PHYLOGENETIC ANALYSIS OF HEAT SHOCK PROTEINS IN THE GLASSY-WINGED SHARPSHOOTER

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

Co-Principal Investigator:
Wayne Hunter
USDA-ARS
Fort Pierce, FL 34945
wayne.hunter@ars.usda.gov

Researchers:
Henry L. Schreiber IV
Department of Biology
University of Texas-Tyler
Tyler, TX 75799
henrylschreiber@gmail.com

Daymon Hail
Department of Biology
University of Texas-Tyler
Tyler, TX 75799
daymon.hail@gmail.com

Reporting Period: The results reported here are from work conducted January 2009 to December 2009.

ABSTRACT
The glassy-winged sharpshooter (GWSS; *Homalodisca vitripennis* (Germar); Hemiptera: Cicadellidae) is the major vector of *Xylella fastidiosa (Xf)*, the causal agent of Pierce's disease (PD) of grapes. As genomic information becomes available more research on leafhopper stress responses are possible. Due to the importance of the GWSS in transmission and spread of *Xf*, a cDNA library was constructed from adult and fifth-instars GWSS, resulting in 5,906 expressed sequence tags (ESTs). After quality scoring, 4,445 sequences underwent assembly which produced a set of 2,123 sequences that putatively represented distinct transcripts. BLASTX analysis identified four significant homology matches to heat shock proteins, (HSP) which are the focus of this study. The overall importance and function of HSPs lie in their ability to maintain protein integrity and activity during stressful conditions, such as extreme heat, cold, drought, or other stresses. Phylogenetic analyses using these four HSP sequences provided further support of transcript by the identification of specific motifs. This study shows that highly conserved genes like HSPs are a viable alternative to ribosomal DNA in elucidating phylogenetic relationships.

LAYPERSON SUMMARY
In this study, we generated and analyzed a cDNA library, which is a representation of all the protein coding genes in an organism, of the insect pest glassy-winged sharpshooter (GWSS). We isolated four incomplete sequences from a large family of proteins called heat shock proteins (HSP). Although they are called HSP, these proteins actually perform many activities in the cell that allow for GWSS to survive stresses like extreme temperatures, pesticides, and even viral infection. This study compares the HSP sequences from GWSS to those of other insect species to help describe the relationship and history of GWSS to find ways to track changes in GWSS territory and life history.

INTRODUCTION
Organisms respond to heat shock or other environmental stress by inducing the synthesis of proteins some of which are known as heat shock proteins (HSP) (Lindquist 1986, Sorenson 2003). Infections, temperature changes, inflammation, toxins, hypoxia, starvation and even exercise can result in increased production of heat shock proteins (Sorenson 2003). HSP aid in folding, targeting and tracking of nascent proteins, promote transcription, are involved in cellular division and can be up regulated via cell signaling in addition to environmental stimuli (Feder and Hofmann 1999).

Small heat shock proteins (sHSP) have an approximate molecular weight of less than 30kDa and are molecular chaperones, maintaining proper protein structure by blocking aggregation of denaturing proteins, aiding nascent protein folding and assisting construction of quaternary structure (Fu and Chang 2004, Gu et al. 2002, Bova et al. 2002, Sobott 2002).

Among the HSP families is a group of well-conserved proteins with an approximate molecular weight of 70 kDa, known as the HSP70 family. Most species have several proteins belonging to this family. Some of these members are only expressed under stress conditions (strictly inducible), while some are present in cells under normal growth conditions (Craig 1989) and are not heat-inducible (Pelham 1986), and are known as heat shock cognates (HSC). In eukaryotes, HSP70 can work with sHSP to restore functionality to heat-denatured proteins (Lee and Vierling 2000) or co-chaperone with HSP40 to fold nascent proteins into proper tertiary structure by temporarily binding to hydrophobic domains until sequence translation is complete (Douglas et al. 1994).

