DEVELOPING A MICROARRAY-PCR-BASED IDENTIFICATION AND DETECTION SYSTEM FOR *XYLELLA FASTIDIOSA* STRAINS IMPORTANT TO CALIFORNIA

Project Leaders:
Jianchi Chen
USDA, ARS, PWA
San Joaquin Valley Agricultural Sciences Center
Exotic and Invasive Diseases and Pests Research
Parlier, CA 93648

Edwin L. Civerolo
USDA, ARS, PWA
San Joaquin Valley Agricultural Sciences Center
Exotic and Invasive Diseases and Pests Research
Parlier, CA 93648

Reporting Period: The results reported here are from work conducted from January 2004 to September 2004**.**

ABSTRACT
From the analysis of the 16S rDNA sequence of *Xylella fastidiosa*, we have identified four single nucleotide polymorphisms (SNPs). The combination of these four SNPs placed all of the known *X. fastidiosa* strains into four groups. With a few exceptions, the four SNP groups are very similar to those based on other genetic analyses such as RAPD analysis, whole 16S rDNA sequence analysis, and the combination of phenotypic characterization, particularly pathogenicity tests. Of particular interest is the PD group. All eight PD strain 16S rDNA sequences from different labs clustered into the same group characterized by two SNPs. Utilizing the SNP information, primer sets, Teme150fc-Teme454rg, specific to PD strain group, and Dixon454fa-Dixon1261rg, specific to non-PD almond leaf scorch group, were designed. More than 200 *X. fastidiosa* strains isolated from California have been tested for the specificity of these SNPs and the results were quite consistent. A microarray system, initially based on the characteristic SNPs from the 16S rRNA locus, is under construction. Coupled with PCR using universal 16S rDNA primers, the microarray-PCR based system has a high potential for quick and accurate *X. fastdiosa* strain identification.

INTRODUCTION
The need to accurately identify and differentiate *X. fastidiosa* strains is becoming more apparent considering the coexistence of different pathotypes in the same crop (Chen et al., 2004a, b). This prompted us to research on improvement of pathogen detection. Polymerase chain reaction (PCR) has been a common technique for *X. fastidiosa* identification. There are, however, technical problems limiting the application of PCR. False positive amplifications can occur among related organisms in the environment sharing similar genetic sequences. Specific primers may fail to amplify DNA from a particular isolate if there is a spontaneous mutation(s) in the primer-binding site, leading to a false negative result. The sensitivity and specificity of PCR amplification tend to be inversely related.

The rationale of this project is to maximize the sensitivity of PCR technology. To increase pathogen detection specificity, microarray methodology based on the principle of DNA hybridization is applied to further confirm the accuracy of the amplified DNA fragments (Chen and Civerolo, 2003). Conceptually, the development of high-density oligonucleotide arrays allows massively parallel hybridizations to occur on the same surface, permitting high levels of probe redundancy and multiple independent detections of a diagnostic DNA sequence. Because of the taxonomic value and available large genomic sequence database, single nucleotide polymorphisms in the 16S rRNA gene are particularly useful. Other genes and intergenic regions could also be the targets due to the availability of complete genome sequences from four different *X. fastidiosa* strains.

OBJECTIVES
The overall goal of this project is to develop and evaluate a microarray-PCR-based system for accurate and quick identification of *X. fastidiosa* strains. A particular emphasis is on strains currently important in California. Two specific objectives are:

1. Using the complete and annotated genome sequence of *X. fastidiosa* Temecula strain as a guide, select appropriate DNA sequences and evaluate their potential for pathotype / genotype identification. Design and construct a DNA microarray; and
2. Evaluate the effectiveness of the constructed microarray through hybridization experiment. Using the microarray as a reference, analyze genomic variation of different pathotypes with multiple strains collected from broad geographical areas and hosts.

