MULTI-LOCUS SIMPLE SEQUENCE REPEATS AND SINGLE NUCLEOTIDE POLYMORPHISM MARKERS FOR GENOTYPING AND ASSESSING GENETIC DIVERSITY OF *XYLELLA FASTIDIOSA* IN CALIFORNIA

Project Leader:
Hong Lin
Crop Disease, Pest, &
Genetics Research Unit
SJVAC, ARS-USDA
Parlier, CA 93648
hlin@fresno.ars.usda.gov

Cooperators:
Harshavardhan Doddapaneni
Dept. of Vitic. & Enology
University of California,
Davis, CA 95616

Andrew Walker
Dept. of Vitic. & Enology
University of California
Davis, CA 95616

Edwin Civerolo
SJVAC, ARS-USDA
Parlier, CA 93648

Russell L. Groves
Department of Entomology
University of Wisconsin
Madison, WI 53706

Reporting Period: The results reported here from work conducted October 2004 to September 2006

ABSTRACT
In California, information regarding the population structure and genetic diversity as well as the genetic evolutionary and epidemiological relationships among *Xylella fastidiosa* (*Xf*) strains in agricultural populations is not clear. To develop effective management strategies, we need to understand pathogen population structure and genetic diversity in the agricultural ecosystem. Here we report development and utilization of two multilocous marker systems, Simple Sequence Repeats (SSR) and single nucleotide polymorphism (SNP) for genotyping and assessing genetic diversity of *Xf* in California. Strain diversity studies using SSRs on samples from different geographic populations and/or from different hosts demonstrated that host selection plays an important role in *Xf* genetic differentiation among agricultural populations in California. Whole-genome comparison of four sequenced strains identified 12,754 potential SNPs in coding sequences and 20,779 SNPs in non-coding regions across four *Xf* strains. Small scale validation (16 loci) tests showed that SNP genotype is tightly linked to the hosts from which the strains were derived. Together, SNP and SSR marker systems appear to be useful tools for pathogen detection and population genetic analyses.

INTRODUCTION
Host plant resistance is a critical component of integrated crop management. However, durability of resistant grape plants depends upon the variability and adaptability of the pathogen population as well as disease resistance gene(s). Population genetics research demonstrates that the evolutionary potential of a pathogen is reflected in its genetic diversity and its genetic structure. Pathogen populations with higher evolutionary potential are more likely to overcome host resistance than pathogen populations with a lower evolutionary potential (MacDonald and Linde, 2002). The resulting changes in population structure or virulence can lead to host resistance breakdown. Therefore, understanding pathogen genetic diversity is critical in developing an effective disease control strategy. To characterize population structures of *Xylella fastidiosa (Xf)* and to understand genetic diversity of *Xf* in agricultural systems, sensitive and accurate marker system(s) are required. The goal of this project is to develop a reliable marker system(s) that unambiguously identifies *Xf* strains from various geographic locations and host plants and further to understand the pathogen dynamics. Previously, we reported the development of multilocus simple sequence repeats (SSR) markers for *Xf* population genetic analysis (Lin et al., 2005). This marker system appeared to be sensitive in detection and powerful in discriminating *Xf* genotypes. This marker system also provides high throughput capability for a large scale population sample analyses. Recently, we developed a new marker system; single nucleotide polymorphism (SNP). We performed whole-genome sequence analysis of CVC, PD, ALS and OLS strains and identified potential SNP loci in both coding and non-coding regions (Doddapaneni et al., 2006). This marker system has proven to be powerful and reliable for distinguishing genetic relatedness. This marker system is very sensitive, has a high degree of specificity, and is quite powerful in detecting genetic polymorphism. Further, adaptability to high a through-put diagnostic platform makes this system an ideal tool for large scale studies of *Xf* population genetics and epidemiological risk assessment analyses.

OBJECTIVES
1. Analyze *Xf* seasonal population dynamics; spatial and temporal disease development and genetic diversity.
2. Comparative whole genome analyses of the *X. fastidiosa* strains to identify SNP loci and develop SNP based marker system for fingerprinting *Xf* strains.

