TOWARDS THE DEVELOPMENT OF A SEQUENCE-BASED PCR SYSTEM FOR DETECTION AND GENOMIC STUDY OF *XYLELLA FASTIDOSA* STRAINS IMPORTANT TO CALIFORNIA

Project Leader:
Jianchi Chen
USDA – ARS
Parlier, CA 93648

Researcher:
Sam Livingston
Department of Plant Pathology
University of California
Davis, CA 95616

Cooperators:
Edwin Civerolo
USDA – ARS
Parlier, CA 93648

Bruce Kirkpatrick
Department of Plant Pathology
University of California
Davis, CA 96616

Reporting Period: The results reported here are from work conducted October 2005 to September 2006.

ABSTRACT
We report two of our recent efforts for improvement of PCR detection and genomic analysis of *Xylella fastidiosa (Xf)*. We evaluated the use of PCR primers from a hyper-variable locus to monitor diversity within strains of the same pathotype. The *psp*B (PD1208) locus, encoding a serine protease, was selected and analyzed. It was observed that tandem repeat sequences in *psp*B locus were highly variable from strain to strain. The biological significance of this hyper-variation is unknown. We also evaluated a simple sample preparation method for template DNA. The pulverized freeze-dried tissue PCR (PFT-PCR) test was compared to the "gold standard" pathogen isolation method. Our results indicated that PFT-PCR had a high predictive value (90.8%) for true positive samples, but a low predictive value for true negative results (29.7%), indicating that a PFT-PCR result is best suited to confirm the presence of *Xf* in a sample.

INTRODUCTION
The complete sequencing of *Xylella fastidiosa (Xf)* genome and their availability allow easy access to every genomic locus of the pathogen. With the availability of many primer design softwares in the Internet, it becomes highly feasible for many plant pathology laboratories to design PCR primers and explore their applications. These form the foundation for developing a genome based detection system for *Xf* using an array of primers from different loci on the platform of polymerase chain reaction (PCR). PCR is theoretically a highly sensitive and versatile technique for pathogen detection. A PCR experiment can be illustrated in the following formula:

PCR SUCCESS = PRIMER * SAMPLE PREPARATION (1)

Where a PCR success is the function of utilization of primers designed based on genomic information and sample preparation that provides DNA template. Either factor is equally critical in affecting PCR outcome.

The evolution and nucleotide variation rates of each gene or genomic locus vary. Therefore, primers designed from different genomic loci have different value in evaluating the bacterial population diversity and usage to define *Xf* strains. Several specific PCR primer sets are currently available for *Xf* detection including the most thoroughly tested RST31/33 primer set (Minsavage, et al., 1994), derived from the RNA polymerase genomic locus, and those derived from 16S rRNA gene (Chen et al., 2005), an important taxonomical character for the description of *Xf* (Wells et al., 1987). These primers target the conserved genes. Variations in these gene sequences are closely associated with pathogen pathotypes. For example, in San Joaquin Valley of California, the 16S rDNA G-genotype (G-genotype) strains cause both Pierce's disease (PD) of grapevine and almond leaf scorch disease (ALSD). The 16S rDNA A-genotype (A-genotype) strains cause only ALSD. Few studied have been performed on the use of less conserved or highly variable loci to study *Xf*. Information from the more variable loci could facilitate our understanding of the bacterial pathogenicity and environmental adaptations.

In contrary to PCR primers, sample preparation methods have been subjected to much less vigorous evaluation. The most common PCR detection procedures for *Xf* detection involve DNA extraction to generate template DNA. However, this severely reduces the high throughput capacity of a PCR procedure. Efforts were made to simplify or omit the DNA extraction procedure by using expressed plant sap as PCR DNA template. However, the results were inconsistent. To address this problem, we previously reported the development of a procedure using pulverized freeze-dried almond tissues for PCR detection of *Xf* (Chen and Civerolo, 2005).

In this report, we present the results of our recent analyses on using primers from a hyper-variable locus to evaluate the population diversity of *Xf* strains within the same pathotype/genotype. We also evaluated the procedure using pulverized freeze-dried almond tissues for PCR (PFT-PCR) detection of *Xf*. These are part of our effort in developing a comprehensive genome sequence-based detection system for *Xf* strains important to California.

