EVALUATION OF SINGLE NUCLEOTIDE POLYMORPHISMS AND SAMPLE PREPARATION PROCEDURES IN THE DETECTION OF *XYLELLA FASTIDIOSA* STRAINS IMPORTANT TO CALIFORNIA

Project Leaders:
Jianchi Chen and Edwin L. Civerolo
USDA, ARS, PWA
Crop Diseases, Pests, & Genetics
Parlier, CA 93648

Reporting Period: The results reported here are from work conducted October 1, 2004 to September 30, 2005.

ABSTRACT
The main objective of this project is to develop a PCR-microarray-based system for accurate and quick identification of *Xylella fastidiosa* (*Xf*) strains important to California crops. The major part of the current work focused on identification of important DNA sequences based on single nucleotide polymorphisms (SNPs) for the detection of *Xf* Pierce's disease (PD) and almond leaf scorch disease strains. We expanded the previous SNP analysis in 16S rDNA to other house-keep genes, such as those coding for TCA cycle enzymes. *Xf*-specific primers were designed and real-time PCR was employed. Melting point analysis was used to confirm the presence of SNPs in the amplicons and to detect different genotypic strains. A second effort was to develop and improved procedure of sample preparation, the bottleneck of PCR detection of *Xf in planta*. We developed two simple sample preparation procedures for PCR amplification of *Xf* DNA from infected almond and grape petioles using freshly squeezed petiole sap and freeze-dried tissue. The detection efficiency of the two PCR methods was similar to that of the pathogen isolation procedure.

INTRODUCTION
Because of the availability of the complete genome sequences, considerable bioinformatics knowledge has been obtained by comparing four *Xf* strains, including strain Temecula for Pierce disease (PD) of grapevine, strain 9a5c for citrus variegated chlorosis, strain Dixon for almond leaf scorch disease (ALSD) and strain Ann-1 for oleander leaf scorch disease. One of the most direct benefits from the study of the genome sequences is the development of a highly accurate and efficient systems for pathogen identification and detection. We have proposed a PCR-microarray approach to develop such a system. In this study, we focused on two sub-components: identification of unique and informative sequences for each pathogen strain (genotype) and development of efficient procedures to prepare samples for detection of *Xf, in planta*.

DNA sequence variations include insertions, deletions, transitions, transversions, sequence rearrangements, distribution of random repeats (restriction fragment length polymorphism or RFLP, random amplified polymorphic DNA or RAPD), variable tandem repeats, etc. Sequence analyses showed that single nucleotide polymorphisms (SNPs) are commonly found in almost all genes and intergentic regions in *Xf* populations. We previously used the SNPs in the16S rDNA locus to define two genotypes correlated to *Xf* pathotypes associated with ALSD in California (Chen et al., 2005). The A-genotype only causes ALSD and the G-genotype causes both ALSD and PD. While the sequences of 16S rDNA are of taxonomic value, the reliability of a pathogen detection system will be improved based on comparisons of more genetic loci, ideally the whole genome.

Sequences of house-keeping genes are conserved and present in all bacterial strains, making them good candidates for bacterial species identification. For *Xf*, SNPs exist within these genes. They could be used for strain differentiation in a PCR or/and microarray format. The value of each SNP, however, varies. Validation is necessary to avoid the selection of false SNPs and to determine the taxonomic specificity of the target SNPs. On the other hand, for a nucleic acid-based pathogen detection system, sample preparation can be a bottleneck for a successful detection. Low efficiency of DNA extraction and the presence of DNA polymerase inhibitors are among the many factors leading to the failure of pathogen detection using PCR.

OBJECTIVES
The objectives of this project are to:
1. Use the complete and annotated genome sequences of the four *Xf* strains to identify SNPs in selected house-keeping genes.
2. Evaluate the potential of these SNPs for *Xf* pathotype / genotype identification.
3. Developed a simple but effective sample preparation procedure that minimize the presence of inhibitors and maximize the availability of nucleic acid.

