GENETIC ANALYSIS OF THE *ZONULA OCCLUDENS TOXIN* (*ZOT*) GENE IN TEXAS ISOLATES OF *XYLELLA FASTIDIOSA*

Principal Investigator:
Blake Bextine
Dept. of Biology
University of Texas-Tyler
Tyler, TX 75799
BBextine@uttyler.edu

Co-Principal Investigator:
Lisa Morano
University of Houston-Downtown
Houston, TX 77002-1001
MoranoL@uhd.edu

Researcher:
Henry L. Schreiber IV
Dept. of Biology
University of Texas-Tyler
Tyler, TX 75799
HenryLSchreiber@gmail.com

ABSTRACT
Multiple subspecies of the phytopathogenic bacterium *Xylella fastidiosa (Xf)* exist which are pathogenic to distinct plant hosts, such as grapes, oleander, almonds, and citrus. Previously, DNA sequence analysis of the *mopB* and *gyr*B genes has been used to separate *Xf* strains into their subspecies groups. In this study, DNA sequence analysis of the *Zonula occludens toxin (Zot)* gene was used to corroborate the genetic variation found between 44 Texas strains of *Xf*. This approach provided variable gene sequences that allow for categorization of *Xf* at both the subspecies and population level. *In silico* translation of the *Zot* gene sequence, and subsequent protein model predictions showed conserved secondary structure, transmembrane regions, and signal cleavage sites despite differences in amino acid code. Results of the analysis of this gene were consistent with the phylogeny found using the more conserved *mop*B and *gyr*B genes at the subspecies level and can be used to differentiate populations within subspecies. The analysis of these variable genes and gene regions provide additional opportunities for new diagnostic and disease management techniques.

LAYPERSON SUMMARY
In this study, we sequenced one of the genes, the *Zot* gene, in the *Xylella fastidiosa* (*Xf*) genome. *Xf* has been implicated as the cause of several plant diseases that cause plant death and crop loss, including Pierce's disease (PD), almond leaf scorch, and citrus variegated chlorosis. By identifying and comparing the sequences of *Zot*, which has been implicated as a disease causing gene, we are able to determine the relationship of the different *Xf* subspecies, and may also be able to identify different populations of *Xf*, such as those from California versus those from Texas. This will allow researchers to track the spread of PD, among others, and may be useful in detailing the mechanisms by which *Xf* causes disease.

INTRODUCTION
Xylella fastidiosa (Xf) is a Gram-negative, xylem-limited, fastidious, insect-transmitted, gammaproteobacteria (Wells et al. 1987). Five subspecies of *Xf* exist, including *Xf fastidiosa* which causes Pierce's disease (PD), *Xf sandyi* which cases oleander leaf scorch (OLS), *Xf multiplex* which causes almond leaf scorch (ALS), *Xf pauca* which causes citrus variegated chlorosis (CVC), and *Xf tashke* (Purcell 1997, Schaad et al. 2004, da Silva et al. 2007, Randall et al. 2009). *Xf* has distinctly different host ranges; though some strains of *Xf* are only pathogenic in a single host species, others cause disease in a variety of hosts (Hopkins and Purcell 2002, Almeida et al. 2003). As much as 30% of the *Xf* genome is prophage in origin(Simpson et al. 2000, Van Sluys et al. 2003, Monteiro-Vitorello et al. 2005). Other research has shown that most of the sequence variation in *Xf* subspecies occurs in coding regions derived from bacteriophages (de Mello Varani et al. 2008).