RESULTS AND CONCLUSIONS
Selected sequences in the genome of *X. fastidiosa* Temecula were used as preliminary queries to identify diagnostic sequences. Because of the sequence availability, most comparisons were made to the four complete genome sequences including PD-Temecula, citrus variegated chlorosis-9a5c, almond leaf scorch disease-Dixon and oleander leaf scorch disease-Ann-1. In general, the tested genome DNA sequences showed high level of similarity as expected. However, single nucleotide polymrphisms were found in most cases. Yet, the number of SNPs varied from gene to gene. Genes of evolutionary importance were particularly emphasized because they could provide a more stable and, therefore, a more consistent base for strain identification. Thus, special efforts were made on DNA sequences from *rrn* operons. In addition,

16S rDNA is by far the most sequenced locus in bacteria including *X. fastidiosa* that has at least 38 sequences currently available. These 38 16S rDNA sequences from eight different sources were retrieved from the GenBank database. The sequences were aligned using CLUSTAL-W program. Nucleotide variations were examined manually. Only the variations supported by multiple sequences were considered as true SNPs. The nucleotide order in the 16S ribosomal RNA gene, PD0048, in the *X. fastidiosa* strain Temecula genome sequence was used as reference to standardize the nucleotide number (Table 1).

Currently, the microarray system is still being established. The evaluation of SNPs for strain identification was done using PCR methodology. The Primer 3 program was used to facilitate primer designs. All primers were designed with Tm = 60±3 C. The basic strategy of primer design was to arrange the SNPs at the 3' end of the oligo-primers. Two multiplex PCR formats were implemented. For the three primer format, primers Teme150fc - Teme478rg-XF16s1031r generated two dominant amplicons, a 348 bp band for the PD group, and a 700 bp band for non-PD group generated by A non-specific prime paired with Teme150fc. In the four primer format, two primer sets were used. The PD group specific primer set, Teme150fc-Teme454rg was the same as in the three primer format. The other primer set, Dixon454fa-Dixon1261rg generated an 847 bp amplicon for the non-PD almond leaf scorch disease (ALSD) group (Figure 1). For comparison purpose, primer set RST31-RST33 was also included. RST31-RST33 is the most commonly used primer set for PCR identification *X. fastidiosa* at the species level. Primer specificity was also compared to non-reduntant GenBank database through the BLAST network service.

Efforts have also been made to obtain a comprehensive collection of *X. fastidiosa* strains in California with emphasis on grape and almond strains. Over 300 isolation attempts have been made from samples of grapes, almonds and other plants. Samples were collected form San Diego, Kern, Tulare, Kings, Fresno, Stanislaus, Butte, Alemeda and Solano counties. Strains were initially confirmed by biological characters such as slow growing and opalescence colony type and then by PCR with primer RST31/33. Over 200 strains were used to evaluate the specificity of the identified SNPs. Research results obtained by far consistently indicate that SNPs in the 16S rDNA sequence have high potential for *X. fastidiosa* strain differentiations. Current design strategy for microarry experiments is to place these SNPs in the center of the oligomers. Also as shown in Table 1, a total of four SNPs can be considered for oligomer designs to cover all the known strains of *X. fastidiosa*. The advantage of such a microarray identification system becomes even more obvious when 16S rDNA primers of different specificity levels, such as universal primers, are used to generate a large amount of target DNAs from a low titer of bacterial cells.

