Open Access

Psyllid Biology: Expressed Genes in Adult Asian Citrus Psyllids, *Diaphorina citri* Kuwayama

W.B. Hunter*[,1], S.E. Dowd[2], C.S. Katsar[3], R.G. Shatters[1] Jr., C.L. McKenzie[1] and D.G. Hall[1]

[1]*United States Department of Agriculture, Agricultural Research Service, United States Horticultural Research Laboratory, Fort Pierce, Florida, USA;* [2]*United States Department of Agriculture, Agricultural Research Service, Cropping Systems Research Laboratory, Lubbock, Texas, USA;* [3]*United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine, Fort Lauderdale, Florida, USA*

Abstract: Where it occurs the Asian citrus psyllid, *Diaphorina citri*, Kuwayama (Hemiptera: Psyllidae) is considered the primary vector of Huanglongbing, HLB, disease of citrus trees. The plant pathogenic bacterium associated with HLB causes economic losses to citrus industries worldwide. To better understand the general biology of *D. citri*, we undertook a sequencing project from adult psyllids. Few genes have been isolated from psyllids however several insect genomic datasets are available for comparisons. We compared the psyllid data to genomic datasets of nematode, *C. elegans*, fruit fly, *D. melanogaster*, honey bee, *A. mellifera*, mosquito, *A. aegypti*, and human, *H. sapiens* since these have completed more thorough levels of annotation. We describe the first data set of ESTs from *D. citri*, the Asian citrus psyllid. A total of 5,906 cDNA clones were sequenced, resulting in 4,595 high-quality ESTs. Electronic removal of 1,487 sequences which matched to bacteria and viruses left an assembly of cDNAs resulting in a total of 636 psyllid sequences (544 contigs plus 92 singlets). The sequences underwent BLAST analyses using (Swissprot-Tremble 03-2007) and NCBI, nr databases which returned 53% with 'No significant match' in either the non-redundant protein or nucleic acid databases, providing new information to the scientific community. The *D. citri* gene expression data set advances current research efforts in the identification of genes and physiological processes of psyllids. Knowledge of these genes and proteins are being used in the development of novel management strategies against psyllids, and other sap feeding insects within the Order: Hemiptera.

Keywords: ACP, AsCP, *Diaphorina citri*, EST, Genome, Hemiptera, Huanglongbing, Psyllid.

INTRODUCTION

The Asian citrus psyllid, *Diaphorina citri*, Kuwayama (Hemiptera: Psyllidae), is a major pest of citrus causing damage through feeding, but more importantly as the vector of the most serious disease of citrus, referred to as Huanglongbing, HLB [1]. Huanglongbing disease is associated with the plant pathogenic bacterium, *Candidatus* Liberibacter asiaticus, [2, 3]. *Diaphorina citri* was first reported in Florida in 1998 and has since spread throughout Florida [4]. In 2005, HLB disease was reported to be in Florida [5]. The pathogen, a phloem-restricted alpha-proteobacterium [6] is associated with HLB which has been referred to by several names: in India as 'Citrus Dieback', more commonly referred to as 'Citrus Greening' a term used more broadly around the world, and 'Huanglongbing', HLB, or 'Yellow Dragon Disease' in China [1, 7]. The importance of *D. citri* as the vector of HLB prompted the need for a better understanding of psyllid biology. Since the genomic sequence of *D. citri* was not available, we produced expressed sequence tags (ESTs) from adult *D. citri*. Sequence tags offered an invaluable resource for the identification of genes associated with the specific biology of psyllids [8-10]. Application of a transcriptome approach has also led to other important discoveries such as insect pathogens, as in the discovery of the first virus in *D. citri* [11], the first viral pathogens of the red imported fire ant [12, 13] and viruses in the glassy-winged sharpshooter, *Homalodisca coagulata virus 1*, HoCV- 1 [14, 15], the discovery and identification of genes linked to the development and reproduction of aphids [16, 17], and in discoveries made in the studies of the specialized cells that harbor aphid endosymbionts [18, 19]. Additionally, ESTs and their accompanying cDNAs provide the means to construct arrays that can be used for transcript profiling on a genome-wide scale [20-23]. However, even without subsequent array analyses, a relatively large number of ESTs from a specific life stage provides clues toward the expression of specific genes important to the functions connected with that life stage [24-26]. Examination of transcriptome data sets, within statistical limitations, and in most cases, have shown that the abundance of a specific cDNA in the EST collection is a measure of gene expression [27] providing a "digital or electronic northern" that has been utilized to gauge relative gene expression in various tissues and can often direct the efforts of researchers towards genes of interest. Genetic information on hemipterans still lags behind that for other insects, thus it was essential to build an expression library de-