The *Zonula occludens toxin* (*Zot*) in *Xf* strains has been suggested as a new potential virulence factor in CVC caused by *Xf 9a5c*, a member of subspecies *pauca* (da Silva et al. 2007). *Zot* genes are also found in the genomes of several other gammaproteobacteria, including *Vibrio cholera*, *Xanthomonas campestris*, *Stenotrophomonas maltophilia* and *Ralstonia solanacearum* (Koonin 1992, Johnson 1993, Chang et al. 1998, Hagemann et al. 2006). The *Zot* gene found in *V. cholerae* has great sequence similarity to the protein product I (pI) of the filamentous phage Pf1, and is most likely derived from a Pf1-like phage (Koonin 1992). The pI protein, which shares similarity to many *Zot* proteins in *Xf*, has both an extracellular and intracellular region, and is necessary for phage packing and transport across the cell membrane in many filamentous phages (Koonin 1992, Di Pierro et al. 2001, Schmidt et al. 2007). A homologous protein of the *Zot* family is found in many *Vibrio cholerae* strains and has been linked to disruption of tight junctions (Johnson 1993), and diarrheagenicity in *V. cholerae* that lack the cholerae toxin (Di Pierro et al. 2001). A *Zot* gene can also be found in strains of pathogens that, like *Xf*, are found in the Xanthomonadaceae family, namely *Xanthomonas campestris*, which causes lesions and loss of water in plant tissue (Block et al. 2005), and a *Stenotrophomonas maltophilia* strain, which can cause severe health problems such as endocarditis and bacteremia (Hagemann et al. 2006). A search of available *Xf* genomes in NCBI reveals that each *Xf* strain possesses multiple copies of *Zot* genes (Schreiber et al. 2010). Three distinct subgroups exist amongst these *Zot* genes. Most abundant are the members of the *Zot1* subgroup, which are found in PD strains *Temecula1*, *M23*, *GB 514*, and *Ann-1* (Schreiber et al. 2010).

OBJECTIVES
1. Sequence the *Zot1* gene in Texas strains of *Xylella fastidiosa*
2. Translate the *Zot1* nucleotide sequence *in silico* to identify amino acid changes
3. Identify changes in predicted protein structure resulting from changes in amino acid sequence
4. Analyze the phylogeny of the *Zot1* gene in Texas strains of *Xylella fastidiosa*

RESULTS AND DISCUSSION

Subspecies identification was performed using *gyrB* and *mopB* (Morano et al. 2008) (**Table 1**). Direct sequencing yielded 41 *Zot1* sequences useful for phylogenetic analysis. Quality trimming yieldegd 861bp sequences that shared 96.0% sequence identity. The sequences were fairly divergent, with XX synonymous substitutions and XX nonsynonymous substitutions. The Texas strains identified as subspecies *fastidiosa* differed from California *fastidiosa* strains in six fixed, synonymous substitutions. These substitutions were found in the middle region of *Zot1*, and may prove to be useful in identifying populations of subspecies *fastidiosa* in the future. Subspecies *fastidiosa* sequences shared 98.9% identity, subspecies *sandyi* sequences shared 99.7% identity, and subspecies *multiplex* sequences shared 96.1% identity. The increased divergence between subspecies *multiplex Zot1* sequences may be a result of the larger number of hosts that the *multiplex* strains were collected from.

Table 1. Sample ID, Collection Site, County of Collection, Host Plant, Scientific Name of Host Plant, and Subspecies ID.

