## MULTIPLEX PCR MARKERS: STEPS TOWARDS SATURATING THE *XYLELLA FASTIDIOSA* GENOME

**Principal Investigator:**
Jeff Brady
Texas AgriLife Research
Stephenville, TX 76401
j-brady@tamu.edu

**Co-Principal Investigator:**
Forrest Mitchell
Texas AgriLife Research
Stephenville, TX 76401
f-mitchell@tamu.edu

**Researchers:**
Jennifer Faske
Texas AgriLife Research
Stephenville, TX 76401
JBFaske@ag.tamu.edu

Blake Myers
Health Science Center
University of North Texas
Ft. Worth, TX 76107
Blake.myers@unthsc.edu

Rebecca Ator
Dept. of Mathematics
Tarleton State University
Stephenville, TX 76402
rebecca.ator@go.tarleton.edu

**Cooperator:**
Harold Rathburn
Dept. of Biological Sciences
Tarleton State University
Stephenville, TX 76402
rathburn@tarleton.edu

**Reporting Period:** The results reported here are from work conducted March 2010 to December 2010.

**ABSTRACT**
*Xylella fastidiosa* (*Xf*) is a bacterium that can cause leaf scorch disease in many plant species, is an asymptomatic colonizer of many plant species, and cannot successfully colonize other plants (Hopkins et al. 2002). Although some molecular determinants of successful colonization have been discovered (Chatterjee et al. 2008), much work remains in order to unravel the genetics of host specificity in various *Xf* strains. Additionally, the same can be said of virulence determinants. Many questions remain about why certain strains cause disease in one plant species but not in other species (Almeida et al. 2008). This project is an attempt to develop a high-density DNA marker system that can quickly and efficiently screen for variation in the *Xf* genome at hundreds of loci. Hundreds of highly conserved *Xf* genes have been aligned using the seven fully sequenced *Xf* genomes currently available, and the alignments have been screened for informative DNA polymorphisms. A multiplex PCR strategy using amplification tags has been employed (Boutin-Ganache et al. 2001), and electrophoresis in capillary sequencing instruments is being utilized to accurately assess DNA amplicon size differences. A 50 marker proof-of-concept test has been conducted and marker number is being increased towards a goal of 400 total. An evaluation of *Xf* strain collections with a high-density marker system should facilitate the identification of additional genetic factors influencing host specificity and virulence, and should provide additional information about recombination frequency among *Xf* strains.

**LAYPERSON SUMMARY**
Different strains of the bacterium *Xylella fastidiosa (Xf)* cause many different plant diseases of economic significance, such as Pierce's disease of grape and citrus variegated chlorosis. Different strains of *Xf* are specific to certain host plants and cause disease in a small subset of potential host plants. Many of the genetic factors that provide host specificity and development of disease remain unknown. This project is an effort to develop enough molecular markers in the *Xf* genome so that additional genetic determinants of host specificity and disease can be mapped using *Xf* strain collections.

**INTRODUCTION**
Genetic tests for *Xylella fastidiosa (Xf)* range from single gene tests that only detect presence or absence of any strain of the bacterium, either by conventional PCR (Minsavage et al. 1994) or qPCR (Schaad et al. 2002), to complete genome sequencing projects that sequence every base pair (Simpson et al. 2000). At one end of the spectrum, the single gene tests produce almost no genetic information except that the target gene is conserved and present. Either conventional PCR or qPCR are rapid, inexpensive, and simple to analyze. At the other end of the spectrum, complete genome sequencing is expensive, though cost is decreasing rapidly, and time-consuming. Although automated annotation of a genome sequence can be accomplished with a limited number of keystrokes, the important biological data can be difficult to discern due to the sheer volume of information to contend with. Indeed, each of the seven completely sequenced *Xf* genomes contain important answers to *Xf* biology that are now unexplained, even though many of the sequences have been available for years, and these answers will be explained in months and years to come as proper questions are framed. Genetic tests that are intermediate in information content include multiplex PCR assays (Hernandez-Martinez et al. 2006) and multilocus sequence typing (MLST) (Scally et al. 2005). Multiplex PCR is as rapid and inexpensive as conventional PCR and provides information about both presence/absence and also produces a limited amount of genetic information to distinguish subspecies as well. Multiplex PCR lacks the sensitivity of detection that single gene assays can achieve, and thus far has been used to amplify a limited number of loci. MLST involves sequencing seven housekeeping genes and it produces a few kilobases of sequence data. The method is more time-consuming, more laborious, and more expensive than multiplex PCR, but it produces much more genetic information than any method other than complete genome sequencing. The amount of data produced by MLST is easily analyzed, will easily discriminate between subspecies and strains, and can even detect genetic recombination.