*Address correspondence to this author at the United States Department of Agriculture, Agricultural Research Service, United States Horticultural Research Laboratory, Fort Pierce, Florida, USA;
E-mail: Wayne.Hunter@ARS.USDA.GOV

1874-4079/09

rived from adult *D. citri* so that analyses of metabolism on a genome-wide scale could begin.

MATERIALS AND METHODS

Psyllid Rearing and Collection

Asian citrus psyllids, *D. citri*, were obtained from a colony established from field caught adults, maintained by C.L. McKenzie at the USDA, ARS, U.S. Horticultural Research Laboratory, Ft. Pierce, FL. Insects were reared on *Murraya paniculata* (L.) 'Orange-jasmine' seedlings in screen cages contained in an insectary, and held at 25˚C, 16 L: 8 D. Plants with new flush were cycled into cages on a weekly basis. High-density psyllid populations produced psyllids that were collected by aspiration; the adults collected were immediately submerged into liquid nitrogen prior to total RNA isolation.

cDNA Library Construction

Approximately 3,500 adult psyllids of mixed age and genders were used in the construction of an expression library. Whole psyllids were ground in liquid nitrogen and total RNA extracted using guanidinium salt-phenol-chloroform procedure as previously described by Strommer *et al*. [28]. Poly(A) + RNA was purified using two rounds of selection on oligo dT magnetic beads according to the manufacturer's instructions (Dynal, Oslo, Norway). cDNA was synthesized using Stratagene's ZAP-cDNA Synthesis Kit (Stratagene, La Jolla, CA, USA). Mass excision of the amplified library was carried out using Ex-Assist helper phage (Stratagene, La Jolla, CA, USA) and bacterial clones containing excised pBluescript SK(+) phagemids were recovered by random colony selection.

Sequencing of Clones

pBluescript SK(+) phagemids were isolated from cell cultures grown overnight in 96-deep well plates containing 1.8 mL of LB, supplemented with 100 µg/mL ampicillin. DNA was extracted using the Qiagen 9600 liquid handling robot and the QIAprep 96 Turbo miniprep kit according to the recommended protocol (QIAGEN Inc., Valencia, CA, USA). Sequencing reactions were performed using the ABI PRISM® BigDye™ Primer Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). Reactions were prepared in 96-well format using the Biomek2000™ liquid handling robot (Beckman Coulter, Inc., USA). Sequencing reaction products were precipitated with 70% isopropanol, resuspended in 15 µL sterile water and loaded onto an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Computer Analysis

ESTs were generated from single-pass 5′ end sequencing of the *D. citri* expression library. Base calling, quality trimming, vector trimming and sequence fragment alignments were performed by SeqMan Pro (DNAstar, USA). Low-quality bases (quality score <12) were trimmed from both ends of sequences. Assembly parameters were set using a minimum overlap of 30 bp, match spacing of 150bp minimum sequence length and 90% identity. Putative sequence identity was determined based on BLAST similarity searches (BLASTX and BLASTn) using WND.BLAST Dowd, against Uniprot (Dec2007), and NCBI, nr database. We compared the psyllid data to genomic datasets of nematode, *C. elegans*, fruit fly, *D. melanogaster*, honey bee, *A. mellifera*, mosquito, *A. aegypti*, and human, *H. sapien* since these have a higher level of completed annotations. (http://www.ncbi.nlm.nih.gov/genomes/; http://www.vectorbase.org/index.php;). A set of the originally described ESTs (4,595 were used for the current analyses) from adult *D. citri* which is available at GenBank, dbEST, for further studies, Accession numbers: DN201110-DN470410. GenBank http://www.ncbi.nlm.nih.gov