Sample ID	Collection ID	County	Host Plant	Scientific Name	Subspecies ID
A	MCC CER 040	McCulloch	Vigonier grape	*Vitis vinifera*	*fastidiosa*
B	VAL VAL 041	Val Verde	Herbemont grape	*Vitis sp. cross*	*fastidiosa*
C	LLA FAL 747	Llano	Chardonnay grape	*Vitis vinifera*	*fastidiosa*
D	XFJK 13.87	Erath	Glassy-winged sharpshooter	*Homalodisca vitripennis*	*fastidiosa*
E	LLA FAL 634	Llano	SO4 Rootstock for grape	*Vitis sp. cross*	*fastidiosa*
F	XFJK 12.57	Erath	Cabernet Sauvignon grape	*Vitis vinifera*	*fastidiosa*
G	XFJK 12.69	Erath	Zinfandel grape	*Vitis vinifera*	*fastidiosa*
H	XFJK 14.11	Erath	Ruby Cabernet grape	*Vitis vinifera*	*fastidiosa*
I	GIL GRA 315	Gillespie	Wine grape	*Vitis vinifera*	*fastidiosa*
J	2018 GIL 007	Gillespie	innoc. Chardonnay, reisolated	*Plantanus sp. (Vitis sp.)*	*fastidiosa*
K	BAN POL 055	Bandera	Black Spanish grape	*Vitis sp. cross*	*fastidiosa*
L	HEN GRA 038	Henderson	Blanc du Bois grape	*Vitis sp. cross*	*fastidiosa*
N	LLA FAL 738	Llano	SO4 Rootstock for grape	*Vitis sp. cross*	*fastidiosa*
O	LLA FAL 745	Llano	SO4 Rootstock for grape	*Vitis sp. cross*	*fastidiosa*
1	GIL BEC 514	Gillespie	Wine grape	*Vitis vinifera*	*fastidiosa*
2	GIL BEC 519	Gillespie	Wine grape	*Vitis vinifera*	*fastidiosa*
3	GIL BEC 528	Gillespie	Wine grape	*Vitis vinifera*	*fastidiosa*
4	GIL GRA 316	Gillespie	Wine grape	*Vitis vinifera*	*fastidiosa*
5	MCC CER 011-1	McCulloch	Cabernet Sauvignon grape	*Vitis vinifera*	*fastidiosa*
7	TRA FLA 338	Travis	Muscat Blanc grape	*Vitis vinifera*	*fastidiosa*
8	TRA FLA 380	Travis	Tinta Madiera	*Vitis vinifera*	*fastidiosa*
9	VAL VAL 033	Val Verde	Black Spanish grape	*Vitis sp. cross*	*fastidiosa*
10	XFJK 21.4	Erath	Ruby Seedless grape	*Vitis vinifera*	*fastidiosa*
11	MED PRI 023	Medina	Oleander	*Nerium oleander*	*sandyi*
12	MED PRI0 45-1	Medina	Oleander	*Nerium oleander*	*sandyi*
13	MED PRI 047	Medina	Oleander	*Nerium oleander*	*sandyi*
14	MED PRI 049-2	Medina	Oleander	*Nerium oleander*	*sandyi*
15	MED PRI 054	Medina	Oleander	*Nerium oleander*	*sandyi*
18	BAN POL 039	Bandera	Golden Rod	*Solidago sp.*	*multiplex*
20	GIL BEC 626B	Gillespie	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*
21	GIL BEC 627	Gillespie	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*
22	GIL BEC 628A	Gillespie	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*
24	GIL GRA 281	Gillespie	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*
27	KIM 001	Kimble	Redbud	*Cercis canadensis*	*multiplex*
28	KIM 004	Kimble	Redbud	*Cercis canadensis*	*multiplex*
30	LLA FAL 651	Llano	Heart leaf Peppervine	*Ampelopsis cordata*	*multiplex*
31	LLA FAL 718A	Llano	Narrow leaf Sumpweed	*Iva texensis*	*multiplex*
33	LLA FAL 752	Lamar	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*
34	MCC CER 044	McCulloch	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*
36	UVA 122A	Uvalde	Sycamore	*Plantanus sp.*	*multiplex*
37	UVA 521-2B	Uvalde	Red Bud	*Cercis sp.*	*multiplex*
38	UVA TAM 115	Uvalde	Western Soapberry	*Sapindus saponaria L. var. drummondii*	*multiplex*
39	VAL VAL 072A	Val Verde	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*

The *Zot1* nucleotide sequences were then translated *in silico* to identify amino acid differences. This produced sequences of 281 amino acids long that shared 96.4% identity. These sequences were submitted to Pfam for identification (Finn et al. 2008). Each sequence was noted as belonging to the Zot protein family with an error value greater than $5e^{-10}$. Tertiary structure homology was predicted using the PHYRE program (Kelly and Sternberg 2009). The PHYRE search identified the N-Terminus of the Zot protein of *Neisseria meningitidis* (PDB code 2R2A) (data not shown) as the most similar structure to the sequences submitted, with a minimum error value of $2e^{-10}$. This, too, points to a conserved structure in the Zot1 protein in *Xf*. The Zot1 protein sequences were then subjected to secondary-structure prediction, transmembrane region prediction, and signal sequence cleavage site prediction (**Figure 1**) (Drummond et al. 2009). The predicted secondary structures, signal sequence cleavage sites, and transmembrane regions were very similar, indicating a degree of conservation amongst the structure of Zot1 proteins, despite the divergent nucleotide sequences. Particularly promising, the signal cleavage site predictions indicate two areas where cleavage may occur. The Zot protein is known to have a cleaved, exotoxic C-Terminus in *Vibrio cholerae* (Di Pierro et al. 2001). The predicted cleavage sites closest to the C-Terminus in the Zot1sequences from *Xf* correspond to the expected cleavage site of the Zot protein in other bacteria. Additionally, these regions of the Zot1 protein exhibits 98.7% sequence identity, higher than the average for the entire sequence. Further research to determine the exact cleavage site of the Zot1 protein.

