *Xylella fastidiosa* genes used for strain identification.

Gene ID		Name	χ^2 Value for shift in Ka/Ks
XF0136	holC	DNA polymerase III holoenzyme chi subunit	4.58
XF0257	rfbD	dTDP-4-dehydrorhamnose 3,5 epimerase	6.60
XF0316	nuoL	NADH-ubiquinone oxidoreductase, NQO12 subunit	5.95
XF0318	nuoN	NADH-ubiquinone oxidoreductase, NQO14 subunit	8.82
XF0656	gltT	glutamate symport protein	4.58
XF0832	cysG	siroheme synthase	4.71
XF0910	petC	ubiquinol cytochrome C oxidoreductase, cytochrome C1	3.88
XF1291	eno	enolase	5.29
XF1632	pilU	twitching motility protein	8.88
XF1818	leuA	2-isopropylmalate synthase	6.35
XF2447	lacF	ABC transporter sugar permease	4.28

Objectives 2 and 4

We have so far sequenced 9 of the 11 target genes (a total of 11014 bases) in 13 strains of *Xylella fastidiosa*. Strains have (and will continue to be) selected to represent the geographic distribution of the infections, with a goal of regional replication to enable tests of geographical hypotheses (Table 2).

Table 2. *Xylella fastidiosa* strains so far examined.

	Strain	Host	Location	Strain Identification**
1*	Temecula	Grape	Temecula	PD
2*	Ann-1	Oleander	Palm Springs	OLS
3*	Dixon	Almond	Solarno	ALS
4	STL	Grape	Napa	PD
5	Tulare	Almond	Tulare	PD
6	Conn Creek	Grape	Napa	PD
7	Traver	Grape	Tulare	PD
8	Texas	Oleander	Texas	OLS
9	Riverside	Oleander	Riverside	OLS
10	I03	Grape	Temecula	PD
11	237	Almond	Temecula	PD
12	276	Almond	Temecula	ALS (?)
13	187	Almond	Temecula	ALS

* Strains 1,2 and 3 were the strains used for the PD, OLS and ALS genomes.
** See Table 3

The sequences for 9 of the genes have been compared for the 13 strains using Multi-Locus Sequence Typing (Maiden *et al*. 1998). Any difference in a sequence results in the assignment of a new type (letter) for that gene (Table 3). All the PD strains except Traver in gene XF1632 were identical to the genome sequence for PD strain #1 (and scored the maximum of 100 in the row labeled "Score PD type"). The strain #11 (237) was isolated from almond but was identical to the "type" Temecula PD strain #1. Similarly #5 Tulare, also isolated from almond, showed only a single base difference with the #1 PD strain. The oleander strains showed more variation. The Riverside strain (#9) differed from the Ann-1 genome sequence (#2) at two genes, XF0832 and XF1632. Almond strains had the most variation. Dixon (#3) and 187 (#13) were identical except for one gene, XF0257. Strain #12 (276) had differences from Dixon in nearly all the genes, but still had more in common with Dixon and 187 than with any other strain.

We can also look at the MLST pattern from the perspective of the genes (i.e. rows in Table 3). For the design of efficient and effective strain identification it is important to note that the genes in the first 4 rows show identical patterns (and hence the same information). Similarly, the genes in the next 2 rows have an identical pattern.

Table 3. Multi-Locus Sequence Typing (MLST) for *Xylella fastidiosa*

	Strain												
Gene	1	4	6	10	11	5	7	2	8	9	3	13	12
XF0136	a	a	a	a	a	a	a	b	b	b	c	c	d
XF0656	a	a	a	a	a	a	a	b	b	b	c	c	d
XF1818	a	a	a	a	a	a	a	b	b	b	c	c	d
XF2447	a	a	a	a	a	a	a	b	b	b	c	c	d
XF0318	a	a	a	a	a	a	a	b	b	b	c	c	c
XF0910	a	a	a	a	a	a	a	b	b	b	c	c	c
XF0257	a	a	a	a	a	a	a	b	b	b	c	d	d
XF0832	a	a	a	a	a	a	a	b	b	d	c	c	e
XF1632	a	a	a	a	a	d	e	b	b	d	c	c	e
% Score PD type	**100**	**100**	**100**	**100**	**100**	**89**	**89**	0	0	0	0	0	0
% Score OLS type	0	0	0	0	0	0	0	**100**	**100**	**78**	0	0	0
% Score ALS type	0	0	0	0	0	0	0	0	0	0	**100**	**89**	**22**

Using the sequence data from the 13 strains provides a large dataset to construct a phylogenetic tree showing the interrelationships of the strains. There are several techniques that are used for the construction of phylogenetic trees, principally maximum likelihood, parsimony, and distance methods. All of these methods gave the same overall relationships among PD, OLS, and ALS strains (using CVC to root the tree). The maximum likelihood tree shown (Figure 1) was derived using the software package PAUP. The tree shows 3 well-defined clades, PD, OLS and ALS. These three clades are statistically very robust, with 100% bootstrap support for PD and OLS, and 98% for ALS. The PD strain's closest relative is always the OLS strain.