OBJECTIVES

1. To identify a hyper-variable locus in the genome of *Xf* and analyzed the variability of the hyper-variable locus among different *Xf* strains within the same pathotype/genotype
2. To evaluate the procedure of using PFT PCR for the detection of *Xf* ALSD strains with an array of PCR primers from the genomic loci important to pathogenicity.

RESULTS

Evaluation of intra-pathotype/genotype variations

Using bacteria like the *Haemophilus influenzae* (De Bolle et al., 2000) as a reference, that changes the number of tandem repeats through several genetic mechanisms to regulate gene expressions during the course of environmental adaptation; we, with the help of internet softwares, identified the locus of *psp*B (PD1208), encoding a putative serine protease, from the genome of *Xf* Temecula strain causing PD. The N terminus of *psp*B is characterized by the presence of 23 tandem repeats of TP (threonine and proline). This converts the tandem nucleotide repeat of $[ACG(A)CCA]_{23}$. Orthologs are also found in the genomes of *Xf* strain Dixon, causing ALSD, strain Ann-1, causing oleander leaf scorch disease (OLSD) and strain 9a5c, causing citrus variegated chlorosis disease (CVCD). The number of the repeat varied from strain to strain. PCR primers flanking the tandem repeat region were designed and used to amplify DNA from over 90 *Xf* strains isolated from different hosts and geographical regions. PCR amplicons were sequenced. The difference in tandem repeat numbers among different strains were estimated by amplicon sizes and/or counted from the sequences.

Evaluation of PFT-PCR

The evaluation process involved two experiments. 1). Symptomatic samples were collected from 102 almond trees in an orchard in Kern County, California, in November, 2004. Small branches showing leaf scorching symptoms were excised, placed in labeled plastic bags, and transported in an ice cooler to the SJV Agricultural Sciences Center, Parlier, California. Upon arrival, samples were stored at 4 º C and processed within 24 hours. One symptomatic leaf was selected to represent one tree. Approximately, two cm petioles were removed from the leaves and used for both *Xf* isolation and for PFT-PCR. 2). During the growing season of 2006, two almond orchards in Fresno County, California, were selected based on the presence of ALSD in the previous year. One previously known ALSD infected tree and one asymptomatic tree were selected from each almond orchard. The branching patterns of each tree were mapped. Leaf samples were collected from labeled scaffolds starting in May when no leaf scorching symptoms were seen and throughout August when symptoms were obvious. Sample collection and processing were identical to experiment 1.

PCR procedure

PCR (25 µl) was carried out in 1x reaction buffer (10 mM Tris-HCl, pH8.3; 50 mM KCl; and 1.5 mM $MgCl_2$) with the addition of: 100 µM of dNTPs, 400 mM of each primer, 1 U of *Taq* DNA polymerase (TaKaRa taqTM Hot Start Version, Takara Bio Inc., Seta 3-4-1, Otsu, Shiga, 520-2193, Japan), and 2 µl of DNA templates. For PFT-PCR, The multiplex PCR procedure (Chen et al., 2005) was adapted for *Xf* detection. Amplification was performed in an MJ Research Tatrad II DNA engine with an initial denaturation at 96 º C for 10 min, followed by 30 cycles consisting of: denaturation at 96 º C for 30 s, annealing at 55 º C for 30 s, and extension at 72 º C for 30 s. The amplification products were then stored at 4 º C. The amplified DNAs were resolved through 1.5 % agarose gel electrophoresis and visualized by ethidium bromide staining.

For PCR array, primers were designed and synthesized from 30 putative *pil* genes and 10 putative *gum* genes based on the strain Temecula genome sequence. *Pil* genes encode proteins needed for Type IV pili formation partially responsible for the bacterial motility. The *gum* genes encode enzymes related to the production of extracellular polysaccharides. Both *pil* and *gum* genes are believed to be related in *Xf* pathogenicity. All primers were used to construct a PCR array to amplify DNA from different strains of *Xf*. All PCR amplicons were planned to be sequenced and sequence variations will be determined.

