PROTEIN IDENTITIES FROM BLUE-GREEN SHARPSHOOTER EXPRESSED SEQUENCE TAGS: EXPANDING LEAFHOPPER VECTOR BIOLOGY

Project Leader:
Wayne B. Hunter
USDA, ARS
U.S. Hortic. Research Lab
Fort Pierce, FL 34945
Wayne.hunter@ars.usda.gov

Cooperators:
Alexander H. Purcell
Division of Insect Biology
University of California
Berkeley, CA 94720

C.M. Wistrom
Division of Insect Biology
University of California
Berkeley, CA 94720

Laura E. Hunnicutt
Genomics Sciences
North Carolina State Univ.
Raleigh, NC

Reporting Period: The results reported here are from work conducted 2006 to 2007.

ABSTRACT
Although *Graphocephala atropunctata* (Signoret) (Hemiptera: Cicadellidae) is the native 'blue-green sharpshooter,' BGSS, which has been a major vector of Pierce's disease in vineyards in California for nearly a century, only recently has any genomic information become available. Due to the importance of the BGSS as the principal native vector of Pierce's disease (Almeida et al., 2005), we chose to examine the biology of the BGSS using a genomics approach. A cDNA library was made from adult BGSS, and 8,160 expressed sequence tags, ESTs, were produced. After quality scoring 6,836 sequences underwent assembly which produced a set of 1,915 sequences that putatively represented distinct transcripts. Initial annotation of this dataset identified 44 putative protein sequences were characterized through *in silico* analyses, and published in the NCBI database (Accession numbers are listed in Table 1). BLASTX analysis identified 10 significant homology matches to heat shock proteins, HSP, which are the focus of this study due to their overall importance and functions in maintaining protein integrity and activity during stressful conditions, such as extreme heat, cold, drought or crowding. A putative full-length small heat shock protein was produced NCBI database accession DQ445538.1. Many other genes of interest which have various functions in leafhopper biology and physiology have also been identified but are not reported herein. The EST sequences reported in this study have been deposited in GenBank's dbEST under accession numbers EH655849-EH662328 and EH662332.

INTRODUCTION
Sharpshooter leafhoppers are vectors of a number of destructive plant diseases caused by the plant infecting bacterium, *Xylella fastidiosa*. The native leafhopper, *Graphocephala atropunctata* (Signoret) (Hemiptera) 'blue-green sharpshooter,' BGSS, is a major vector of Pierce's disease, PD, in vineyards in California and has been spreading PD for nearly a century. Unlike invasive glassy-winged sharpshooter (GWSS), the BGSS is smaller in size and prefers to feed in riparian habitats, thus keeping most infections in grapevine to the bordering plants. The crops grown in the San Joaquin Valley face extremely hot temperatures in the summers, often over 38°C, and freezing temperatures during the winter months which creates a highly stressful environment for sharpshooters and other insects. The importance of BGSS as the principal native vector of PD in grapes (Almeida et al., 2005), led us to examine the biology of BGSS using a genomics approach to determine how these sharpshooters are dealing with these harsh environmental conditions. The benefits gained from expressed sequence tag, EST, studies have been definitively demonstrated through many studies on insects (*Drosophila,* Honey bee, Aphids, Silk worm) and other organisms. Current production of genomic information on the BGSS is now available (Hunter et al., NCBI) which was derived from single-pass sequencing of cDNA clones. The identification of genes associated with leafhopper biology continues to expand as more ESTs are produced from different species. Annotation of these datasets advances current understanding of leafhopper biological pathways while providing clues to the genetic basis of such processes as insect-pathogen, and insect-plant interactions. The availability of genomic data on BGSS, which is one of three sets of genomic data on sharpshooters (Hunter, NCBI) provides a solid foundation for future studies in functional genomics to advance the creation of novel genomics-based management strategies for this and other leafhopper vectors of plant diseases. Herein we report on the production and annotation of 44 putative proteins from BGSS, and the annotation of heat shock proteins from BGSS.

OBJECTIVES
To produce, annotated, and identify genes critical to sharpshooter survival, such as heat shock proteins. Genomics advances our understanding of the genetic basis of leafhopper biology. The results build the foundation for functional genomic studies, aimed towards development of better leafhopper management strategies to reduce the spread and impact from Pierce's disease.

RESULTS & DISCUSSION
Adult *G. atropunctata* were obtained from a colony managed by Alexander Purcell at the University of California (Berkeley, CA). Founder BGSS were field-collected from mugwort (*Artemisia douglasiana* L) in Guerneville, CA (Sonoma Co.) and subsequently reared on sweet basil (*Ocimum basilicum* L.) at 25°C (+10°C/-5°C), 14 L: 10 D. First-generation progeny were macerated in RNA*later*® RNA Stabilization Reagent (Ambion, Austin, TX) and stored at -40ºC prior to shipment.