The 90 kDa heat shock proteins (HSP90) is one of the most prolific proteins in eukaryotic cells, constituting 12% of cellular proteins under baseline conditions (Sreedhar 2004). Their functions and morphological evolution have been extensively studied and include signal transduction, protein folding and degradation of denatured proteins (Nadeau 1993, Jakob 1994). Increased functionality of HSP90 is acquired when associated with its co-chaperones, playing an important role in the folding of newly synthesized proteins. Apart from its co-chaperones, HSP90 binds to an array of substrate proteins, where the necessary co-chaperones varies and depends on the actual substrate (Jakob 1995). Understanding heat shock proteins in insects, especially leafhoppers, will provide insights into the biological adaptive elasticity of these important agricultural pests to stressors such as insecticides, parasitization, and temperature.

The glassy-winged sharpshooter (GWSS; *Homalodisca vitripennis* (Germar); Hemiptera: Cicadellidae), is an insect pest that occurs throughout most of the southern USA and is endemic to most regions of Texas (Young 1958; Turner and Pollard 1959). Without naturally occurring forms of biological control, GWSS have established populations in new areas and have

negatively affected the yields of the grape industry (de Leon et al. 2004). GWSS is a voraciously-feeding, xylem-limited pest that has been reported to feed on host plants from at least 35 families, including both woody and herbaceous types (Hoddle et al. 2003), and can impact the plant's health directly by depriving the plant of nutrients and damaging the xylem sufficiently to preclude vascular flow. However, indirect plant damage occurs during feeding and subsequent transmission of the xylem-limited bacterium *Xylella fastidiosa (Xf)* Wells (Xanthomonadales: Xanthomonadaceae). The invasion of GWSS into grape growing regions of California, namely the Temecula valley, produced an enormous risk to the California wine and table grape industry by spreading the phytopathogen *Xf,* the causative agent of Pierce's disease (PD) (Purcell 1997, de Leon et al. 2004). Additionally, many other economically important plants including citrus, almond and oleander are affected by separate strains of *Xf* resulting in a multitude of plant diseases including citrus variegated chlorosis (Chang et al. 1993; Pooler and Hartung 1995), almond leaf scorch (Mircetich et al. 1976) and oleander leaf scorch (Purcell 1999).

A search of the National Center for Biotechnology Information (NCBI) for GWSS genes or protein sequences revealed less than 25 complete, non-mitochondrial genes or complete proteins. Although the complete mitochondrial sequence of GWSS has been described (Genbank AY875213), the genomic DNA sequence is incomplete. Over 20,000 Expressed Sequence Tags (ESTs) from GWSS have been submitted to NCBI; however, many of these EST's are duplicates and this study is an initial step in examining the potential use of this information by examining the utility of these heat shock proteins to describe the phylogeny of leafhoppers in relation to other insects. Further management approaches have been proposed to disrupt HSP in insects as a means to suppress leafhopper populations.

OBJECTIVES

1. Identify the phylogenic relationship of important leafhopper species using heat shock protein sequences.
2. Develop a methodology to distinguish between populations of GWSS.

RESULTS AND DISCUSSION

Mining of the 5,906 cDNA clones produced from cDNA library constructed from 140 adult and fifth-instar GWSS using Stratagene's ZAP-cDNA Synthesis Kit (Stratagene, La Jolla, CA, USA) resulted in 4,445 high-quality (i.e., ≥200 bases with a TraceTuner™ score of 20 or better) GWSS ESTs sequenced by ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence alignment of these ESTs resulted in a Unigene set of 2,123 total assembled sequences, at Phred 20 score, 40bp overlap, 100 bp minimum length, using Sequencher™ 8.0, (Gene Codes Corp, Ann Arbor, MI, 48108). Translated proteins were analyzed with National Center for Biotechnology Information's (NCBI) BLAST*p*, Pfam (www.pfam.org), InterProScan (www.ebi.ac.uk) and Expert Protein Analysis System (www.expasy.org). Four partial protein sequences were analyzed for phylogenetic relationships with homologous HSP sequences. A BLAST*p* and BLAST*n* analysis showed that WHWTC-contig[1627] and WHWTC-contig[1325] showed significant homology with the sHSP, while WHHC-contig[1333] displayed homology with HSP70, and WHWTC-contig[1285] was homologous with HSP90 (**Table 1**). A homology search conducted in the Pfam database identified protein sequences with their respective HSP families (**Table 2**). A functional analysis and homology search using PantherDB annotated and classified the sequences as belonging to the HSP superfamily (data not shown).