Unlike the conserved 16S rDNA locus, *psp*B was a hyper-variable chromosomal region among the *Xf* strains. In general, A-genotype strains was found to have higher number of repeats (>20) than that of the G-genotype strains (<20) (Figure 1, Table 1). However, strain Temecula and 59-3 (causing PD) and a strain isolated from muscadine in Georgia showed higher repeat numbers, similar to some of the A-genotype strains (Figure 1 and Table 1). On the other hand, the A-genotype strain Ann-1 showed a low repeat number of 13 similar to those of G-genotype strains. All primers from *pil* and *gum* amplified DNA from pure bacterial culture and from pulverized freeze-dried almond tissues.

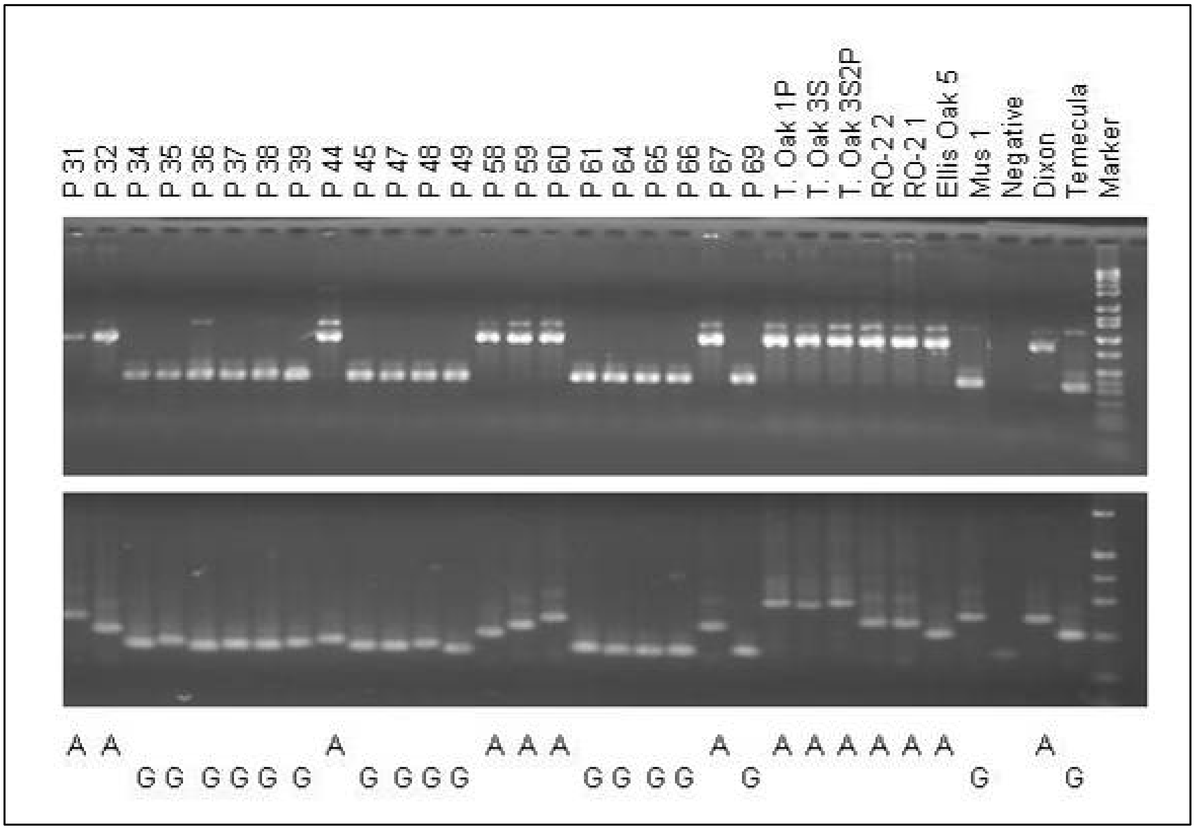

Figure 1. Comparison of DNA variations between the 16S rDNA locus (top) and the pspB locus (bottom) of *Xf* strains

Table 1. Comparison of tandem repeats of *Xf* G-genotype strains at the *psp*B locus.