RESULTS

Identification of ESTs

Of the 5, 906 ESTs from the *D. citri* expression library, 4,595 ESTs were identified with lengths greater than 150 nucleotides subsequent to quality and vector trimming with an average length of 553 bases. Because multiple ESTs can be derived from a single gene, sequences were assembled into contigs to estimate the number of genes giving rise to the ESTs. Electronic removal of 1,487 sequences which matched to bacteria and viruses left a total assembly of 636 cDNAs from psyllid (544 contigs plus 92 singlets) (refer to Table **1** for the most abundant contigs). Thus this represents the maximal number of 'Psyllid unique' assembled sequences. Of that total, 299 of the contigs and singlets corresponded with a putative match in GenBank, while 337 of the cDNAs had 'No Significant Homology' to any sequence currently listed in GenBank. 'No Significant Homology' denotes a 'no match' to GenBank's database when similarity or identity searches were performed using BLASTX, and BLASTN, with an E value $<1^{-10}$. The unknown 53% of the cDNAs is due to the large number of unknown proteins still to be characterized in insects, especially in Hemiptera which are a more ancient lineage thus they may have a more diverse transcriptome (Fig. **1**). To examine the uniqueness within this psyllid dataset a comparison to five other genomic datasets (Table **4**); nematode, *Caenorhabditis elegans*, fruit fly, *Drosophila melanogaster*, honey bee, *Apis mellifera*, mosquito, *Aedes aegypti*, and human, *Homo sapiens* provided a distribution across established classes of putative functions based on homology to proteins that had been previously characterized in the NCBI database (Tables **1**,

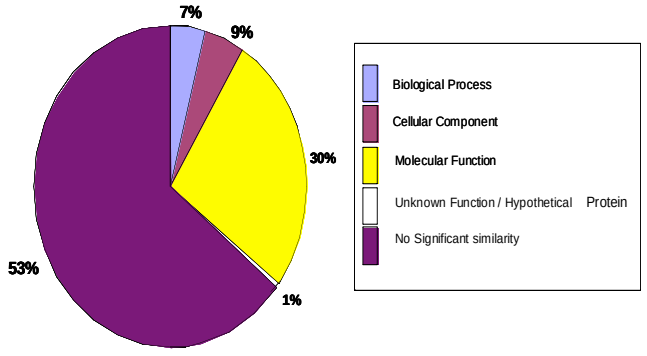

Fig. (1). Proportion of unassembled sequences from the Asian Citrus Psyllid EST dataset sorted using Gene Ontology. 53% of the cDNA's had 'No Significant Homology' to any known protein or nucleotides in the GenBank non-redundant database, BLASTX analysis at *E*-value of ≤-10. NCBI, http://www.ncbi.nlm.nih.gov/.

Table 1. Twenty Most Frequently Represented EST Sequences in Asian Citrus Psyllid