Table 1. Protein sequence similarities from GWSS contigs. Nucleotide matches (accesion|protein description organism) and e-values for query contigs using National Center for Biotechnical Information (NCBI)'s BLAST*x*.

GWSSclones	**Descriptor**	**E-Value**
WHWTC-Contig[1627] 694 bases	gb|ABC84494.1| heat shock protein 20.7 *Locusta migratoria*	3.00E-43
WHWTC-Contig[1325] 954 bases	gb|ACH85196.1|heat shock protein 20 *Bemisia tabaci*	2.00E-44
WHHC-Contig[1333] 1047 bases	gb|AAZ17399.2| 70 kDa heat shock protein *Bemisia tabaci*	2.00E-149
WHWTC-Contig[1285] 1144 bases	gb|AAZ17403.1| 90 kDa heat shock protein *Bemisia tabaci*	6.00E-179

Table 2. Hidden Markov models (HMM) homology search of *in silico*-translated protein sequences using Pfam protein database (www.pfam.org) with protein family description, Pfam identification, sequence coverage and corresponding e-value.

Contig Number	Description	Pfam Family ID	Sequence		HMM		E-value
			Start	End	From	To	
WHWTC-Contig[1627]	Hsp20/alpha crystallin family	PF00011	86	182	1	109	1.50E-40
WHWTC-Contig[1325]	Hsp20/alpha crystallin family	PF00011	63	159	1	109	2.30E-38
WHHC-Contig[1333]	Hsp70 protein	PF00012	1	325	291	619	2.20E-201
WHWTC-Contig[1285]	Hsp90 protein	PF00183	3	380	101	489	0

The heat shock proteins (HSPs) from GWSS had homology to the HSP from other insects, and grouped most closely with other Hemiptera when subjected to phylogenetic analysis. Phylogenetic trees illustrated accurate grouping of taxa into clades relative to known HSP from closely related Hemipteran species (**Figures 1-3**). The clades were separated according to taxonomic Order. Two small heat shock protein sequences (sHSP) from GWSS grouped with another sharpshooter *Graphocephala atropunctata* sHSP (**Figure 1**). The HSP70 sequence from GWSS grouped with two HSP70 sequences from the pea aphid (*Acyrthosiphon pisum*) (**Figure 2**). Finally, the HSP90 sequence from GWSS grouped with three sequences from the pea aphid (*A. pisum*) (**Figure 3**).

These phylogenetic analyses corroborate evidence from Pfam and PantherDB protein databases that describe the GWSS partial protein sequences as HSP. Additionally, the phylogenetic trees created using these protein sequence comparisons show that HSP can be used to determine phylogenetic and cladistical associations.

Heat shock proteins, HSPs have a variety of functions within the cell including the prevention of protein aggregation and denaturation due to heat and are well conserved across all taxa, and are present in every species analyzed (Feder and Hofmann 1999, Sorensen 2003). HSP families are organized via their level of expression in the cell (i.e. inducible or constitutive expression) as well as the complexes formed by the HSP; however, the greatest organization criteria is the molecular weight of the HSP which include families of 20kDa, 40kDa, 60kDa, and 90kDa proteins (Gething 1997). Finally, conserved domains exist in these families, including alpha-crystalline structure in sHSP, an N-terminal pentapeptide sequence in HSP70, and a highly conserved N-terminal domain in HSP90. The HSP sequences collected from GWSS contain these conserved domains permitting significant *in silico* comparisons (data not shown).