	CCGCCA	ACGCCA	ACACCA	ACGCCA	ACACCA	Total
R77	1	3	9	0	0	13
R20	1	3	10	0	0	14
R29	1	3	10	0	0	14
R53	1	3	10	0	0	14
R10	1	3	11	0	0	15
R58	1	3	11	0	0	15
R75	1	3	11	0	0	15
R76	1	3	11	0	0	15
R14	1	3	12	0	0	16
R23	1	3	12	0	0	16
R63	1	3	12	0	0	16
R82	1	3	12	0	0	16
R27	1	3	13	0	0	17
Temecula	1	3	8	2	9	23
59-3	1	3	8	2	9	23

From the symptomatic trees, 85 out of 102 samples (83.3%) were positive based on pathogen isolation (Table 2). PFT-PCR detected 65 (63.7%) positive samples. Among the *X. fastidiosa* positive samples, 59 were shared by both methods. The true positive rate of PFT-PCR was 69.4%, 59/85) Table 2). Only 6 samples were PFT-PCR positive but isolation negative. These were considered to be false positive. The predictive value of PFT-PCR for a positive test was, therefore, calculated to be 90.8%. On the other hand, 11 *X. fastidosa*-negative samples by isolation and PFT-PCR were defined as true negative (64.7%, 11/17). Twenty-six PFT-PCR negative samples were in fact isolation positive and were considered to be false negative. The predictive value of the PFT-PCR method for a negative test was 29.7% (Table 2). The results of ALSD temporal development in four almond trees during 2006 are presented in Figure 2.

Table 2. Evaluation of PCR detection of *Xf* in pulverized freeze-dried almond petiole tissue.

		Pathogen isolation	
		Positive = 85	Negative = 17
PCR detection	Positive = 65	59 (a, True +)	6 (b, False +)
	Negative = 37	26 (c, False -)	11 (d, True -)

Predictive value of a Positive test = a/(a+b) = 59/(59+6) = 90.8%
Predictive value of a Negative test = d/(c+d) = 11/(26+11) = 29.7%

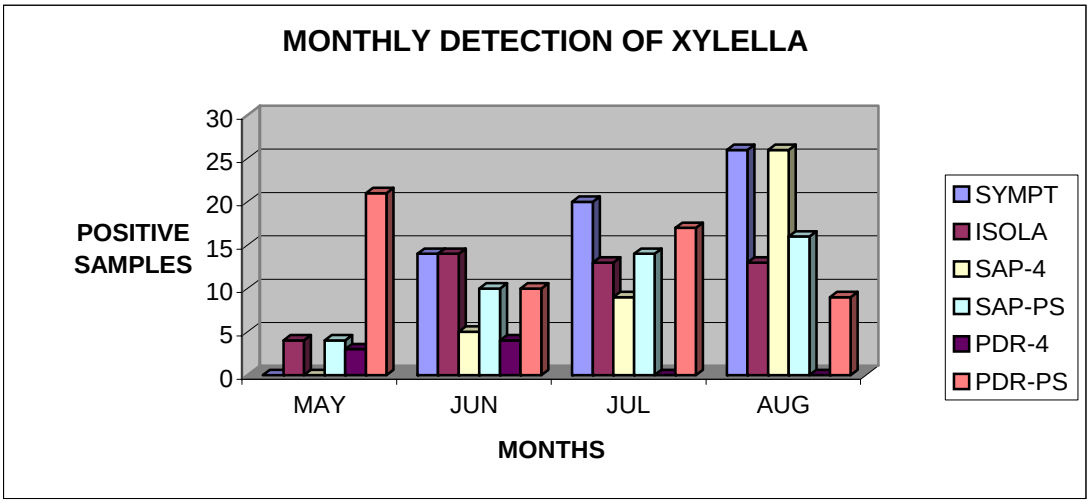

Figure 2. ALSD temporal development in during 2006. SYMPT: leaf scorch symptoms observed; ISOLA: isolation of *Xf* on PW-G media; SAP-4: sap PCR positive with 4 primers; SAP-PS: sap PCR positive with pspb primers; PDR-4: Freeze-dried powder positive with 4 primers; AND PDR-PS: Freeze-dried powder positive with pspb primers.