Sequence analysis - Base calling was performed using TraceTuner™ (Paracel, Pasadena, CA) and low-quality bases (quality score <20) were stripped from both ends of each EST. Quality trimming, vector trimming, and sequence fragment alignments were executed using Sequencher™ software (Gene Codes, Ann Arbor, MI). Sequencher contig assembly parameters were set using a minimum overlap of 50 bp and 90% identity. Contigs joined by vector sequence were flagged for possible miss-assembly and manually edited. Putative sequence identity was determined based on BLAST similarity searches using the NCBI BLAST server (www.ncbi.nlm.nih.gov) with comparisons made to both non-redundant nucleic acid and protein databases using BLASTN and BLASTX, respectively. Matches with an E-value ≤-10 were considered significant and were classified according to the Gene Ontology (GO) classification system. The 5′-single pass sequencing of a cDNA library derived from adult BGSS yielded 8,160 ESTs, of which 6,836 were designated as "high quality". Forty putative proteins identified from BGSS are listed in Table 1. Homologous matches to heat shock proteins, HSP 20, 40, 70, and 90, are shown in Table 2. A putative full-length protein was matched to a small heat shock protein Table 3. Protein sequence for the BGSS small heat shock protein, sHSP, to: Locust, Pink hibiscus mealybug, Honey bee, Parasitoid Nasonia, and the Mosquito showed most similar to Locus and mealybug.

Small heat shock protein - Small heat shock considered α-crystallin proteins, sHSP, are defined by a conserved sequence of approximately 90 amino acid residues, termed the α-crystallin domain (MacRae, 2000; Taylor & Benjamin, 2005). Functionally, most sHSP display *in vitro* chaperone-like activity, that is, the capacity to interact with other HSP to prevent aggregation and to keep proteins in a folded, competent state (Franck et al., 2004), they occur in all Kingdoms, but not in all organisms. Small HSP have been implicated in an astounding variety of processes, such as enhancing cellular stress resistance (Feder & Hofmann, 1999), regulating actin and intermediate filament dynamics (Wieske et al., 2001), inhibiting apoptosis, modulating membrane fluidity (Tsvetkova et al., 2002), and regulating vasorelaxation (Flynn et al., 2003). Amino acid sequence comparisons of the BGSS sHSP with other sHSP showed the common motif of the alpha-crystallin domain, NCBI GenBank database (http://www.ncbi.nlm.nih.gov/blast). The α-crystallin domain is a hallmark of the α-crystallin/small HSP superfamily. The putative α-crystallin domain was present at amino acid positions 64–146. The percentage identity among BGSS to other insect sHSP deduced amino acid sequences varied from 44% to 56%, with the highest similarity between *Locusta migratoria* HSP 20.7 and *Maconellicoccus hirsutus*, sHSP, (Table 3) and the lowest to *Rattus norvegicus* Alpha-crystallin A chain (not shown).

The occurrence of sharpshooters in high densities during summer months which may reach extremely hot temperatures as in CA and FL produce similar conditions of stress on sharpshooters. The insects must be able to prevent the crosslinking or deformation of proteins to maintain their function and life. Comparative genomics permits us to examine the full length cDNAs of HSP 20.5, 20.6, 20.7, 40, 70 and HSP 90 of the migratory locust which have been cloned and sequenced to make reasonable associations to similar proteins in the BGSS. The functions of HSP are well studied and further comparisons between the BGSS and other organisms provide key information for the examination and characterization of HSP in BGSS. We are using these findings in other insects, like locusts, to expand our understanding of the roles and pathways HSP play in BGSS survival.

CONCLUSIONS

The information gained from this study represents the first investigation regarding the transcriptome of *G. atropunctata*, BGSS. The resultant sequence data has produced valuable information on sharpshooter heat shock proteins, and identified many other physiologically important transcripts. The data has been made available to the public to facilitate the use of this information in further studies on sharpshooters. The important role of heat shock proteins to sustain protein integrity and other critical functions make them suitable for further examination as potential critical genetic targets which may be altered to reduce leafhopper populations. Collectively, these genetic sequences provide the strong foundation needed for further functional genomics studies which will enable the development of more biorational management strategies to reduce losses from the diseases spread by this and other leafhopper pests.