Sequence ID	Description[a]	Best Match[b]	E.C.#	E Value
WHDc 0014	Cytochrome c oxidase subunit I	Q69HD6	1.9.3.1	1.5035E-150
WHDc 0030	ATP synthase F0 subunit 6	Q69HD3		2.70664E-32
WHDc 0080	Cytochrome c oxidase subunit III	Q69HD2	1.9.3.1	5.07862E-46
WHDc 0021	Putative ferritin GF2	Q6PPI2		7.47398E-63
WHDc 0035	Cytochrome b	Q69HC6		2.7731E-119
WHDc 0023	Cytochrome c oxidase subunit II	Q69HD5	1.9.3.1	9.85564E-77
WHDc 0065	myosin light chain	Q7PUV3		8.0000E-48
WHDc 0076	ATP synthase F0 subunit 6	Q69HD3		6.18037E-06
WHDc 0086	NADH dehydrogenase subunit 4	Q69HC9	1.6.5.3	2.8032E-99
WHDc 0592	NADH dehydrogenase subunit 2	Q69HC4	1.6.5.3	1.86387E-36
WHDc 0125	Ejaculatory bulb specific prot. III precursor	Q9W1C9		7.48136E-32
WHDc 0187	Translationally controlled tumor protein	Q75VN3		1.85962E-80
WHDc 0020	NADH dehydrogenase subunit 5	Q69HD0	1.6.5.3	2.06965E-80
WHDc 1078	Ribosomal protein L32	Q6F449	3.6.5.3	2.37581E-49
WHDc 0049	S3e ribosomal protein	Q6EV05	3.6.5.3	2.6657E-100
WHDc 0175	NADH dehydrogenase subunit 1	Q68RL0		7.91356E-87
WHDc 0177	Thioredoxin-like protein	Q9U515		2.38394E-19
WHDc 0441	Ribosomal protein S27	Q962Q3	3.6.5.3	3.31106E-41
WHDc 0472	40S ribosomal protein S9	Q9VT06		5.0000e-86
WHDc 0502	Abnormal wing disc-like protein	Q8MUR5	2.7.4.6	9.19534E-73

[a]Description derived from GenBank nr database BLASTX search. [b]EST match indicates the gi number of the most similar annotated cDNA in GenBank nr database, BLASTX analysis matches had an *E*-value of $\leq$-10. NCBI.

Table 2. Biological Process

Gene Ontology Term	# Hits
ATP biosynthesis	2
ATP synthesis coupled proton transport	4
Biological process unknown	3
Carbohydrate metabolism	2
cation transport	1
cell adhesion	3
cell cycle	1
compound eye morphogenesis (*sensu* Endopterygota)	1
Cytokinesis	3
detection of virus	1
Electron transport	9
fatty acid alpha-oxidation	1
fatty acid metabolism	1

(Table 2). Contd…..

Gene Ontology Term	# Hits
Immune response	1
intracellular protein transport	4
larval or pupal development (*sensu* Insecta)	1
Metabolism	5
polysaccharide catabolism	1
protein deneddylation	1
protein import into nucleus, docking	1
Proteolysis	1
Regulation of progression through cell cycle	1
ribonucleoside diphosphate catabolism	1
Ribosomal large subunit assembly and maintenance	1
rRNA modification	1
Salivary gland cell autophagic cell death	1
selenocysteine incorporation	1
signal transduction	1
tetrahydrobiopterin biosynthesis	1
Transport	1
Two-component signal transduction system (phosphorelay)	2
Type I hypersensitivity	1
Total	59

Functional assignments of Psyllid EST unassembled sequences described are inferred from electronic evidence using top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. Definition term associated with each sequence defined according to The Gene Ontology Consortium.

Table 3. Molecular Function

Gene Ontology Term[a] (Parent)	# Hits	Gene Ontology Term[a] (Child)	# Hits
2-dehydro-3-deoxygluconokinase activity	1		
3'-5'-exoribonuclease activity	2		
4-alpha-hydroxytetrahydrobiopterin activity	1		
5-aminolevulinate synthase activity	1		
6-pyruvoyltetrahydropterin synthase activity	1		
acid phosphatase activity	1		
Actin binding	1		
acyl-CoA dehydrogenase activity	1		
Acyltransferase activity	1		
adenine phosphoribosyltransferase activity	1		
adenosylhomocysteinase activity	1		
adenyl-nucleotide exchange factor activity	1		
alpha-amylase activity	1		

(Table 3). Contd…..