RESULTS AND CONCLUSIONS
A BLAST search with randomly selected ORFs or intergenic sequences from strain Temecula against the four *Xf* whole genome sequence database showed that SNPs existed in every section of the bacterial genomes. One question was whether some SNPs might simply be due to sequence errors. The use of published genome sequences reduced the possibility of sequence error because of the high number of base coverage during the sequencing effort. Further confirmation was

performed using a large number of *Xf* strains. We found that some regions of the genome have more SNPs than other regions, suggesting different rates of sequence evolution. The strategy for *Xf* detection was to use the more conserved regions to secure the correct identification of the bacterial species (Chen et al., 2005). For *in planta* detection, this is particularly important because bacterial species of unknown taxonomy were expected to occur in the host tissue along with the pathogen. For this reason, we focused on the house-keeping genes.

Among the many house-keeping genes, five genes involving in tri-carboxylic acid cycle were selected. They were: PD0492 for malate dehydrogenase, PD0750 for citrate synthase, PD0234 for aconitase, and PD2056 for isocitrate dehydrogenase. Sequence lengths of the five genes ranged from 873 nt to 2232 nt with the number of SNPs ranging from19 to 55. As expected from existing phylogenetic data, there were more SNPs between strain 9a5c (CVC) and the three other North American strains (Temecula, Ann-1 and Dixon) as a group than among the three North American strains. PD0234 (aconitase) showed the highest number of SNPs (55). PD2056 (isocitrate dehydrogenase) ranked second with 44 SNPs. These two genes were selected for further evaluation.

Although higher in number, SNPs in PD0234 (aconitase) were more evenly distributed along the gene sequences in strains Temcula and Dixon. All the SNPs occurred singly with the exception of a doubled SNP between strain Ann-1 and the group of strains Temecula and Dixon. In PD2056 (isocitrate dehydrogenase), most SNPs were singly present. Positions 1,225 and 1,226 were doubled SNPs with an additional SNP occurring at position 1,228. To test the value of these SNPs, PCR primes were made by placing the double or near-triple SNPs in the center of the amplicons with the size of 97 and 90 nt for PD0234 and PD2056, respectively, using Primer 3 software. Primer specificity was checked by performing BLASTn search against all published DNA sequences in GenBank using concatenated primer sequences as queries (Chen et al., 2005). The results confirmed that these primers were specific to *Xf* with respect to known bacterial species.

Table 1. Comparative DNA amplification and melting temperature (Max T_m) of *Xf* A- and G-genotypes using PCR Primers targeting SNPs in the aconitase gene.

Strain	Genotype	C_t	Max T_m
Ju04	A	21.959	83.1
RL A2	A	21.772	83
Dixon	A	23.356	82.8
M12	A	21.496	83
RL 47	A	21.117	83.1
McC 9BL	A	25.974	82.7
MTO 4L	G	18.840	82.4
Price 19	G	20.466	82.3
M23	G	18.828	82.1
RS	G	19.938	82.2
Neg. Ctrl.		None	81.3

Table 2. Comparative DNA amplification and melting temperature (Max T_m) of *Xf* A- and G genotypes using PCR primers targeting SNPs in the isocitrate dehydrogenase gene.

Strain	Genotype	C_t	Max T_m
Ju04	A	21.846	79
RL A2	A	22.856	79.2
Dixon	A	24.232	79.2
M12	A	22.755	79
RL 47	A	21.889	79
McC 9BL	A	24.679	79
MTO 4L	G	20.790	77.5
Price 19	G	23.059	77.5
M23	G	24.413	77.5
RS	G	26.157	77.4
Neg. Ctrl.		None	83.4

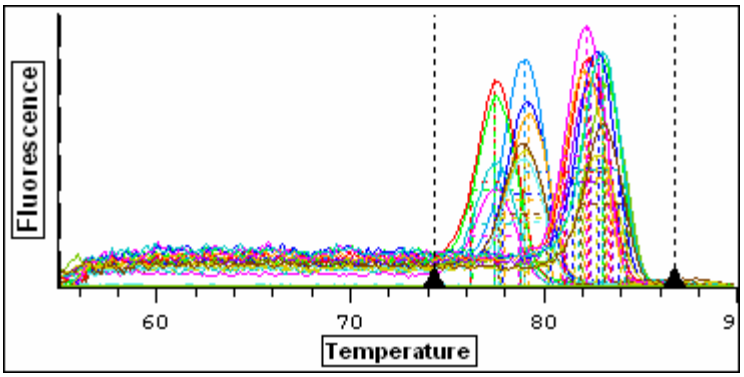

Figure 1. Capacity of genotype differentiation between the ICO primer set (two peaks on the left) and the ACO primer set (two peaks on the right) in melting point analysis. Also refer to Table 1 and 2.