CONCLUSIONS

While SNPs in the 16S rDNA separate *Xf* strains into two genotypes/pathogtypes (Chen et al., 2005), sequences in *psp*B locus were found to have more variability (Figure 1). If repeat number (DNA fragment size) is considered, the distinction between the bacterial genotypes/pathotypes could be blurred. This means that the variation in *psp*B locus is independent from the bacterial pathotype. The *psp*B locus alone does not clearly define the two genotypes/pathotpes.

The continuous increment of hexomer number from strains in the same orchard (Table 1) implies that the *psp*B locus could be under a constant change. In other bacteria, DNA slippage during replication is one of the few mechanisms responsible for the change of tandem repeats. DNA slippage has not been reported in *Xf*. More interestingly, *psp*B encode a serine protease. Serine proteases participate in a wide range of cellular functions possibly including pathogenicity. Protease genes have been the research target for many bacterial pathogen including *Xf*. It is of high interest to investigate how the variation in tandem repeats will affect the expression of the serine protease.

Since PFT-PCR and pathogen isolation were performed simultaneously from the same petiole, we were able to evaluate PFT-PCR quantitatively by calculating the predictive value of a positive test and the predictive value of a negative test. As shown in Table 2, PFT-PCR had a high predictive value (90.8%) for the true positive samples, but a low predictive value for a true negative result (29.7%). These values suggested that the power of PFT-PCR is, in inferring a positive result rather than a negative result. That is, a PFT-PCR positive result suggests with high confidence, the presence of *Xf* in the sample. However, a PFT-PCR negative result does not appear to be a reliable indication of the absence of *Xf* in the sample.

As shown in Table 2, the pathogen isolation method detected 20 % more *Xf* positive samples than PFT-PCR. Pathogen isolation method is simple and is still a routine test in our laboratory when sample size is small and time is not a constraint. It should be noted that the bacteria were isolated from fully symptomatic leaves in September. Detection accuracy or reliability of both pathogen isolation and PFT-PCR are similar when asymptomatic samples were used (data not shown). For *Xf* detection in asymptomatic samples which is usually associated with a large sample size, PFT-PCR is advantageous. PFT-PCR is more rapid and much less labor intensive than *Xf* isolation and better suited for high throughput pathogen detection for other applications, such as epidemiological studies and evaluating germplasm for *Xf* resistance. The slight decrease in

detection sensitivity is justified by its throughput capacity. Our focus in future will be on the role of tandem repeat variation on the expression of *psp*B and its effects in the host-pathogen interactions using PFT-PCR techniques.

REFERENCES

Chen, J., and E.L. Civerolo. 2005. Evaluation of single nucleotide polymorphisms and sample preparation procedures in the detection of *Xylella fastidiosa* important to California, pp 152-154. In California Department of Food and Agriculture Pierce's Disease Research Symposium, 2005,

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

De Bolle, X., C.D. Bayliss, D. Field, T. van de Ven, N.J. Saunders, D. W. Hood, E.R. Moxon. 2000. The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. Mol Microbiol. 35:211-22.

Minsavage, G.V., C. M. Thompson, D.L. Hopkins, R.M. V. B.B. Leite, R.E. Stall. 1994. Development of a polymerase chain reaction protocol for detection of *Xylella fastidiosa* in plant tissue. Phytopathology 84: 456-461.

FUNDING AGENCIES

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

CHARACTERIZATION OF REGULATORY PATHWAYS CONTROLLING VIRULENCE IN *XYLELLA FASTIDIOSA*

Project Leader:
Donald A. Cooksey
Department of Plant Pathology
University of California
Riverside, CA 92521

Cooperators:
Xiangyang Shi
Department of Plant Pathology
University of California
Riverside, CA 92521

Korsi Dumenyo
Department of Plant Pathology
University of California
Riverside, CA 92521

Reporting Period: The results reported here are from work conducted October 2005 to September 2006.