Figure 1. Protein structure prediction produced using EMBOSS tool plug-in for Geneiousv. 5.1. A representative sequence from each phylogenetic cluster is represented to display structure homology across groups. Cellular region is indicated by pink arrows, transmembrane regions are indicated by red arrows, signal cleavage sites are indicated by green arrows. Secondary structure prediction is above the amino acid sequence. Blue arrows indicate turns, yellow arrows indicate alpha helices, and purple cylinders indicate beta sheets.

Phylogenetic analysis yielded a tree topology that was predicted by the *gyrB* and *mopB* identification (**Figure 2**). The Texas samples of this study were supplemented with sequences retrieved from NCBI's GenBank, including two sequences from both *Xf* M23 and Temecula1, one sequence from *Xf* Ann-1, and three *Zot1* sequences previously sequenced in house. Though recombination blurred distinctions between groups, sequences separated into complexes according to their subspecies. These complexes are highly supported, as evidenced by their high bootstrap values. Additionally, the California *fastidiosa* strains (*Xf M23 and Temecula1*) separated from the Texas *fastidiosa* strains (*Xf Texas 1* and Samples 10, K, N, and O). The group composition of the complexes was not reflective of geographic proximity or host plant identity (data not shown). The high degree of similarity between sequences despite geographic range, and the high levels recombination indicate a large amount of mixing between different strains of *Xf*.