Table 1. Proteins from *Graphocephala atropunctata*, the blue-green sharpshooter, 44 Putative Protein Sequences, Published 03-03-06, NCBI. http://www.ncbi.nlm.nih.gov/sites/entrez

Definitions			Clone Accesion Number	
file	WHGA0016	(similar to CG2210).sqn:	WHGA0016	DQ445499
file	WHGA0091	(ribonuclease).sqn:	WHGA0091	DQ445500
file	WHGA0096	(cytochrome C oxidase polypeptide:	WHGA0096	DQ445501
file	WHGA0097	(s9e ribosomal protein).sqn:	WHGA0097	DQ445502
file	WHGA0105	(ubiquitin fusion protein).sqn:	WHGA0105	DQ445503
file	WHGA0114	(tropomyosin 1).sqn:	WHGA0114	DQ445504
file	WHGA0124	(thioredoxin-like protein).sqn:	WHGA0124	DQ445505
file	WHGA0140	(CSF signaling molecule).sqn:	WHGA0140	DQ445506
file	WHGA0151	(mitochondrial ATP synthase).sqn:	WHGA0151	DQ445507
file	WHGA0169	(cytochrome c reductase).sqn:	WHGA0169	DQ445508
file	WHGA0271	(LIM protein).sqn:	WHGA0271	DQ445509
file	WHGA0283	(tumor protein):	WHGA0283	DQ445510
file	WHGA0301	(oligomyocin sensitivity protein):	WHGA0301	DQ445511
file	WHGA0310	(cytochrome oxidase Va):	WHGA0310	DQ445512
file	WHGA0380	(ferritin):	WHGA0380	DQ445513
file	WHGA0381	(calmodulin):	WHGA0381	DQ445514
file	WHGA0392	(ADP-ATP translocase):	WHGA0392	DQ445515
file	WHGA0411	(NADH dehydrogenase 1 alpha):	WHGA0411	DQ445516
file	WHGA0412	(ribosomal protein L23):	WHGA0412	DQ445517
file	WHGA0430	(Histone3A):	WHGA0430	DQ445518
file	WHGA0449	(vacuolar ATPase subunit E):	WHGA0449	DQ445519
file	WHGA0585	(ribosomal protein 4e):	WHGA0585	DQ445520
file	WHGA0587	(ribosomal protein L37Ae):	WHGA0587	DQ445521
file	WHGA0762	(ribosomal protein S23e):	WHGA0762	DQ445522
file	WHGA0689	(elongation factor 1d):	WHGA0689	DQ445523
file	WHGA0199	(ribosomal protein 49).sqn:	WHGA0199	DQ445524
file	WHGA0225	(V-ATPase).sqn:	WHGA0225	DQ445525
file	WHGA0228	(NADH-ubiquinone reductase).sqn:	WHGA0228	DQ445526
file	WHGA0230	(cytochrome oxidase VIa).sqn:	WHGA0230	DQ445527
file	WHGA0257	(ribosomal protein L27Ae).sqn:	WHGA0257	DQ445528
file	WHGA0270	(mitochondrial ATP synthase).sqn:	WHGA0270	DQ445529
file	WHGA0783	(cytochrome oxidase Vb):	WHGA0783	DQ445530
file	WHGA0824	(cytochrome c):	WHGA0824	DQ445531
file	WHGA0900	(tropomyosin):	WHGA0900	DQ445532
file	WHGA0927	(PPIase):	WHGA0927	DQ445533
file	WHGA1072	(ribosomal protein L19e):	WHGA1072	DQ445534
file	WHGA1215	(GABA):	WHGA1215	DQ445535
file	WHGA1242	(mito. ATP synthase gamma):	WHGA1242	DQ445536
file	WHGA1340	(mito. porin):	WHGA1340	DQ445537
file	WHGA1462	(small heat shock protein):	WHGA1462	DQ445538
file	WHGA1611	(ribosomal protein L18A):	WHGA1611	DQ445539
file	WHGA2669	(mito. ATP synthase e):	WHGA2669	DQ445540
file	WHGA2689	(reductase complex QP-C):	WHGA2689	DQ445541
file	WHGA3412	(ribosomal protein S7e):	WHGA3412	DQ445542

Table 2. Heat shock protein homologs to transcripts in *Graphocephala atropunctata*, the blue-green sharpshooter, cDNA library. The full-length cDNA to a small heat shock protein, HSP 20.1, was sequenced and posted in NCBI database accession DQ445538.1. Partial sequences were identified homologous to HSP 40, HSP 70, and HSP 90.

Clone	Descriptor	E-value
Contig[1462]928 bp	gb|EAA04497.3| small HSP 20 *Anopheles gambiae*	8e-040
Contig[1012] 737 bp	gb|AAG42838.1| heat shock 70 kDa protein*Leptinotarsa decemlineata* (Colorado potato beetle)	6e-028
Contig[0130]1076 bp	ref|XP_623939.1| 90 kDa heat shock protein *Apis mellifera*	3e-084
Contig[1050] 741 bp	dbj|BAE44308.1| heat shock cognate protein 70 *Chilo suppressalis*	e-104
Contig[1901] 833 bp	gb|EAA08691.3| HSP cognate 70 *Anopheles gambiae*	6e-050
Contig[2381] 857 bp	gb|EAA03148.2| heat shock 40kD Anopheles gambiae	9e-022
WHGA057-76 560 bp	gb|AAL27404.1| 70 kDa heat shock protein *Artemia franciscana* brine shrimp	2e-045
WHGA051-87 578 bp	gb|AAO65964.1| heat shock protein 70 *Manduca sexta*	1e-054
WHGA079-33 749 bp	gb|AAO21473.1| hsp70 family member [*Locusta migratoria*]	2e-082
WHGA008-42 812 bp	dbj|BAD74196.1| heat shock protein hsp20.1 *Bombyx mori*	2e-037

