THE RECOMBINATION RATE OF THE *ZOT* AND *GYRASE B* GENES 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
Xylella fastidiosa (Xf) is a Gram-negative gamma proteobacteria that is responsible for several economically important plant diseases. The *Zonula occludens toxin* (*Zot*) is an exotoxin produced and secreted by *Xf* that has been suggested as a potential virulence factor in other research. This report is a description of the recombination rates, nucleotide diversity, and rates of linkage disequilibrium of both the *Zot* gene and the housekeeping gene *gyraseB* (*gyrB*). The *Zot* gene has a much higher degree of nucleotide diversity, recombination rate, and less intragene linkage disequilibrium. This indicates that the *Zot* gene is undergoing more selection pressure than the *gyrB* gene. Additionally, this report suggests that *Xf* has higher than reported rates of recombination, but that this recombination is masked by similar sequence identity.

LAYPERSON SUMMARY
In this study, we examined the rates of recombination in two genes in the *Xylella fastidiosa (Xf)* genome. Recombination occurs when strands of DNA interact, and sometimes switch. These instances occur in bacteria when bacteriophages, viruses that infect bacteria, insert their genetic material into a bacterial cell, when bacteria undergo a form of mating, called conjugation, or when bacteria uptake foreign DNA from outside of their cell. Recombination plays an important role in evolution, by rearranging chromosomes, inserting new genetic sequences, or exchanging bits of genes from one strand of DNA to another. It is an important source of mutation in bacteria, and little work has been done to study recombination in *Xf*.

INTRODUCTION
Xylella fastidiosa (Xf) is a Gram-negative, xylem-limited, fastidious, insect-transmitted, gamma proteobacteria (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 , *Xf multiplex* which causes almond leaf scorch, *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). 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). High rates of chromosomal rearrangements, recombinations, and gene loss has been detected in the prophage regions of *Xf* (Monteiro-Vitorello 2005). The *Zonula occludens toxin* (*Zot*) 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 many other pathogenic bacteria, including *Vibrio cholera*, *Xanthomonas campestris, Stenotrophomonas maltophilia*and *Ralstonia solanacearum* (Koonin 1992, Johnson 1993, Chang et al. 1998, Hagemann et al. 2006). The *Zot* genes in *Xf* appear in prophage regions of the genome (Monteiro-Vitorello et al. 2005, de Mello Varani et al. 2008). Several of these *Zot*genes share sequence homology with the *Zot*gene found in *Vibrio cholera*, which is derived from the pI protein of a bacteriophage (Johnson 1993). The prophagic pI protein is integral to proper virus packaging and export (Change et al. 1998). 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).

Recombination events can affect bacterial evolution (Maynard Smith et al. 1994), but little work on the recombination of prophage regions of *Xf* has been done. Recombination rates have been shown to affect clonal complex composition and influence the phylogenetic structure of *Xf*(Scally et al. 2005). However, short divergence times, and a low rate of mutation has led to a high degree of clonality amongst *Xf* strains (Schuenzel et al. 2005). This high degree of similarity means reduces the chances of accurately identifying recombination rates, as recombination events between identical sequences are undectable via sequence analysis (Posada et al. 2002). As such, only a small fraction of recombination events are accurately identified in sequences with high degrees of similarity (>99%), resulting in underestimates of recombination rates (Hudson and Kaplan 1985).

This study is a presentation of materials describing the differences in recombination between a housekeeping gene, *gyrB*, and a prophage gene with significant sequence divergence, *Zot1*. The *gyrB*gene was chosen because of its use in phylogenetic analysis and its conserved nature (Morano et al. 2008), while the *Zot1* gene was chosen because it is most prevalent amongst the PD strains of *Xf*, is significantly divergent, and is prophage in origin.

OBJECTIVES

1. Sequence the *Zot1*and *gyrB*genes in Texas strains of *Xf*.
2. Identify areas of recombination using visual inspection methods as well as *in silico* analysis.
3. Compare rates of recombination between a prophage gene, *Zot1* and a housekeeping gene, *gyrB.*

RESULTS AND DISCUSSION

Subspecies identification was performed using *gyrB* and *mopB* (Morano et al.2008) (**Table 1**). Quality trimming of the *Zot1* sequences yielded 861bp sequences that shared 96.0% sequence identity. The sequences were fairly divergent, with 69 synonymous substitutions and 35nonsynonymous substitutions. Quality trimming of the *gyrB* sequences yielded 631bp sequences that shared 99.0% sequence identity. The *gyrB* sequences contained 15 synonymous substitutions and 3nonsynonymous substitutions. This indicates that the *Zot1* gene is more divergent than the *gyrB* gene.