RESULTS AND DISCUSSION
Objective 1.
Previously, we analyzed genetic diversities and geographic population structures of *Xf* in California vineyards (Napa, Sonoma and Kern and Riverside counties). Results based on multi-locus SSR marker systems and hierarchical sampling

showed that a larger proportion of total genetic diversity (68.89 %) was attributed to genetic variation among different host plants. These results suggest that host selection plays an important role in *Xf* genetic differentiation among agricultural populations in California. Using the same SSR marker system, we extended our investigation to *Xf*-induced, almond leaf scorch (ALS) disease in California's San Joaquin Valley (SJV). Survey data for two years of ALS disease incidence was analyzed to characterize the progress of almond leaf scorch disease development within selected orchards and to evaluate the seasonal population dynamics associated with *Xf* adaptation and host selection in almond-associated populations. The seasonal collection and detection studies showed that *Xf* populations were low in early season (March and April), when *Xf* is less likely detectable by PCR. *Xf* populations quickly increased after May which parallels increased plant growth and the activity of *Xf* vectors. Successful *Xf* isolation/culture and PCR detection were comparable after July. To further characterize if the spatial and temporal aspects of disease development associated with *Xf* genetic differentiation among geographic locations and/or with host adaptation, two almond orchards, Richline and McCall, were studied in Kern and Fresno Counties, respectively. In the Richline orchard, two almond cultivars including Sonora and Nonpareil were affected with ALS symptoms while only the cultivar Sonora was *Xf*-infected at the McCall Orchard. ALS samples from these three populations; Richline-Sonora, Richline-Nonpareil and McCall-Sonora were collected from May to October months. Using five SSR markers, allelic types and allelic frequencies of haplotypes among these three populations were analyzed (Table 1). Genetic relationships among three populations showed that *Xf* population derived from Richline-Sonora was closely related to the population of McCall-Sonora even though they are geographically separated while the *Xf* population from the Nonpareil cultivar in Richline orchard was genetically distant from the population derived from Sonora at either location (Figure 1). This result again confirms our earlier findings that suggests host selection plays an important role in determining population differentiation.

Our next experiment is to determine host-pathogen-vector interactions. In this study we will address questions of host selection, pathogen adaptation, and the role of vectors in transmitting *Xf* strains. Specifically, a vineyard adjoining an almond orchard was chosen. We will analyze gene flow between populations resulting from the movement of the vectors. *Xf* strains will be collected from PD-affected vines and ALS affected almond trees. We will also collect samples of adult green sharpshooters which are actively dispersing in these sites. The first sampling was conducted from September to October in 2006 and the early sampling (April-May) and late sampling (August-October) will be conducted in 2007. That way, over-wintering populations will be included for determination the effect of genetic drift or initial inoculum sources.

Table 1. Allele frequencies of three ALS *Xf* populations at five SSR loci.

	Population I (Site = Richline, Host = Sonora)			Population II (Site = Richline, Host = NonPareil)			Population III (Site = McCall, Host = Sonora)		
Locus	Allele (bp)	# of Observations	Allele Frequency	Allele (bp)	# of Observations	Allele Frequency	Allele (bp)	# of Observations	Allele Frequency
ASSR 19	279	16	0.64	279	2	0.08	279	27	0.93
	288	5	0.20	288	21	0.92	288	1	0.03
	294	4	0.16	294	0		294	1	0.03
OSSR 12	279	9	0.36	279	23	1.00	279	2	0.07
	308	16	0.64	308	0		308	27	0.93
CSSR 12	282	9	0.36	282	22	0.95	282	2	0.07
	291	16	0.64	291	1	0.05	291	27	0.93
GSSR 20	304	10	0.40	304	22	0.95	304	2	0.07
	332	4	0.16	332	1	0.05	332	20	0.69
	341	11	0.44	341	0		341	7	0.24
ASSR 4	366	5	0.20	366	22	0.95	366	2	
	372	4	0.16	372	0		372	2	0.07
	393	16	0.64	393	1	0.05	393	22	0.75
	397	0		397	0		397	5	0.17

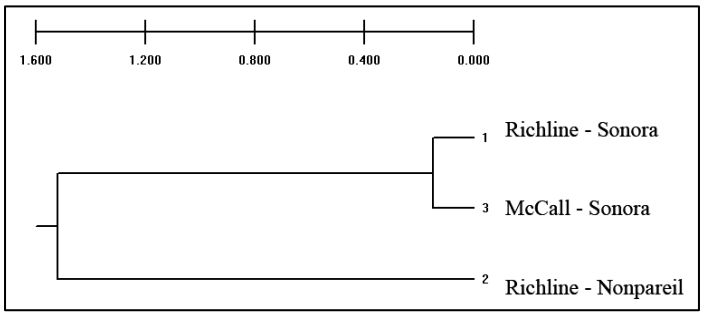

Figure 2. Genetic distances among three ALS *Xf* populations using UPGMA cluster analysis with Nei's coefficient.