While each of the genotyping methods described above may be appropriately used to answer different questions related to *Xf* genetics, no method currently available can rapidly survey a large number of genetic loci. This project is an effort to create a

large number of genetic markers to saturate the *Xf* genome so that a genetic fingerprint can be rapidly generated at minimal expense. Presently the goal is to generate 400 informative DNA markers. Each individual marker will have low information content, but together, due the number of markers, a highly informative genetic fingerprint will be generated. Roughly assuming an average *Xf* genome size of 2.5 Mb, and roughly assuming an average of 2,500 genes per *Xf* genome, 400 markers would give average genome coverage of one marker every 6250 bp, or one marker every 6.25 genes among *Xf* strains.

A densely-saturated marker system can be utilized in several ways. There are a number of questions concerning *Xf* host specificity determinants and *Xf* virulence determinants that could use additional tools such as these to help identify important genomic regions in given strains that may be involved. Very little is known about the genetics of host specificity in *Xf*. Any new information would be beneficial. While there are a number of genes that have been identified as generally important in the colonization of plants (Reddy et al. 2007; Chatterjee, Wistrom et al. 2008), few are plant host specific. The possibility exists that the genetic components of disease response are entirely in the plant genome. Given the highly interconnected nature of other plant-microbe interactions, this seems unlikely, and a high-density marker system may help identify important genomic regions common to most *Xf* infecting grape, for example. Additionally, this marker system should provide a substantive assessment of the importance of horizontal gene transfer in *Xf* strains. In other gamma proteobacteria such as *Escherichia coli*, horizontal gene transfer has been shown to be much more prevalent among virulent strains than among commensal strains (Wirth et al. 2006). Virulent *E. coli* often possess a mosaic genome composed of genetic segments from several other *E. coli* groups, and additionally show higher levels of recombination and mutation than do non-pathogenic groups. The genes responsible for horizontal gene transfer exist in the *Xf* genome (Monteiro-Vitorello et al. 2005), and recombination is detected in MLST studies using only seven genes and a limited number of isolates. This project should, on a broader scale, complement previous studies documenting genetic recombination in *Xf* strains, and could potentially help identify additional genomic regions and genes associated with pathogenicity.

**OBJECTIVES**

1. Identify and align 450 informative sequence polymorphisms using all seven fully sequenced *Xf* genomes.
2. Design 200 insertion/deletion (indel) and 200 single nucleotide polymorphism (SNP) assays for subspecies and strain identification.
3. Use multiplex PCR amplification to efficiently create highly informative genetic fingerprints in a single day.

**RESULTS AND DISCUSSION**

Complete genomes for the seven fully sequenced *Xf* strains have been downloaded and are being used to create gene alignments for conserved genes. Previous bioinformatics projects have identified conserved genes among all the sequenced genomes (Doddapaneni et al. 2006). The conserved genes are being aligned and screened visually for informative polymorphisms using VectorNTI software (Invitrogen, Carlsbad CA).

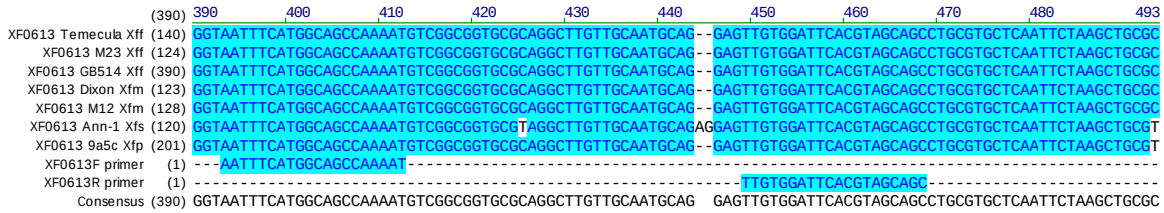


**Figure 1**. DNA sequence alignment of the 7 available *Xf* genomes at the XF0613 locus. A representative 2 bp indel polymorphism that would allow discrimination of *Xfs* from other subspecies is included in this portion of the alignment. Forward and reverse primer binding sites are also indicated.