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	*Vitisvinifera*	*fastidiosa*
B	VAL VAL 041	Val Verde	Herbemont grape	*Vitis sp. cross*	*fastidiosa*
C	LLA FAL 747	Llano	Chardonnay grape	*Vitisvinifera*	*fastidiosa*
D	XFJK 13.87	Erath	Glassy-winged sharpshooter	*Homalodiscavitripennis*	*fastidiosa*
E	LLA FAL 634	Llano	SO4 Rootstock for grape	*Vitis sp. cross*	*fastidiosa*
F	XFJK 12.57	Erath	Cabernet Sauvignon grape	*Vitisvinifera*	*fastidiosa*
G	XFJK 12.69	Erath	Zinfandel grape	*Vitisvinifera*	*fastidiosa*
H	XFJK 14.11	Erath	Ruby Cabernet grape	*Vitisvinifera*	*fastidiosa*
I	GIL GRA 315	Gillespie	Wine grape	*Vitisvinifera*	*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	*Vitisvinifera*	*fastidiosa*
2	GIL BEC 519	Gillespie	Wine grape	*Vitisvinifera*	*fastidiosa*
3	GIL BEC 528	Gillespie	Wine grape	*Vitisvinifera*	*fastidiosa*
4	GIL GRA 316	Gillespie	Wine grape	*Vitisvinifera*	*fastidiosa*
5	MCC CER 011-1	McCulloch	Cabernet Sauvignon grape	*Vitisvinifera*	*fastidiosa*
7	TRA FLA 338	Travis	Muscat Blanc grape	*Vitisvinifera*	*fastidiosa*
8	TRA FLA 380	Travis	TintaMadiera	*Vitisvinifera*	*fastidiosa*
9	VAL VAL 033	Val Verde	Black Spanish grape	*Vitis sp. cross*	*fastidiosa*
10	XFJK 21.4	Erath	Ruby Seedless grape	*Vitisvinifera*	*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	*Cerciscanadensis*	*multiplex*
28	KIM 004	Kimble	Redbud	*Cerciscanadensis*	*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	*Sapindussaponaria L. var. drummondii*	*multiplex*
39	VAL VAL 072A	Val Verde	Giant Ragweed	*Ambrosia trifida L. var. Texana Scheele*	*multiplex*

The sequenced genes were aligned using Geneious v. 5.1. Analysis of the sequences was performed using DnaSP to identify sequence characteristics (Rozas et al. 2003).Nucleotide Diversity, per nucleotide (π) and average number of nucleotide changes per gene (θ) were calculated to identify the differences in nucleotide variability between the *Zot1* and *gyrB* genes (**Table 2**). These numbers show that the *Zot1* gene has a much higher rate of variability than the *gyrB* gene. The rate of recombination per gene, R, was calculated using the minimum number of recombinations statistic (Rm) following the protocol of Hudson and Kaplan (1987), (**Table 2**). Although the minimum number of recombinations was much higher for the *Zot1* gene, the *gyrB*gene displayed a higher rate of recombination overall. Finally, the rate of linkage disequilibrium of each gene was calculated, as an indirect measure of recombination (**Table 3**). The ZZ statistic of Rozas et al. (2001) was used in place of the Z_{nS} proposed by Kelly (1999) for greater accuracy in determining the rates of recombination. The higher

ZZ value of the *Zot1* gene indicates a greater amount of recombination is apparent in the *Zot1* gene than in the *gyrB*genes. The much larger number of informative sites, corrected using the Bonferroni calculations, for the *Zot1* gene supports the conclusion derived from the ZZ statistic.

Table 2. Results of sequence analysis using DnaSP v. 5.1.

Sequence Characteristics	**Test**	***Zot gene***	***gyrB gene***
Nucleotide Diversity	**Nucleotide Diversity, per nucleotide (π)**	**0.0355**	**0.0112**
	Average number of nucleotide changes per gene (θ)	**29.31**	**2.394**
Recombination	**Minimum number of recombinations, Rm**	**20**	**0**
	Estimate of recombination per gene, R	**4.1**	**6.3**
Linkage Disequilibrium	**Fisher's Exact Test**	**1444**	**12**
	with Bonferroni correction	**129**	**11**
	Chi squared Test	**2408**	**12**
	with Bonferroni correction	**545**	**11**
	ZZ Value	**0.2371**	**0.2004**

Coalescent simulations were performed using DnaSP v. 5.1 based on the θ statistic and the observed rate of recombination for each gene to predict the Rm statistic and ZZ statistic in a hypothetical population (Rozas et al 2003). Simulations were run 1000 times in order to obtain a predicted average, and a 95% confidence interval. This average was then compared to the observed value to identify the probability that the observed value lies outside the predicted bell curve. The coalescent simulations show that the observed values of the *Zot1* gene lie outside the predicted bell curve for both the Rm and ZZ statistic. This indicates either relaxed negative selection or positive selection pressures are working on the *Zot1* gene to increase genetic diversity by overcoming negative selection sweeps that are common to gene recombination.

Table 3. Coalescent simulations performed using DnaSP v. 5.1. Simulations were run using observed values of Rm and the θ statistic.

Rm	**Observed Value**	**Simulated Average**	**95% confidence interval**	**p-value of observed Rm statistic**
Zot gene	20.00	3.49	1.00 to 7.00	0.000***
gyrB gene	2.00	2.20	0.00 to 3.00	0.681(ns)

ZZ Statistic	**Observed Value**	**Simulated Average**	**95% confidence interval**	**p-value of observed ZZ statistic**
Zot gene	0.237	0.053	-0.015 to 0.159	0.001 **
gyrB gene	0.200	0.041	-0.077 to 0.213	0.07(ns)

CONCLUSIONS

The results of the experiments performed in this project suggest that the *Zot1* gene is evolving rapidly and is prone to recombination events. As Zot proteins have been identified as potential virulence factors, this phenomenon deserves greater scrutiny. Previous reports have identified relatively low rates of recombination in *Xf*. The high sequence similarity between strains of *Xf*, as much as 98% between subspecies, may be masking high rates of recombination that leave no genetic trace when they occur between highly similar strains. Further research into divergent regions of the *Xf* genome to determine actual rates of recombination is warranted, given the rate of cohabitation common to *Xf* strains.

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

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 zonulaoccludens toxin (*zot*) does not occur independently from cholera enterotoxin genes (ctx) in *Vibrio cholerae*. J. Clin. Microbiol.31:732–733.

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 zonulaoccludens 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 *Xylellaf astidiosa* 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 *Xylellafastidiosa*. 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.