Objective 2.
Comparative whole genome analyses of the *X. fastidiosa* strains: *Xf* Temecula-1 (grapevine strain), *Xf* 9a5c (citrus strain) *Xf* Ann1 (oleander strain) and *Xf* Dixon (almond strains) were carried out to identify SNP loci. The analyses identified 1,579 genes and 194 non-coding homologous sequences present in the genomes of all four strains, representing a 76.2% conservation of the entire genomic sequence. Our analysis shows that 51 Ann1 and Dixon genes were present at the end of sequenced contigs (partially sequenced) and were excluded from further analysis. Therefore, SNPs were identified in 1528 conserved gene alignment files. There were 12,754 potential SNPs in coding sequences and 20,779 SNPs in the non-coding sequences in these conserved regions. The average SNP frequency was 1.08×10^{-2} per base pair, which translates to approximately one SNP for every 93 bp of the DNA. SNPs were defined according to their nucleotide types as transversion or transition types. On average, 85.2% of SNPs are transversion type suggesting that transversion was the major type of SNP in *Xf*. In *Xf* 60.33% of the SNPs cause synonymous changes while 39.67% cause non-synonymous changes. For this analysis, those gene pairs that show internal INDELs that cause frame shift mutations were not included. A database called "Xfbase" has been developed under an IIS6.0 server using CGI scripts is available at http://cropdisease.ars.usda.gov/CVC_index.htm. A summary of whole genome SNP analyses is presented (Doddapaneni et al., 2006). We designed SNP detection primers for 16 SNP loci (Table 2) and validated by using 18 strains representing four strains from grape, citrus, almond and oleander plants. Strains were validated as *Xf* using HL5 and HL6 primers (Francis et al., 2006). The results clearly demonstrate that the SNP genotypes are associated with the hosts from which the strains were derived (Table 3). Next, 700-1,000 potential SNP loci located in interesting gene regions will be selected for screening using a set of 5-10 grape, almond, oleander and citrus *Xf* strains to confirm that these loci are truly polymorphic (to eliminate SNP caused by sequence error or isolated mutation event). To increase data throughput and reduce operational costs, we optimized multiplex PCR primers that can amplify up to 5 SNP loci in a single reaction. Our recently designed HTP format protocol allows analysis of up to 10 SNP loci per sequencing run. In the future, genome wide multiplex SNP detection primers based on screening results (~100 loci) will be analyzed. This *Xf* SNP genetic analysis system appears to be a powerful new marker system for *Xf* genetic study.

Table 2. Details of the loci and primers used for SNP validation. Allele frequency information is mentioned in the next table.

Gene ID	Description	Left primer (5'-3')	Right primer (5'-3')
XF0450	Two-component system, regulatory protein	ATTACCGCAACCGATGG	TCGTTCGCTTTTGCTTTTG
XF0677	Type 4 fimbriae assembly protein	GTACGCGCCAGGGTATTCT	CAAGCAACGTCTCAATGC
XF0845	Family 3 glycoside hydrolase	TTCTTCCGTCAAGACAACG	GCCGGAGTTTTTCAAGAGG
XF1267	1,4-beta-cellobiosidase	TTACGAAGAAGGCATAAAA TATG	GCAAAACCATTGACACTAGC
XF1532	Oxidative stress transcriptional regulator	GTGCGCGTAGCATTGTTG	ACACGAACGGCTCCTCAA
XF2352	Cold shock protein	ATGCAGAGCGGTACAGTTA AG	TATTGGCGTGATATTCGATG
XF2545	Two-component system, regulatory protein	ACGTATGGGGCTGCGTAT	CGATTGTTCAATTCCAAAGC

CONCLUSIONS
SSR and SNP marker systems appear to be useful for strain identification and for analysis of genetic diversity. The multi-locus SSR marker system is particularly suitable for *Xf* population genetic study due to its powerful and unambiguous discrimination ability of genetically related strains across independent genetic loci. Information generated from SNP markers will advance our knowledge in the understanding of *Xf* variation associated with functional gene sets which can be used to define the genome-wide Linkage Disequilibrium blocks and in linking SNP genotype to the *Xf* phenotype. Results from this project will be used to generate comprehensive SSR allele frequency and SNP association databases. These two databases will be complementary in strengthening the power of SSR marker in strain discrimination and the power of SNP marker for functional-related genotyping which can aid in *Xf* epidemiological and strain virulence studies.