Phylogenetic analysis and alignment searches of the four HSP sequences were confounded by the overwhelming number of HSP sequences isoforms submitted to NCBI. However, with careful consideration paid to HSP isoforms, phylogenetic comparisons showed accurate clades of GWSS HSP with those of other closely related insect taxa (**Figures 1-3**). Phylogenetic trees formed on HSP comparison verified other phylogenetic analyses based on mitochondrial DNA. Although many ESTs in this study were found to be HSP homologues, there remain many as yet unanalyzed HSP in the GWSS genome, including members of the small Heat Shock Proteins (sHSP), HSP60, and HSP70 families. Additionally, many members of the HSP90 family and its co-chaperones have yet to be sequenced. The need for more in depth sequencing of GWSS is evident by the paucity of HSPs currently identified in the GWSS genomic database. In *Drosophila melanogaster*, whose genome is completely sequenced, over 200 HSPs have been identified and submitted to the National Center for Biotechnology Information (NCBI). GWSS has a predicted genome size of ~1.24 pg (Hunter, unpublished), similar to the haploid male Whitefly, *Bemisia argentifolii* at ~1.1pg (Leshkowitz et al. 2006), three times the size of the Asian citrus psyllid ~0.35pg (Hunter et al., 2009) and roughly five times the size of the fruitfly *D. melanogaster*, which is ~0.18pg (Brown et al. 2005). Thus, we suspect that sharpshooters will have a number of HSPs that would approximate the number in other insect genomes.

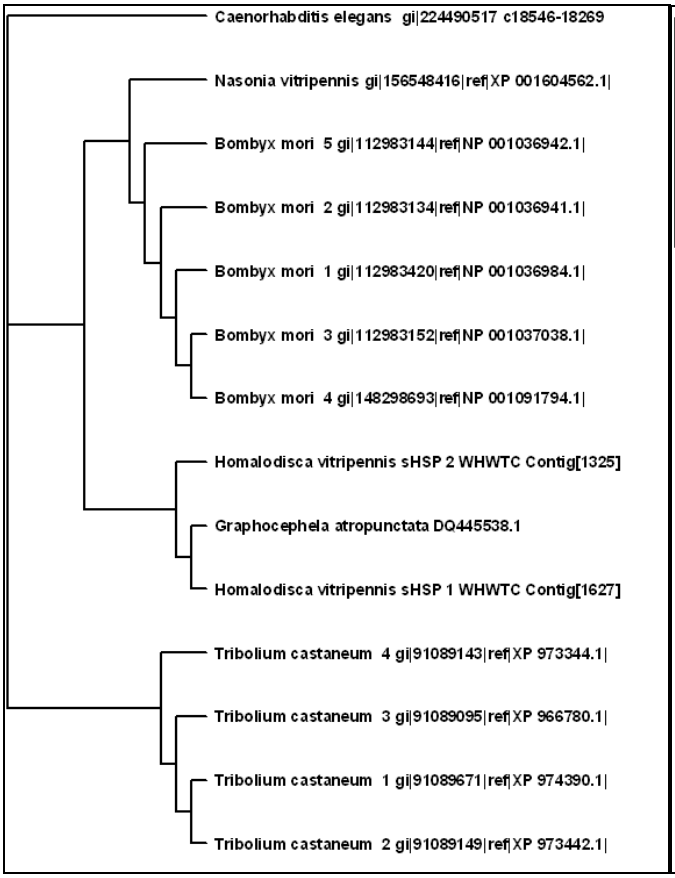

Figure 1. Cladogram of small heat shock proteins constructed using Glassy-winged sharpshooter sequences WHWTC-Contig[1627] and WHWTC-Contig[1325]. Subject sequences were analyzed using NCBI BLAST*p* search and aligned using T-Coffee multiple alignment tool (www.tcoffee.org) and visualized using Treeview v1.6.6. (Organism name|accession number| reference number).