In the first iteration of the project 200 indel markers are being developed. A typical alignment is shown in **Figure 1**, with the indel in the center of the sequence. Primers are designed that flank the indel marker in areas of conserved sequence using the primer3 website. Primers are designed to fit in size groups so that PCR amplicons are between 100 bp and 400 bp in size. The 200 indel markers are being amplified in multiplex PCR reactions with approximately 10 loci per reaction. The entire 200 gene indel set will require 20 PCR reactions for marker generation. Forward primers are labeled with an amplification tag as previously described (Boutin-Ganache, Raposo et al. 2001). Depending on the capillary electrophoresis instrument, several color channels can be used for efficient data collection. We are currently using ABI 3130 sequencing instruments capable of three colors for samples and one color for size markers. Amplification reactions labeled with FAM, HEX, and NED can be pooled into the same well following PCR and data can be collected simultaneously for three different bacterial strains. Gene designations follow those of the *Xfp* genome sequencing project (Simpson and ONSA 2000). A 10-plex multiplex grouping with expected amplicon sizes for the seven sequenced strains is shown in **Table 1**. A representative electropherogram from a 11-plex PCR reaction is shown in **Figure 2**. A 50 marker proof-of-concept experiment has already been carried out with both local isolates and fully sequenced type strains.

**Table 1**. Expected amplicon sizes for 10 loci amplified by multiplex PCR.

| Locus | Temecula *Xff* | M23 *Xff* | GB514 *Xff* | Ann-1 *Xfs* | Dixon *Xfm* | M12 *Xfm* | 9a5c *Xfp* |
|---|---|---|---|---|---|---|---|
| XF0053 | 88 | 88 | 88 | 95 | 95 | 95 | 88 |
| XF1419 | 112 | 112 | 112 | 109 | 109 | 109 | 109 |
| XF0294 | 129 | 129 | 129 | 127 | 127 | 127 | 127 |
| XF1470 | 158 | 158 | 158 | 159 | 154 | 154 | 154 |
| XF2759 | 179 | 179 | 179 | 197 | 197 | 197 | 197 |
| XF2316 | 237 | 237 | 237 | 230 | 230 | 230 | 230 |
| XF0388 | 267 | 267 | 267 | 267 | 269 | 269 | 269 |
| XF1972 | 292 | 292 | 292 | 292 | 293 | 293 | 286 |
| XF1267 | 345 | 345 | 345 | 345 | 355 | 355 | 345 |
| XF0454 | 371 | 371 | 371 | 380 | 380 | 380 | 381 |