Table 3. Single Nucleotide Polymorphism data at different genomic loci across multiple strains of *Xf*. A total of 16 genes that are conserved in all four sequenced strains and showing SNPs were selected based on their putative function. Primers were designed to amplify the SNP regions in 4-5 host specific strains of grape, citrus, almond and oleander. Target regions were PCR amplified, sequenced and aligned. The data presented here shows SNPs in functional genes that differentiate host-associated strains.

		Polymorphism at selected gene loci								
Host	**Strain, Location**	**XF 1532** I(77bp) *	**XF 1532** II (101bp)	**XF 1267** I (263 bp)	**XF 1267** II (293 p)	**XF 2545** (106 bp)	**XF 2352** (86 bp)	**XF 0845** (258 bp)	**XF 0677** (125 bp)	**XF 0450** (151 bp)
Grape	Stag leap, Napa, CA	C	C	G	A	C	T	A	T	G
	Temecula, CA	C	C	G	A	C	T	A	T	G
	PD-6,Riverside, CA	C	C	G	A	C	T	A	T	G
	PD-7,Riverside, CA	C	C	G	A	C	T	A	T	G
	Xf-10, Kern, CA	C	C	G	A	C	T	A	T	G
Citrus	9a5c, Brazil	T	G	G	A	C	C	C	C	A
	CVC-10, Brazil	T	G	G	A	C	C	C	C	A
	CVC-12, Brazil	T	G	G	A	C	C	C	C	A
	CVC-14, Brazil	T	G	G	A	C	C	C	C	A
	CVC-16, Brazil	T	G	G	A	C	C	C	C	A
Oleander	Cathedral	T	G	G	A	T	T	A	T	G
	TIB	T	G	G	A	T	T	A	T	G
	T1	T	G	G	A	T	T	A	T	G
	T5	T	G	G	A	T	T	A	T	G
Almond	Manteca	T	G	A	G	C	T	A	T	G
	ALS-2, Costa, CA	T	G	A	G	C	T	A	T	G
	ALS-6, Solano, CA	T	G	A	G	C	T	A	T	G
	ALS-7	T	G	A	G	C	T	A	T	G
Allele frequency		5C:14 T	5C:14 G	4A:15G	4G:15A	4T:15C	5C:14 T	5C:14 A	5C:14 T	5A:14G

* Base pair number in parenthesis indicates the location of the target SNP nucleotide in the predicted coding sequence. Details can be obtained from the multiple alignments at our website.

REFERENCES:

Doddapaneni H, J. Yao, H. Lin, M.A. Walker and E.L. Civerolo. 2006. Analysis of the genome-wide variations among multiple strains of the plant pathogenic bacterium *Xylella fastidiosa*. BMC Genomics. 7:225.

Lin, H., E.L. Civerolo, R. Hu, S. Barros, M. Francis, and M.A. Walker. 2005. Multilocus simple sequence repeat markers for differentiating strains and evaluating genetic diversity of *Xylella fastidiosa*. Applied & Environmental Microbiology 71: 4888-4892.

Francis M, H Lin, J.R. Cabrera-La, H Doddapaneni and E.L. Civerolo. 2006. Genome-based PCR primers for specific and sensitive detection and quantification of *Xylella fastidiosa*. European J of Plant Pathol. 115: 203-213.

McDnald, B., C Linde. 2002 Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40:349-379.

FUNDING AGENCIES

Funding for this project was provided by the CDFA Pierce's Disease and Glassy-winged Sharpshooter Board.