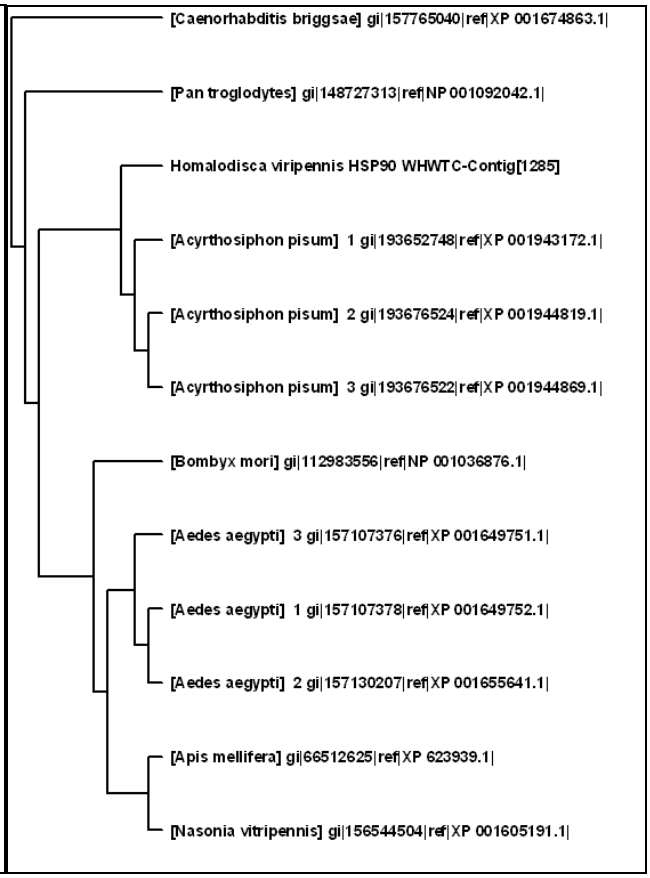

Figure 2. Cladogram of heat shock 70 proteins constructed using sequence WHHC-Contig[1333]. Subject sequences were analyzed NCBI BLAST*p* search and aligned using ClustalW2 multiple alignment tool (www.ebi.ac.uk/Tools/clustalw2/index.html) and visualized using Treeview v1.6.6. (Organism name|accession number| reference number).

Systematic biases can distort evidence via improper gene sampling. Therefore, it is necessary to limit the effects of these entanglements by analyzing multiple genes that undergo relatively uniform evolution. Ribosomal DNA is a useful molecule for examining phylogenetic relationships among many eukaryotes, primarily because no other molecule has been sequenced as extensively (Stechmann 2003). However, phylogenetic analysis using ribosomal DNA can cause artefactual groupings of unrelated genera that have undergone rapid rRNA evolution (Philippe and Adoutte 1998; Philippe et al. 2000). Previous studies have used HSPs to elucidate phylogenetic relationships in eukaryotes (Plesofsky-Vig 1992, Stechmann 2003). The ubiquitous and metropolitan prevalence of HSP allow for comparison of organisms as distantly related as that of bacteria, *Escherichia coli* and flies, *Drosophila melanogaster* (Lindquist 1986). Additionally, the importance of HSP in evolution and speciation has been well documented (Sorensen 2003). Finally, the difference between families of HSPs allows researchers many options in determining precision and resolution in describing phylogenetic relationships by utilizing the more conserved HSP90 domain or the more varied sHSP family to define relationships at any level of categorization, from kingdom to species (Feder and Hofmann 1999). One of the greatest determining factors in host range for an invasive species is stress tolerance; an attribute directly related to HSP expression (Feder and Hoffman 1999). As such, the sequence variation of heat shock proteins offer an excellent resource to apply in defining phylogenetic relationships, and also to aid in revealing pest species range.