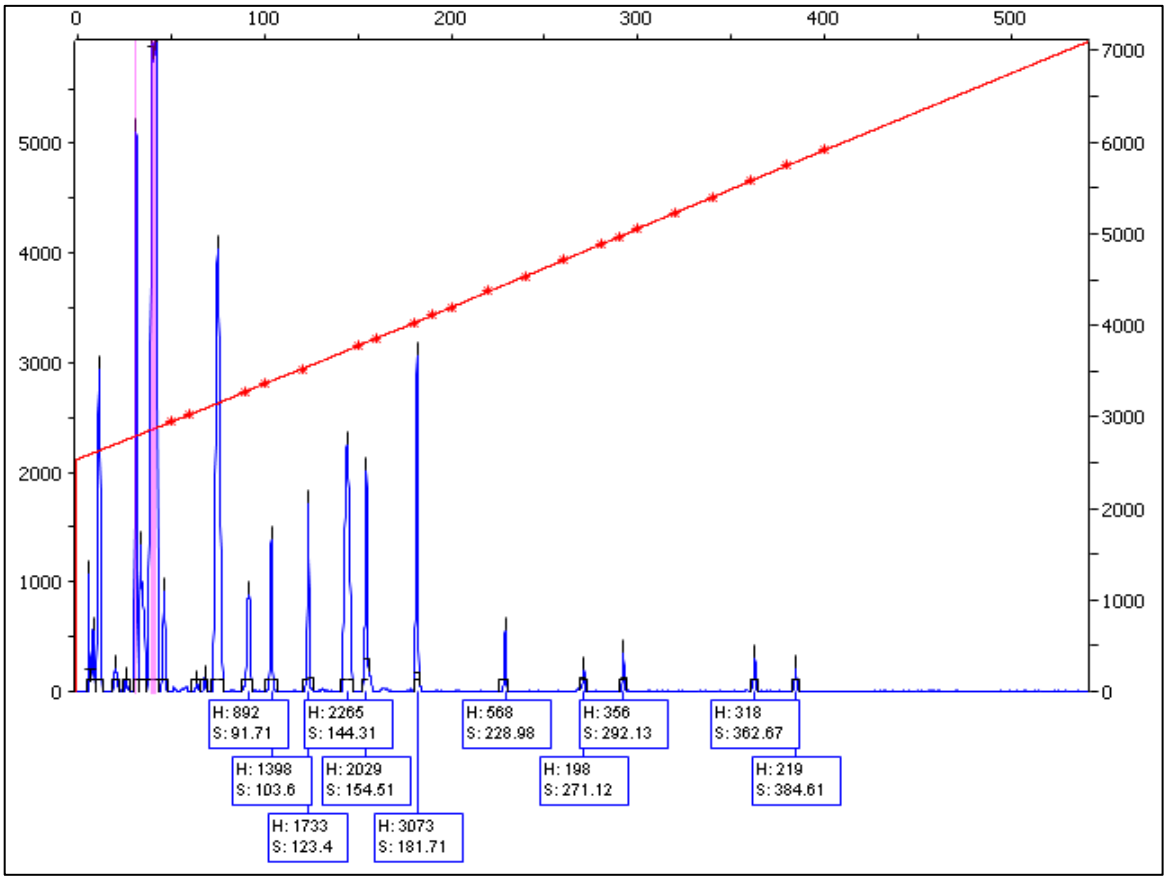


**Figure 2**. A representative 11-plex amplification of *Xf* indel loci. Peak sizes in the electropherogram are determined by comparison to internal size standards contained in every sample. Size standards are indicated by asterisks on the red line above the samples.

The second iteration of the project involves the development of 200 SNP markers. SNP markers are ubiquitous in *Xf* gene comparisons, so that in cases where no suitable indel exists for amplification in a given *Xf* conserved gene, there are invariably several SNPs in the gene that can be used for marker development. The level of multiplexing possible with SNP markers should be higher than is possible with indel markers. By using a similar amplification tag strategy to that used for the indel markers with the addition of size adapters (Lindblad-Toh et al. 2000), an even greater number of SNPs can be assayed per well than is possible with indel multiplex reactions, because the single base extension amplification products are precisely known sizes *a priori*. SNP markers can thus be designed in closer size ranges so that more can be analyzed per lane in single capillary gel runs.

**CONCLUSIONS**

In this project we are developing a large number of molecular markers to screen the *Xf* genome for important traits, to detect recombination levels, and to create a rapid and informative genetic fingerprinting system. Numerous markers have been

developed to date, and proof-of-concept experiments indicate that the system will perform as expected. At this time, marker development is approaching the halfway point.

**REFERENCES CITED**


Almeida, R. P., F. E. Nascimento, et al. (2008). Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. Applied and Environmental Microbiology 74(12): 3690-3701.

Boutin-Ganache, I., M. Raposo, et al. (2001). M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. BioTechniques 31(1): 24-26.

Chatterjee, S., C. Wistrom, et al. (2008). A cell-cell signaling sensor is required for virulence and insect transmission of *Xylella fastidiosa*. Proceedings of the National Academy of Sciences, USA 105(7): 2670-2675.

Doddapaneni, H., J. Yao, et al. (2006). Analysis of the genome-wide variations among multiple strains of the plant pathogenic bacterium *Xylella fastidiosa*. BMC genomics 7: 225.

Hernandez-Martinez, R., H. S. Costa, et al. (2006). Differentiation of strains of *Xylella fastidiosa* infecting grape, almonds, and oleander using a multiprimer PCR assay.

Hopkins, D. L. and A. H. Purcell (2002). *Xylella fastidiosa*: Cause of Pierce's disease of grapevine and other emergent diseases. Plant Disease 86(10): 1056-1066.

Lindblad-Toh, K., E. Winchester, et al. (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nature Genetics 24(4): 381-386.

Minsavage, G. V., C. M. Thompson, et al. (1994). Development of a polymerase chain reaction protocol for detection of *Xylella fastidiosa* in plant tissue. Phytopathology 84(5): 456-461.

Monteiro-Vitorello, C. B., M. C. de Oliveira, et al. (2005). *Xylella* and *Xanthomonas* Mobil'omics. OMICS 9(2): 146-159.

Reddy, J. D., S. L. Reddy, et al. (2007). TolC is required for pathogenicity of *Xylella fastidiosa* in *Vitis vinifera* grapevines. Molecular Plant-Microbe Interactions 20(4): 403-410.

Scally, M., E. L. Schuenzel, et al. (2005). Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. Applied and Environmental Microbiology 71(12): 8491-8499.

Schaad, N. W., D. Opgenorth, et al. (2002). Real-time polymerase chain reaction for one-hour on-site diagnosis of Pierce's disease of grape in early season asymptomatic vines. Phytopathology 92(7): 721-728.

Simpson, A. J. G. and ONSA (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature 406(6792): 151-159.

Wirth, T., D. Falush, et al. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. Molecular Microbiology 60(5): 1136-1151.


**FUNDING AGENCIES**


Funding for this project was provided by a USDA Animal and Plant Health Inspection Service contract under the Texas Pierce’s Disease Research and Education Program.