Gene Ontology Term[a] (Parent)	# Hits	Gene Ontology Term[a] (Child)	# Hits
amino acid-polyamine transporter activity	1		
aminoacyl-tRNA hydrolase activity	1		
apyrase activity	2		
aspartic-type endopeptidase activity	1		
ATP binding	7		
beta-galactosidase activity	1		
Binding	3		
calcium ion binding	9		
carboxypeptidase activity	1		
catalytic activity	26		
		carboxypeptidase A activity	1
		Alpha-trehalose-phosphate synthase activity	1
		amino acid metabolism	1
		argininosuccinate lyase activity	1
		Biotin carboxylase activity	1
		cystathionine beta-synthase activity	1
		Electron transporter activity	1
		Epoxide hydrolase activity	1
		Fatty acid biosynthesis	1
		Glycolysis	1
		GMP reductase activity	1
		Lipid catabolism	1
		Lipid metabolism	1
		Metabolism	2
		molecular function unknown	1
		Phosphoribosylformylglycinamidine synthase activity	1
		pyruvate carboxylase activity	1
		Serine esterase activity	1
		ubiquitin activating enzyme activity	1
chymotrypsin activity	6	UDP-glucose 4-epimerase activity	1
copper ion binding	1	zinc ion binding	1
copper, zinc superoxide dismutase activity	1		
cyclin-dependent protein kinase activity	1		
cysteine-type endopeptidase activity	5		
Cytochrome-c oxidase activity	8		
cytokine activity	1		
D-alanyl-D-alanine endopeptidase activity	1		

(Table 3). Contd…..

Gene Ontology Term[a] (Parent)	# Hits	Gene Ontology Term[a] (Child)	# Hits
diphosphomevalonate decarboxylase activity	1		
DNA binding	11		
electron transporter activity	10		
enzyme inhibitor activity	6		
fructose-bisphosphate aldolase activity	1		
glucan 1,4-alpha-glucosidase activity	1		
glucose-6-phosphate 1-dehydrogenase activity	1		
glucose-6-phosphate isomerase activity	1		
Glutamate decarboxylase activity	1		
glutathione transferase activity	1		
GTP binding	2		
GTPase activity	5		
Guanylate cyclase activity	1		
glycotransferase activity	1		
Hydrolase activity,	5		
ion channel activity	1		
Isocitrate dehydrogenase (NADP+) activity	1		
kinase activity	1		
lactoylglutathione lyase activity	1		
Lysozyme activity	1		
Magnesium ion binding	4		
mannose-6-phosphate isomerase activity	1		
Microtubule motor activity	2		
molecular function unknown	8		
monooxygenase activity	3		
motor activity	7		
N-acetylmuramoyl-L-alanine amidase activity	1		
N-acetyltransferase activity	1		
NADH dehydrogenase activity	5		
nucleic acid binding	21		
		DNA binding	7
		Helicase activity	3
		Intracellular	1
		Nucleus	3
		RNA binding	3
nucleoside diphosphate kinase activity	1	S-adenosylmethionine-dependent methyl-transferase activity	1
nucleotide binding	8	structural constituent of ribosome	1

(Table 3). Contd…..