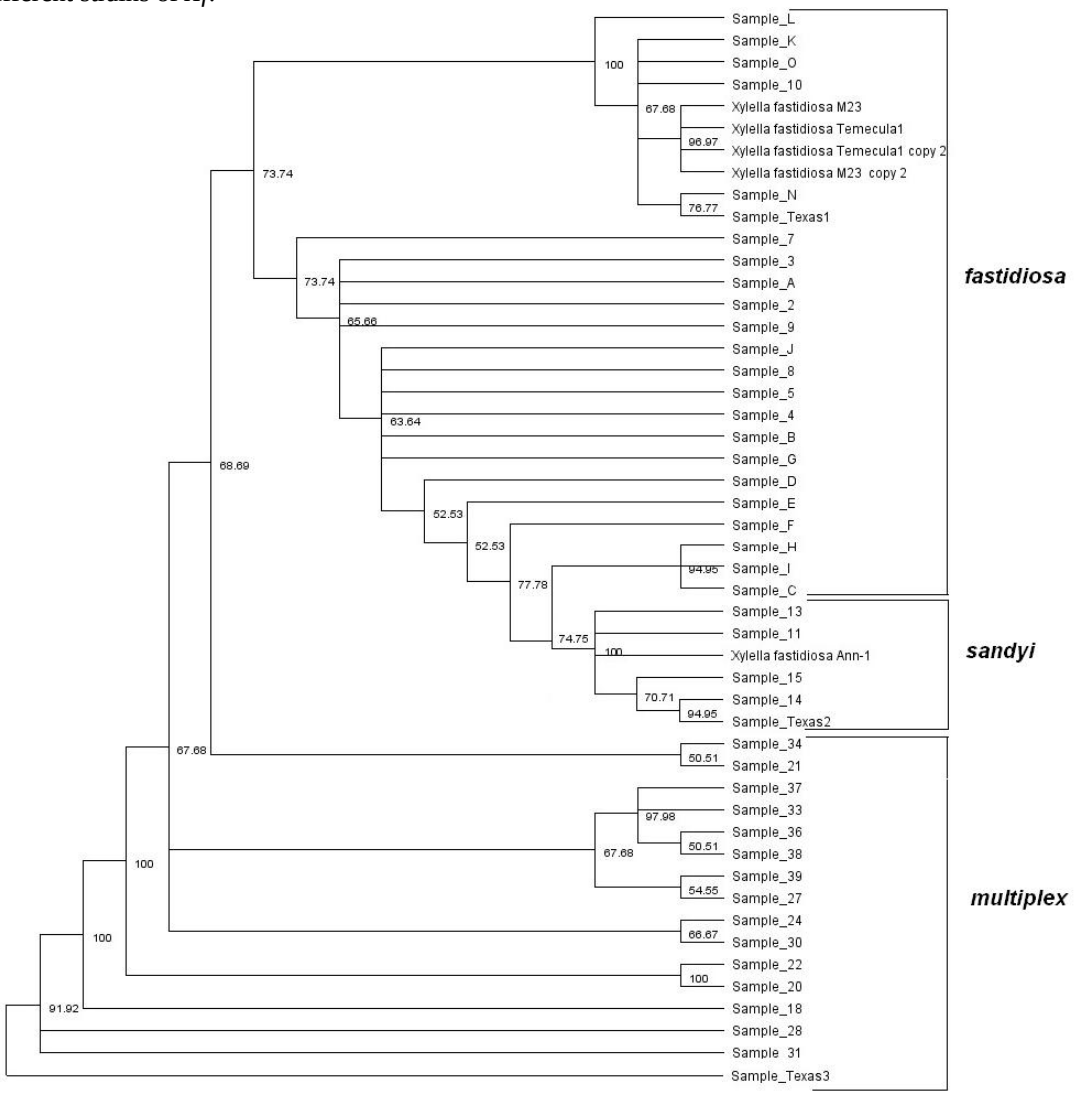

Figure 2. Phylogenetic tree produced using the Maximum-Likelyhood method and the HKY evolutionary model with a ratio of 0.774 invariable sites, four categories of substitution, and a gamma shape of 0.931.

Little is known about the *Zot* protein's structure due to the complexity of isolating *Zot* for analysis as a result of its transmembrane region (Koonin 1992, Hagemann et al. 2006, Schmidt et al. 2007). *In silico* translations and structure predictions offer great insights into protein function and classification and has been found to be accurate and sensitive. By using *in silico* analysis, the sequence variation found in *Zot1* genes between subspecies has been shown to yield differences in protein structure. Predictive modeling shows that the variation found in the amino acid code does not translate to altered protein structure. This similarity in structure, despite changes in both the nucleotide and amino acid code, indicates that the *Zot1* gene may not be evolving due to host-pathogen interactions.

The first step in determining host range in differentially pathogenic bacteria is placing the bacteria into clades (Morano et al. 2008). Many techniques for identification and classification exist; however, the complexity of *Xf* pathogen makes categorization based on morphology or pathogenicity difficult (Almeida et al. 2003). Additionally, *Xf* has been shown to

have limited genomic variability within clades and region, and that the majority of strain specific genes occur in prophage regions, though they contain genomic islands which enable rapid evolution (Van Sluys et al. 2003). Techniques that focus on classification based on well conserved regions shared by all *Xf* subspecies might then miss putative evolutionary growth and adaptation. The *Zot1* gene insertion is an excellent target for QRT-PCR, and the small sequence differences can be targeted by restriction enzyme digestion analysis for quick and accurate identification and classification of *Xf* subspecies and populations.

CONCLUSIONS

Comparative analysis *Zot1*genes and proteins provide accurate, population level differentiation therefore allows researchers greater ability to track the spread of economically important phytopathogens. Additionally, *in silico* translation and analysis of *Zot1* describes in greater detail differences between strains, and describes possible conformation changes that result from sequence changes between strains. Taken together, these results show that *Zot1* is a useful target for differential sequence analysis and can be used to elucidate the phylogenetic history of *Xf*, and its spread through the U.S.

REFERENCES CITED

Almeida, R. P. P., and A.H. Purcell. 2003. Biological traits of *Xylella fastidiosa* strains from grapes and almonds. Appl. Environ. Microb. 69:7447–7452.

Block, A., E. Schmelz, P.J. O'Donnell, J. B. Jones, and H. J. Klee. 2005. Systemic acquired tolerance to virulent bacterial pathogens in tomato. Plant. Physiol. 138:1481–1490.

Chang, K.H., F.S. Wen, T.T. Tseng, N.-T. Lina, M.-T. Yanga, and Y.-H. Tseng. 1998. Sequence analysis and expression of the filamentous phage φLf gene I encoding a 48-kDa protein associated with host cell membrane. Biochem .Biophys. Res. Commun. 245:313–318.

Fasano, A., B. Baudry, D. W. Pumplin, S. S. Wasserman, B. D. Tall, J. M. Ketley, and J. B. Kaper. 1991. *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. Proc. Natl. Acad. Sci. USA 88:5242-5246

Di Pierro, M., R. Lu, S. Uzzau, W. Wang, K. Margaretten, C. Pazzani, F. Maimone, and A. Fasano. 2001. Zonula occludens toxin structure-function analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain. J. Biol. Chem. 276:19160-19165.