Table 3. Alignment of conserved domain for Small Heat Shock Protein, Essential for life, from *Graphocephala atropunctata*, the blue-green sharpshooter. Conserved domain alignments were most similar to *Locusta migratoria* and *Maconellicoccus hirsutus*, sHSP (Expect = 5e-49), BLAST2, NCBI tools.

Similar alignment 182 aa	SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
	Conserved Domain Database hits				
Locusta migratoria	569	27	ABC84492	85816366	HSP 20.5 **Expect = 5e-49**
Graphocephala atropunctata	500	18	AB D98776	90820038	small HSP
Maconellicoccus hirsutus	500	18	ABM55532	121543671	small HSP **Expect = 5e-49**
Apis mellifera	431	18	XP_001...	110750766	Protein lethal(2) essential for life (Protein Efl21)
Aedes aegypti	423	18	XP_001...	157135561	lethal(2)essential for life protein, l2efl
Aedes aegypti	423	18	XP_001...	157135559	lethal(2)essential for life protein, l2efl

REFERENCES

Almeida, R.P.P., Wistrom, C., Hill, B.L., Hashim, J. and Purcell, A.H. 2005. Vector transmission of *Xylella fastidiosa* to dormant grape. Plant Disease 89: 419-424.

Almeida, R.P.P., Blua, M.J., Lopes, J.R.S. and Purcell, A.H. 2005. Vector transmission of *Xylella fastidiosa*: Applying fundamental knowledge to generate disease management strategies. Ann Entomol Soc America 98: 775-786.

Altschul, S.F., Madden, T.L., Schäffer, A.A, Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Feder, M.E., Hofmann, G.E. 1999. Heat shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu Rev Physiol 61: 243–282.
Flynn, C.R., Komalavilas, P., Tessier, D., Thresher, J., Niederkofler, E.E., Dreiza, C.M., Nelson, R.W., Panitch, A., Joshi, L., Brophy, C.M. 2003. Transduction of biologically active motifs of the small heat shock-related protein HSP20 leads to relaxation of vascular smooth muscle. FASEB J 17: 1358–1360.
Franck, E., Madsen, O., Rheede, T.V., Ricard, G., Huynen, M.A., de Jong, W.W. 2004. Evolutionary diversity of vertebrate small heat shock proteins. J Mol Evol 59: 792–805.
Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressivemultiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
Hunter, W.B., Dang, P.M., Bausher, Chaparro, J.X., McKendree, W., Shatters, R.G. Jr., McKenzie, C.L., Sinisterra, X.H. 2003. Aphid Biology: Expressed Genes from Alate *Toxoptera citricida* (Kirkaldy), Brown Citrus Aphid (Hemiptera: Aphididae). J Insect Science. http://www.insectscience.org/3.23
MacRae, T.H. 2000. Structure and function of small heat shock/α-crystallin proteins: established concepts and emerging ideas. Cell Mol Life Sci 57: 899–913.
Marchler-Bauer, A, Bryant, SH. 2004. CD-Search: protein domain annotation on the fly. Nucleic Acids Res 32(W): 327-331.
Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., Altschul, S.F. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res 29: 2994-3005.
Taylor, R.P., Benjamin, I.J. 2005. Small heat shock proteins: a new classification scheme in mammals. J Mol Cell Card 38: 433–444.
Tsvetkova, N.M., Horvath, I., Torok, Z., Wolkers, W.F., Balogi, Z., Shigapova, N., Crowe, L.M., Tablin, F., Vierling, E., Crowe, J.H., Vigh, L. 2002. Small heat-shock proteins regulate membrane lipid polymorphism. Proc Natl Acad Sci USA 99: 13504–13509.
Wieske, M., Benndorf, R., Behlke, J., Dolling, R., Grelle, G., Bielka, H., Lutsch, G. 2001. Defined sequence segments of the small heat shock proteins HSP25 and aB-crystallin inhibit actin polymerization. Eur J Biochem 268: 2083–2090.
Whitfield C.W., Band M.R., Bonaldo M.F., Kumar C.G., Liu L., Pardinas J.R., Robertson H.M., Soares M.B., Robinson G.E. 2002. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Research 12: 555-566.

FUNDING AGENCIES
Funding for this project was provided by the USDA Agricultural Research Service.