Gene Ontology Term[a] (Parent)	# Hits	Gene Ontology Term[a] (Child)	# Hits
odorant binding	1		
ornithine-oxo-acid transaminase activity	1		
Oxidoreductase activity	1		
oxygen transporter activity	1		
Peptidase activity	1		
peptidyl-prolyl cis-trans isomerase activity	2		
peroxidase activity	2		
phosphoglycerate dehydrogenase activity	1		
phosphoglycerate kinase activity	1		
phospholipase A2 activity	2		
Phosphoprotein phosphatase activity	3		
polypeptide N-acetylgalactosaminyltransferase activity	1		
protein binding	5		
protein kinase activity	2		
protein phosphatase inhibitor activity	1		
protein phosphatase type 1 activity	1		
protein tyrosine phosphatase activity	1		
protein-L-isoaspartate O-methyltransferase activity	1		
Pyridoxal kinase activity	1		
receptor activity	2		
RNA binding	14		
RNA polymerase II transcription factor activity	1		
serine-type endopeptidase activity	4		
serine-type endopeptidase inhibitor activity	1		
signal transducer activity	1		
Structural constituent of cell wall	2		
Structural constituent of ribosome	21		
Structural molecule activity	7		
Succinate dehydrogenase activity	1	oxidoreductase activity	1
sugar binding	1		
sugar porter activity	1		
Threonine endopeptidase activity	1		
transcription factor activity	1		
transferase activity	1		
translation elongation factor activity	9		
translation initiation factor activity	3		
translation release factor activity	1		
transporter activity	5		
tRNA binding	1		
tRNA ligase activity	4		
tubulin-tyrosine ligase activity	1		
Total	**241**		

[a]Classification is hierarchical: indented terms are children [c] of parent terms [p] listed above. Functional assignments of Psyllid EST sequences are inferred from electronic evidence using top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. Definition terms according to The Gene Ontology Consortium.

Table 4. Comparison of the Assembled Psyllid ESTs to Five Genomes

Species:	C. elegans	D. melanogaster	H. sapiens	A. mellifera	A. aegypti
E-Values					
≤ e-100	5	6	6	9	8
	3.50%	4.00%	3.90%	5.50%	4.70%
≤ e-50	41	48	43	54	52
	29.10%	32.00%	28.10%	33.10%	30.80%
≤ e-20	61	66	69	63	68
	43.30%	44.00%	45.10%	38.60%	40.20%
≤ e-10	34	30	35	37	41
	24.10%	20.00%	22.90%	22.70%	24.30%
TOTALS	141	150	153	163	169

Counts and percentages of significant matches over four categories of E-values, comparing Psyllid sequences to Nematode, Fruit fly, Human, Honey bee, and Mosquito, respectively. Psyllid data compared to each species separately, BLASTX analysis. No significant differences between species at each E-value.

2-Biological Process, **3**- Molecular Functions). Psyllid EST's in relation to their putative protein homologues (BLASTX) had the greatest overall similarity to the mosquito, *A. aegypti* (homology matches better than E-value ≤ e-10). Individual pairwise comparisons to the five genome databases resulted in similar distribution patterns of homology matches at each of four categories of E-values (ranges were from ≤ e-10 to ≤ e-20 to ≤ e-50 to ≤ e-100). There were no significant differences within each E-value analysis. Chi Square significance probability levels at 0.05, df= 4; 1≤ E-100, 1.588. 1≤ E-50, 2.630, 1≤ E-20, 0.691, 1≤ E-10, 1.842 (Table **4**). The high rate of non significantly matched sequences which estimates potential unique sequences within the *D. citri* cDNA library is most likely to be an overestimation due to several factors, such as computer alignment parameters, as well as low quality internal sequences [29]. Moreover, assembled sequences may have lacked an open reading frame because they were too short causing cDNAs to consist mostly or entirely of a noncoding region (e.g., 3′ untranslated region [16, 30, 31]. Annotation will be improved by obtaining more, and longer assembled sequences resulting in a higher proportion of sequences that contain a protein coding region. Longer assembled sequences result in an increase of significant matches as shown to be true for cDNA's from plants [30].