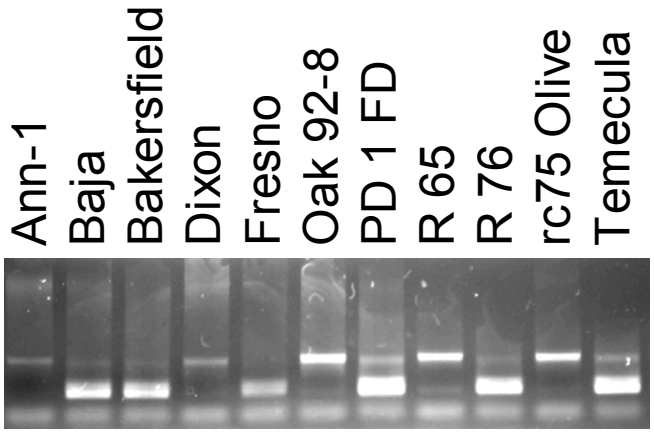

Figure 1. Representative results of multiplex PCR using the four primer format based on single nucleotide polymorphisms in the 16S rDNA sequence. The STRONG presence of the upper band (847 bp) indicates the almond leaf scorch strain group. The STRONG presence of the lower band (348 bp) indicates a grape Pierce's disease strain group.

REFERENCES

Chen, J., and Civerolo, E. 2003. Developing a microarray-PCR-based identification and detection system for *Xylella fastidiosa* strains important to California. Proceedings of Pierce's disease research symposium. California Department of Food and Agriculture. Page 49-50. December 8-11, 2003. Coronado, California.

Chen, J., Groves, R., Civerolo, E. L., Viveros, M., Freeman, M., and Y. Zheng. 2004a. Simultaneous association of more than one *Xylella fastidiosa* genotype with almond leaf scorch disease in California. Phytopath. 94:S17.

Chen, J., Groves, R., Civerolo, E. L., Viveros, M., Freeman, M., and Y. Zheng. 2004b. Simultaneous association of more than one *Xylella fastidiosa* genotype with almond leaf scorch disease in California. *(submitted)*

Table 1. List of four single nucleotide polymorphisms from 38 rDNA sequences of *Xylella fastidosa* and the related information.

Strain Name	Host	Geographic Origin	150	454	1261	1338
R116v11	Grape	Georgia	C	G	A	C
PCE-FG	Grape	Florida	C	G	A	C
PD28-5	Grape	Florida	C	G	A	C
PCE-FF	Grape	Florida	C	G	A	C
Temecula	Grape	California	C	G	A	C
GR.8935	Grape	Florida	C	G	A	C
Temecula	Grape	California	C	G	A	C
Temecula	Grape	California	C	G	A	C
Mul-2	Mulberry	Nebraska	C	G	A	C
Dixon1	Almond	California	T	A	G	C
Dixon2	Almond	California	T	A	G	C
Ann-1 1	Oleander	California	T	A	G	C
Ann-1 2	Oleander	California	T	A	G	C
PWT-22	Periwinkle	Florida	T	A	G	C
PWT-100	Periwinkle	Florida	T	A	G	C
Mul1	Mulberry	Massachusetts	T	A	G	C
Elm	Elm	Washington DC	T	A	G	C
OAK	OAK	Washington DC	T	A	G	C
PLS2-9	Plum	Georgia	T	A	G	C
PLM G83	Plum	Georgia	T	A	G	C
PP4-5	Peach	Georgia	T	A	G	C
RGW-R	Ragweed	Florida	T	A	G	C
ELM-1	Elm	Washington DC	T	A	G	C
ALS-BC	Almond	California	T	A	A	T
MUL-3	Mulberry	Massachusetts	T	A	A	T
P3	Coffee	Brazil	T	A	A	T
B14	Citrus	Brazil	T	A	A	T
SL1	Citrus	Brazil	T	A	A	T
CRS2	Coffee	Brazil	T	A	A	T
CM1	Coffee	Brazil	T	A	A	T
CI.52	Citrus	Brazil	T	A	A	T
CO.01	Coffee	Brazil	T	A	A	T
CVC93-2	Citrus	Brazil	T	A	A	T
9a5c	Citrus	Brazil	T	A	A	T
9a5c	Citrus	Brazil	T	A	A	T
PE.PLS	Pear	Taiwan	T	A	A	C
PL.788	Plum	Georgia	T	A	A	C
OSL92-3	Oak	Florida	T	A	A	C

FUNDING AGENCIES

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