The PD clade includes strains isolated from almond, but the ALS clade does not include strains isolated from grapes. Furthermore, both the OLS and PD strains show surprisingly little variability in our sample so far, and these sequences of over 10K bases show no geographical structure. For example, the cluster of PD strains (1,4,6,10 and 11) derive from Napa and Temecula. Similarly the geographically separated OLS strains 2 and 8, from Palm Springs and Texas (site unknown), are identical across the sequenced genes.

The finding of little variation within our present strain sample simplifies strain identification, if the pattern holds when more strains are added. Using some of the 11 sequenced genes we have developed and tested 9 gene/restriction enzyme reactions. Each reaction identifies one strain (3 PD, 3 ALS, 3 OLS), thus appropriate combinations of three reactions allow unambiguous positive identification of the three strains. One promising combination involves three PCR reactions (genes XF0136, XF0257, and XF0832) using the same restriction enzyme. We are currently testing mixtures of DNA from different strains as a first step to establishing how effective the method is for detecting a mixture, when one strain is rare.

Objective 5

We did not anticipate finding any indication of recombination at this early stage; however examination of the ALS strain #12 revealed strong (though preliminary) evidence of genetic exchange between strains. When the strains are compared on a gene-by-gene basis, the three ALS strains typically cluster very closely. However, in two genes the #12 sequence clusters closely with OLS and in one gene it clusters closely with PD.

CONCLUSIONS

1. In broad agreement with previous work, we have found that PD, OLS, and ALS strains of *X. fastidiosa* form robust clades, with PD being most closely related to the OLS strain.
2. It appears that the PD strain can infect almond, but there is no evidence that the ALS strains can infect grape.
3. We find no evidence of geographical differences within the PD and OLS strains. In particular, the PD strains from Temecula and Napa are extremely similar.
4. The simple technique of multi-locus sequence typing (MLST) is effective at identifying the three strains.
5. We have developed a simple method to uniquely identify the PD, OLS, and ALS strains by looking for characteristic restriction fragments following PCR of targeted genes. We are currently testing the sensitivity of the method in detecting a rare strain in mixtures of DNA.
6. One strain (#12) provides preliminary support for recombination among strains. The importance of this result is that it raises the possibility that recombination can rapidly generate novel pathotypes.

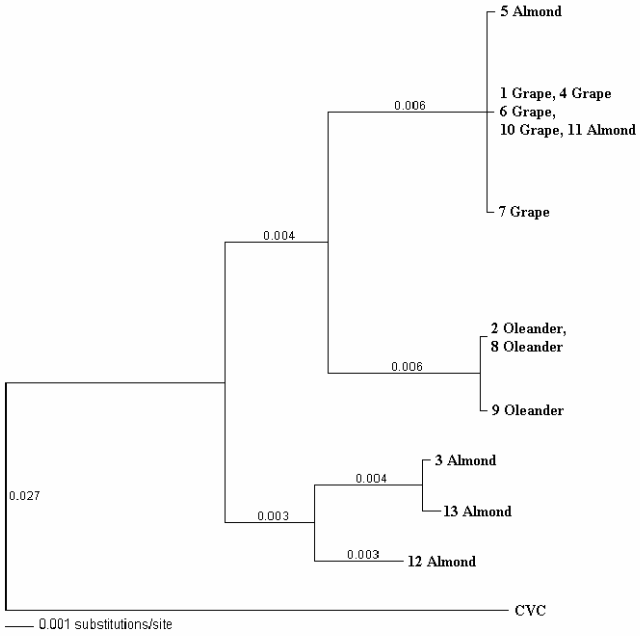

Figure 1: Phylogenetic relationships among 13 strains of *X. fastidiosa*

REFERENCES

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

Hendson, M., A.H. Purcell, D. Chen, C. Smart, M. Guilhabert, and 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, et al. 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, and 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, and 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.

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

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15: 568-573

FUNDING AGENCIES

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