Our working experience indicates that, while primer design and selection are important, sample preparation is currently the bottleneck for *Xf* detection *in planta*. Inhibitor(s) from host tissue and the efficiency of obtaining template DNA can directly

influence the success of PCR detection. Therefore, we developed two efficient sample preparation procedures for PCR amplification of *Xf* DNA from infected almond and grape petioles. A schematic illustration of the two procedures is shown in Figure 2.

In the freeze-dried method, petioles were pulverized into small particles (< 0.1mm in diameter based on our microscopic

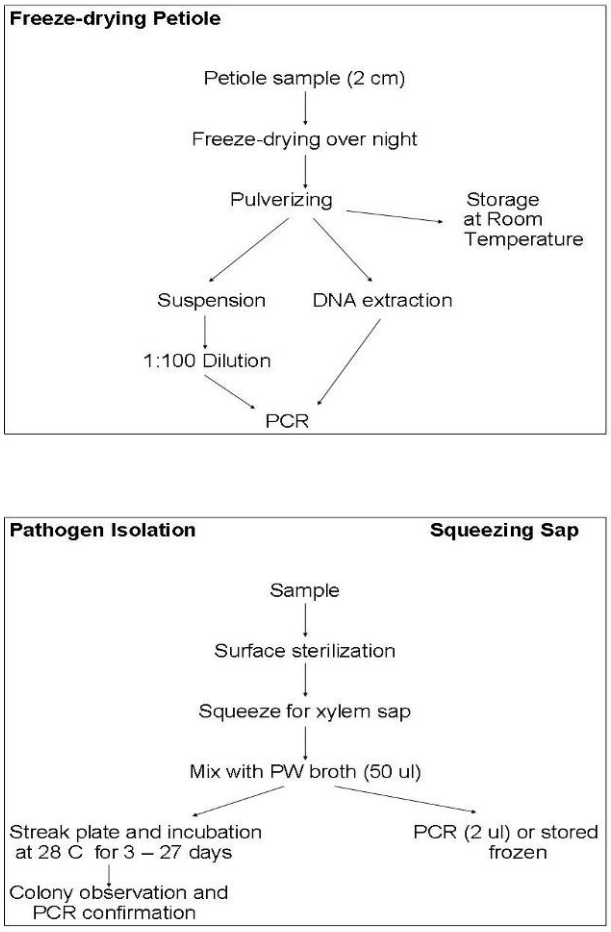

Figure 2. Procedures of sample preparation for PCR detection of *Xf*.

measurement). This increased the efficiency of DNA release from host tissue. Because the petiole was dried, oxidation and enzymatic degradation of samples during pulverization were minimal. Dilution reduced the effect of DNA polymerase inhibitors. But a balance was kept to avoid excessive dilution of the bacterial DNA beyond the detection range of PCR. The squeeze-sap method was often coupled with the bacterial isolation procedure. The amount of PW broth used did not seem to inhibit PCR. The squeeze-sap procedure has been routinely used in our laboratory as a quick check during the pathogen isolation.

In summary, our study indicates that SNPs are a good resource for the study of bacterial genome variation. They are particularly useful when whole genome sequence information becomes available, as in the case of *Xf*. When combined with the phytogenetic data, SNP analysis can be highly versatile and reliable. We have also developed simple procedures to prepare samples for *in planta* pathogen detection. Such procedures will contribute significantly towards the development of an efficient and accurate system for *Xf* identification and detection.

REFERENCES

Chen, J., R. Groves, E.L. Civerolo, M. Viveros, M. Freeman and Y. Zheng. 2005. Two *Xylella fastidiosa* genotypes associated with almond leaf scorch disease on the same location in California. Phytopathology 95:708-714.

FUNDING AGENCIES

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