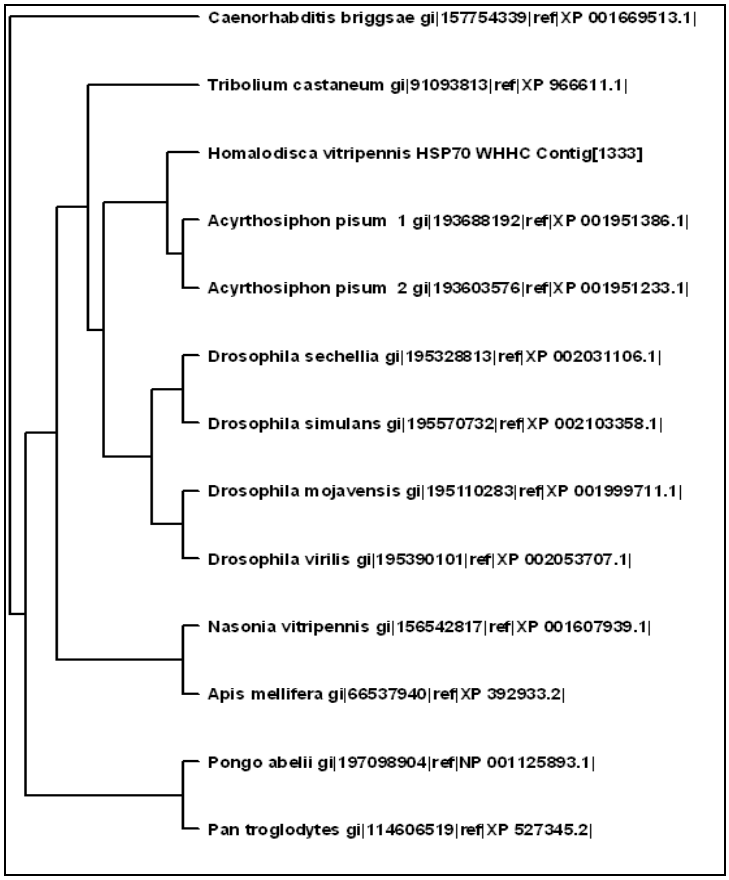

Figure 3. Cladogram heat shock 90 proteins constructed using sequence WHWTC-Contig[1285]. Subject sequences were analyzed using NCBI BLAST*p* search and aligned using ClustalW2 multiple alignment tool (www.ebi.ac.uk/Tools/ clustalw2/ index.html) and visualized using Treeview v1.6.6. (Organism name|accession number| reference number).

CONCLUSION

These results show that HSP can be used to accurately describe the phylogenetic history of GWSS, thus offering a novel target for molecular systematics. Additionally, this study is the first to describe any of the HSP sequences found inGWSS. We believe that understanding and sequencing heat shock protein encoding genes is an important step elucidating the underlying genetic determinants of pest species range and stress tolerance of GWSS.

REFERENCES CITED

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer E L, Studholme DJ, Yeats C, Eddy SR. (2004) The Pfam protein families database. *Nucl. Acids Res.* 32: D138-D141.

Bova MP, Huang Q, Ding L, Horwitz J. (2002) Subunit exchange, conformational stability, and chaperonelike function of the small heat shock protein 16.5 from *Methanococcus jannaschii*. *J Biol Chem* 277: 38468–38475.

Bukau B, Horwich AL. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell.* 92: 351-66.

Burdon RH. (1986) Heat shock and the heat shock proteins. *Biochem J*. 240: 313–324.

Craig EA. (1989) Essential roles of 70kDa heat inducible proteins. *Bioessays.* 11: 48-52.

Chang CJ, Garnier M, Zreik L, Rossetti V, Bove JM. (1993) Culture and serological detection of xylem-limiting bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Curr Microbiol* 27: 137-142.

Coudron TA, Brandt SL, Hunter WB. (2006) Molecular profiling of proteolytic and lectin transcripts in *Homalodisca vitripennis (*Hemiptera: Auchenorrhyncha: Cicadellidae) feeding on sunflower and cowpea. *Arch. of Insect Biochem. and Physiol*. 66: 76-88.

Cyr DM, Langer T, Douglas MG. (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70, *Trends Biochem Sci*. 19: 176-181.

Feder ME, Hofmann GE. (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Ann Rev Physio.* 61: 243–282.

Flaherty KM, DeLuca-Flaherty C, McKay DB. (1990) Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature*. 346 623-628.