ASSESSMENT OF THE PROCESS OF MOVEMENT OF *XYLELLA FASTIDIOSA* WITHIN SUSCEPTIBLE AND RESISTANT GRAPE VARIETIES

Project Leader:
Steve Lindow
Plant and Molecular Biology
University of California
Berkeley, CA
icelab@socrates.berkeley.edu

Cooperators:
Clelia Baccari
Plant and Molecular Biology
University of California
Berkeley, CA
cbaccari@nature.berkeley.edu

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

ABSTRACT
We are studying the process of movement of *Xylella fastidiosa* (*Xf)* cells between xylem vessels and through plants by analyzing the changing proportion of genetically distinct strains, initially introduced into the plants by distance and time from point of inoculation. We are also determining whether bottlenecks in movement of *Xf* cells in plants are more extreme in resistant plants than in susceptible plants, and whether this phenomenon can be exploited as a tool to screen germplasm for resistance to *Xf.* We expect that the process of movement of *Xf* involves a progressive and sequential colonization of a large number of xylem vessels that is limited by anatomical features of plants (nature of pit membranes and other barriers to vessel to vessel movement in the stem). The resulting in bottlenecks practically limit the number of *Xf* cells that can move from one vessel to another, and thus constitute a major factor that confers resistance in plants.

INTRODUCTION
Xylella fastidiosa (Xf) has a rather unique means of colonizing plants and causing symptoms, which make strategies of disease control that are useful in other bacterial diseases ineffective. Many agriculturally important plants besides grapevines, including citrus, almond, alfalfa, and coffee, are susceptible to diseases caused by *Xf* (Hopkins 1989). The bacterium is transmitted to new host plants during xylem sap feeding by sharpshooter vectors and then multiplies and spreads from the site of inoculation to colonize the xylem; a water transport network of vessels composed of dead, lignified cells. Vessels are interconnected by channels, called bordered pits, that allow the passage of xylem sap but block passage of larger objects due to the presence of a pit membrane (Choat et al. 2003, Esau 1977). Bacterial cells attach to the vessel wall and multiply, forming biofilm-like colonies that can, when sufficiently large, occlude xylem vessels, blocking water transport (Alves et al. 2004, Fry and Milholland 1990a, Newman et al. 2003). In susceptible plants, leaf scorching, fruit shriveling and other symptoms result, likely due to the increased stress of xylem blockage as colonization ensues. However, within the majority of host plants, *Xf* behaves as a harmless endophyte (Freitag 1951). The population size of *Xf* in grapevines resistant and susceptible to Pierce's disease (PD) is highly correlated with symptom expression (Alves et al. 2004, Fry and Milholland 1990a, Fry and Milholland 1990b, Hopkins 1981, Newman et al. 2003, Krivanek and Walker 2004). A much higher proportion of vessels are colonized by *Xf* in symptomatic tissues that in non-symptomatic tissues (Newman et al. 2003). However we still lack an understanding of the process of colonization and what specifically about high populations of *Xf* leads to symptom expression. The pathways by which water moves through plants via the xylem are spatially complex. It is simplistic to consider axial water movement in stems via xylem vessels as simple vertical "pipes". Indeed, xylem vessels themselves often follow complicated paths through a tissue with respect to each other (Figure 1). More importantly, the water in a given vessel is in contact with that of different vessels as well as with those in adjacent tracheids via the many pits in the cell walls (Figure 1). Pits are adjacent to one another on either side of the cell wall and thus come in pairs. The pit membrane is composed of the primary cell wall and middle lamella of adjacent cell walls of the pit pairs (Esau 1977). In a bordered pit the secondary cell wall forms a border over the pit membrane leaving a small opening called a pore. While secondary cell walls can be thickened via the intrusion of lignin and other polymers into the cellulosic matrix of the primary cell wall, pits represent local "thinning" of the primary wall with only a minimal amount of cellulose and pectin, which allows relatively free diffusion of water and solutes from one cell to another. Thus instead of a limited number of vertical "pipes' that conduct water through a stem, there are thousands of alternative pathways that water might travel in a tissue. The interconnectivity of the xylem cells is presumably one means by which the plant overcomes injuries or other insults that would disrupt the movement of water via a given xylem element by shunting it to adjacent cells.

In the context of PD it thus becomes obvious that in order for water movement in a stem to be so restricted that disease develops, a large percentage of the xylem pathways must be blocked for disease to occur. Yet, while over 40% of the xylem vessels in a single section of an infected grape stem may be infested with *Xf* (Newman et al. 2003) this alone is unlikely to explain water stress. Serial sections of grape tissue however, demonstrated that different xylem vessels are blocked in different cross sections; the percentage of occluded vessels in one of several sections along 5 mm of petiole was five times that of a single cross section. Given that inoculation of grape with *Xf* must occur in a relatively few sites on a vine, it is clear that the pathogen has the ability to move both axially and radially in xylem tissues. Such extensive movement must take some time, explaining why the disease is "progressive" and appears only several weeks after inoculation.