ABSTRACT
We are pursuing a strategy to identify traits important in virulence of *Xylella fastidiosa* (*Xf*) through the mutagenesis of "global" regulatory genes, which are known to broadly regulate virulence functions in other microbes. In addition to phenotypic characterization of such mutants, we are using whole-genome microarrays to identify which genes are regulated by these global regulators and examine these genes as putative virulence factors. Here we report a specific example of this approach that has helped to define genes involved in aggregation, biofilm formation, and virulence of *Xf*. In previous work with *X. fastidiosa*, we mutated the global regulatory gene *rsmA*, and found a number of genes that were over-expressed in this mutant when grown in vitro, implying that these genes are normally repressed by the post-transcriptional regulator RsmA in the wild-type. In addition, the *rsmA*$^-$ mutant formed much more biofilm than wild type. Among the genes repressed by *rsmA* was another regulatory gene, *algU*, which regulates important virulence factors in *Pseudomonas*. In this study, an *algU::nptII* mutant had reduced cell-cell aggregation, attachment, biofilm formation, and lower virulence in grapevines. DNA microarray analysis showed that 42 genes had significantly lower expression in *algU::nptII* than wild type, including several genes which could contribute to cell aggregation and biofilm formation, as well as other physiological processes that could contribute to virulence and survival. Thus, *rsmA* appears to control biofilm formation and other traits partly through its repression of the positive regulator, *algU*.

INTRODUCTION
Many virulence genes in bacterial pathogens are coordinately regulated by "global" regulatory genes. The gene *rsmA*, for example, is known to regulate pathogenicity and secondary metabolism in a wide group of bacteria (Blumer and Haas, 2000; Mukherjee et al., 1996). Conducting DNA microarray analysis with mutants for such genes, compared with wild-type, can help to refine the list of genes that may contribute to virulence. We have reported on such an analysis with an *rsmA* mutant of *Xylella fastidiosa (Xf)*, and identified a number of genes that were overexpressed in the mutant (Cooksey, 2004). Among those were *pil* genes that have been subsequently confirmed to be important in twitching motility and long-distance spread of *Xf* in grapevines (Hoch and Burr, 2005; Meng et al., 2005), as well as enzymes or other structural proteins. In addition, a few genes controlled by *rsmA* were "secondary" regulatory genes, such as *algU*, which controls exopolysaccharide production in certain human and plant pathogens and contributes to virulence (Schnider-Keel, et al., 2001; Yu et al., 1995; Yu et al., 1999).

AlgU is a an alternative sigma factor whose role in regulation of biosynthesis of the exopolysaccharide (EPS) alginate has been extensively studied in *Pseudomonas aeruginosa* and *P. syringae*. Alginate functions as a virulence factor in *P. aeruginosa* during infection of cystic fibrosis patients (May and Chakrabarty, 1994), and also contributes to both virulence and epiphytic survival of the plant pathogen *P. syringae* (Yu et al., 1999). In *P. aeruginosa,* AlgU activates AlgU-dependent promoters of *algD* and *algR*. AlgR regulates *algC* and *algD* in cooperation with AlgU (Martin et al., 1994). *mucD* is a negative regulator of *algU* activity in *P. aeruginosa*. Homologs of *algU* (PD1284), *algZ* (PD1154), *algS* (PD0347), *algR* (PD1153), *algC* (PD0120), *algH* (PD1276) and *mucD* (PD1286) were detected in the *Xf* genome (Simpson et al., 2000; Van Sluys et al., 2003), but there are no homologs of the alginate biosynthesis genes *algA, algD, algG, algF, algI* and *algJ*. The alginate homolog genes in *Xf* are therefore probably not involved in alginate biosynthesis, but may be involved in synthesis of other EPS or of lipopolysaccharide (LPS), which could play a role in biofilm formation and cell attachment. In *P. aeruginosa*, the *algC* gene encodes a bifunctional enzyme that is involved in alginate production (phosphomannomutase activity) and lipopolysaccharide (LPS) production (phosphoglucomutase activity) (Coyne et al., 1994). We have constructed an insertional mutation in *algU* in *X. fastidiosa*, which reduced cell-cell aggregation, attachment, biofilm formation, and virulence. DNA microarray analysis of the *algU* mutant was then conducted to determine which genes it regulates.