Fu X, Chang Z (2004) Temperature-dependent subunit exchange and chaperone-like activities of Hsp16.3, a small heat shock protein from Mycobacterium tuberculosis. *Biochem Biophys Res Commun* 316: 291–299.
Gething MJ, ed. 1997. Guidebook to Molecular Chaperones and Protein-Folding Catalysts. Oxford, UK: Oxford Univ. Press.
Gu L, Abulimiti A, Li W, Chang Z (2002) Monodisperse HSP16.3 nonamer exhibits dynamic dissociation and reassociation, with the nonamer dissociation prerequisite for chaperone-like activity. *J Mol Biol* 319: 517–526.
Hoddle MS, Triapitsyn SV, Morgan DJW. (2003) Distribution and plant association records for *Homalodisca coagulata* (Hemiptera: Cicadellidae) in Florida. *Florida Entomol* 86: 89–91.
Hunter WB, Dowd SE, Katsar CS, Shatters RG Jr, McKenzie CL, Hall DG. (2009) Psyllid biology: expressed genes in adult Asian citrus psyllids, *Diaphorina citri* Kuwayama. *Open Entomol J*. 3: 18-29.
Jakob U, Buchner J. (1994) Assisting spontaneity: the role of Hsp90 and small Hsps as molecular chaperones. *Trends Biochem Sci.* 19: 205-211.
Jakob U, Lilie H, Meyer I, Buchner J. (1995) Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. *J Biol Chem*. 270: 7288–7294.
Lee GJ, Vierling E. (2000) A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiol*. 122: 189-197.
de Leon, JH, Jones WA, Morgan DJW. (2004) Population genetic structure of *Homalodisca coagulata (*Homoptera: Cicadellidae), the vector of the bacterium *Xylella fastidiosa* causing PD in grapevines. *Ann Entomol Soc Am.* 97: 809-818.
Lindquist S, Craig EA. (1988) The heat-shock proteins. *Annu. Rev. Genet*. 22: 631-677.
Mircetich SM, Lowe SK, Moller WJ, Nyland G. (1976) Etiology of almond leaf scorch disease and transmission of the causal agent. *Phytopath.* 66: 1-24.
Nadeau K, Das A, Walsh CT. (1993) Hsp90 chaperonins possess ATPase activity and bind heat shock transcription factors and peptidyl prolyl isomerases. *J Biol Chem*. 268: 1479-1487.
Notredame C, Higgins DG, Heringa J. (2000) T-Coffee: A novel method for multiple sequence alignments. *J Mol Bio*. 302: 205-217
Page RDM. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp Appl Biosciences.* 12: 357-358.
Pelham HR. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell.* 46: 959-961
Pelham H. (1988) Heat-shock proteins. Coming in from the cold. *Nature.* 332: 776-777.
Phillippe H, and Adoutte A. (1998) The molecular phylogeny of Eukaryota: solid facts and uncertainties. Pp. 25–56 in G. Coombs, K. Vickerman, M. Sleigh, and A. Warren, eds. Evolutionary Relationships Among Protozoa. Chapman & Hall, London.
Plesofsky-Vig N, Vig J, Brambl R. (1992) Phylogeny of the alphacrystallin- related heat-shock proteins. *J Mol Evol*. 35: 537–545.
Pooler MR, Hartung JS. (1995) Specific PCR detection and identification of *Xylella fastidiosa* strains causing citrus variegated chlorosis. *Curr Microbiol.* 31: 377-381.
Purcell AH. (1997) *Xylella fastidiosa,* a regional problem or global threat? *J Plant Pathol.* 79: 99-105.
Purcell AH, Saunders S, Hendson M, Grebus M, Henry M. (1999) Causal role of *Xylella fastidiosa* in oleander leaf scorch disease. *Phytopathol.* 89: 53-58.
Snutch TP, Heschl MF, Baillie DL. (1988) The Caenorhabditis elegans hsp70 gene family: a molecular genetic characterization. *Gene.* 64: 241-55.
Sobott F, Benesch JLP, Vierling E, Robinson CV. (2002) Subunit exchange of multimeric protein complexes: real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J Biol Chem*. 277: 38921–38929.
Sorensen JG, Kristensen GTN, Loeschcke V. (2003) The evolutionary and ecological role of heat shock proteins. *Ecol Lett*. 6: 1025–1037.
Sreedhar AS, Kalmar E, Csermely P, Shen YF. (2004) Hsp90 isoforms: functions, expression and clinical importance. *FEBS Lett*. 562: 11-15.
Stechmann A, Cavalier-Smith T. (2003) Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *J Mol Evol*. 57: 408–419.
Takiya DM, McKamey SH, Cavichioli RR. (2006) Validity of *Homalodisca* and of *GWSS* as the name for glassy-winged sharpshooter (Hemiptera: Cicadellidae: Cicadellinae). *Annals Entomol Soc Am.* 99: 648-655.
Thompson JD, Higgins DG, Gibson TJ. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res*. 22: 4673-4680.
Turner W, Pollard H. (1959) Life histories and behavior of five insect vectors of phony peach disease. *Technical Bulletin of the United States Dept Ag.* 28: 1188.
Young, DA. (1958) A synopsis of the species of *Homalodisca* in the United States. *Bulletin Brooklyn Entomol Soc.* 53: 7–13.

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.