Drummond A.J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, A. Wilson. 2010. Geneious v5.1, Available from http://www.geneious.com

Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. 2008. The Pfam protein families database. Nucleic Acids Res. 36:D281-D288.

Hagemann, M., D. Hasse, and G. Berg. 2006. Detection of a phage genome carrying a zonula occludens like toxin gene (*zot*) in clinical isolates of *Stenotrophomonas maltophilia*. Arch. Microbiol. 185: 449-458.

Hayward, A. C. 1991. Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. Annu. Rev. Phytopathol. 29:65-87

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

Johnson, J. A., J. G. Morris, Jr., and J. B. Kaper. 1993. Gene encoding zonula occludens toxin (*zot*) does not occur independently from cholera enterotoxin genes (ctx) in *Vibrio cholerae*. J. Clin. Microbiol. 31:732–733.

Kelley, L. A., and M. J. E. Sternberg. 2009. Protein structure prediction on the web: a case study using the Phyre server. Nat. Protoc. 4: 363-371

Koonin E. V. 1992. The second cholera toxin, *Zot*, and its plasmid-encoded and phage-encoded homologues constitute a group of putative ATPases with an altered purine NTP-binding motif. FEBS Lett. 312(1):3-6.

de Mello Varani, A., R. C. Souza, H. I. Nakaya, W. C. de Lima, L. G. P. de Almeida, E. W. Kitajima, J. Chen, E. Civerolo, A. T. R. Vasconcelos, and M. V. Sluys. 2008. Origins of the *Xylella fastidiosa* prophage-like regions and their impact in genome differentiation. PLoS One 3(12): e4059. Epub 2008 Dec 31.

Monteiro-Vitorello, C. B., M. C. De Oliveira, M. M. Zerillo, A. M. Varani, E. Civerolo, and M. A. Van Sluys. 2005. *Xylella* and *Xanthomonas* mobil'omics. OMICS. 9:146-159.

Morano, L. D., B. R. Bextine, D. A. Garcia, S. V. Maddox, S. Gunawan, N. J. Vitovsky and M. C. Black. 2008. Initial genetic analysis of *Xylella fastidiosa* in Texas. Curr. Microbi

Purcell, A. H. 1997. *Xylella fastidiosa*, a regional problem or global threat? J. Plant Pathol. 79:99-105.

Randall, J.J., N.P. Goldberg, J.D. Kemp, M. Radionenko, J.M. French, M.W. Olsen, S.F. Hanson. 2009. Genetic Analysis of a Novel *Xylella fastidiosa* Subspecies Found in the Southwestern United States. Appl. Envir. Microbiol. 75: 5631-5638

Schaad, N. W., E. Postnikova, G. Lacy, M. Fatmi, and C. J. Chang. 2004. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. *piercei*, subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. Syst. Appl. Microbiol. 27:290-300. (Erratum 27:763).

Schmidt, E., S. M. Kelly, and C. F. van der Walle. 2007. Tight junction modulation and biochemical characterisation of the zonula occludens toxin C-and N-termini. FEBS Lett. 581(16):2974–2980.

da Silva, V. S., C. S. Shida, F. B. Rodrigues, D. C. D. Ribeiro, A. A. de Souza, H. D. Coletta-Filho, M. A. Machado, L. R. Nunes, and R. C. de Oliveira. 2007. Comparative genomic characterization of citrus-associated *Xylella fastidiosa* strains. BMC Genomics 8:474-489.

Simpson, A. J. G., F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, et al. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature 406:151-157.

Van Sluys, M. A., M. C. de Oliveira, C. B. Monteiro-Vitorello, C. Y. Miyaki, L. R. Furlan, et al. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J. Bacteriol. 185:1018-1026.

Wells, J. M., B. C. Raju, H. Hung, W. G. Weisburg, L. Mandelco-Paul, and D. J. Brenner. 1987. *Xylella fastidiosa* gen. nov., sp. nov.: gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. Int. J. Syst. Bacteriol. 37:136-143.

FUNDING AGENCIES

Funding for this project was provided in part by the USDA Animal and Plant Health Inspection Service, and the Texas Pierce's Disease Research and Education Program.

AKNOWLEDGEMENTS

This material was made possible, in part, by a Cooperative Agreement from the United States Department of Agriculture's Animal and Plant Health Inspection Service (APHIS). It may not necessarily express APHIS' views.