Genes of Interest

The psyllid genome size was estimated using Feulgen densitometry to be 0.35 pg, SD = 0.0291 (Hunter and Ardila-Garcia unpublished results) using five individuals, as in Hardie *et al.* [32](2002). The distribution of EST sequences sorted by enzyme class shows that Hydrolases make up the majority of enzymes identified (64%) with the Oxidoreductase class second (14%) (Table **5**). Within the Hydrolase class identified transcripts occurred across 10 subclasses (Table **6**). Also of interest was the identification of several specific enzymes. Aspartic protease (EC 3.4.23) (Psyllid sequence, partial 1100 bp, DN468794, DN467920), similar to vertebrate Cathepsin D. The percent similarities of amino acid sequence of the aspartic protease from *D. citri* to other organisms, averaged 57% similarity, *E-value*, e^{-84} (BLASTX analyses top three - 57% - Mosquito, *Aedes aegypti*, AAA29350; 57% - Bee, *Apis mellifera*, XM392857; 58% - Japanese flying squid, *Todarodes pacificus*, BAD15111, AB106552.1). Also identified were cDNA's with similarities to Glutathione S-transferase (BLASTX analysis, partial 672 bp, average *E-value*, e^{-46}) (DN466244, DN466232), and to cytochrome P450 monooxygenase, with similarity to the *Cyp4* gene family, (TBLASTX analysis, partial 687 bp, average *E-value*, e^{-21}) (DN466541, DN466559, DN466560, DN468646, DN469747).

DISCUSSION

While the Asian Citrus Psyllid, *Diaphorina citri*, is an economically important pest of citrus, overall very little has been published concerning psyllid genomics. While the results reported here provide an important dataset, albeit a small dataset covering ~4% of the total potential transcripts in psyllids, these data provide the impetus for further functional genomics approaches to be undertaken to advance the understanding of psyllid biology. Currently two more cDNA libraries are being sequenced one from psyllid midguts and another from adult male testes (Hunter in progress). These results have received attention from the Florida and California Citrus Industries, to consider funding a full metagenomic study of the Asian citrus psyllid to identify all associated microbes to address questions surrounding the severe Huanglongbing disease of citrus which threatens citrus production worldwide (Hunter, http://www.ars.usda.gov/pandp/people/people.htm?personid=11768; http://www.doacs.state.fl.us/pi/hlb_conference/Proceedings.pdf). Even within the Hemiptera, only recently are efforts completing full genome projects. In 2005, two Hemiptera genome projects were started which targeted: 1) the Pea aphid, *Acyrthosiphon pisum*, Family *Aphididae*, which was released early 2009 http://www.aphidests.org/; http://www.aphidbase.com/aphidbase/ and 2) the insect vector of Chagas Disease, *Rhodnius prolixus*, Family *Reduviidae*, http://www.vectorbase.org/index.php. Completion of these genomes will greatly aid efforts working on hemipteran pests. One of the major advantages of EST projects is that they provide imme-

Table 5. Enzymatic Classification of Asian Citrus Psyllid Sequences

EC #	Class	Subclass	# sequences
1.1	Oxidoreductase	acting on the CH-OH group of donors	7
1.11	Oxidoreductase	Peroxidases	2
1.14	Oxidoreductase	Oxygen	4
1.3	Oxidoreductase	Acting on the CH-CH Group of Donors	3
1.6	Oxidoreductase	Acting on NADH or NADPH	11
1.7	Oxidoreductase	Acting on other Nitrogenous Compounds as Donors	1
1.9	Oxidoreductase	Acting on a Heme Group of Donors	8
2.1	Transferase	Transferring One-Carbon Groups	2
2.3	Transferase	Acyltransferases	1
2.4	Transferase	Glycosyltransferases	4
2.5	Transferase	Transferring alkyl or aryl groups, other than methyl groups	1
2.6	Transferase	Transferring nitrogenous groups	1
2.7	Transferase	Transferring phosphorus-containing groups	19
3.1	Hydrolases	Acting on ester bonds	11
3.2	Hydrolases	Glycosidases	11
3.3	Hydrolases	Acting on ether bonds	2
3.4	Hydrolases	Acting on peptide bonds (Peptidases)	13
3.5	Hydrolases	Acting on Carbon-Nitrogen Bonds, other than Peptide Bonds	1
3.6	Hydrolases	Acting on Acid Anhydrides	127
4.1	Lyases	Carbon-Carbon Lyases	4
4.2	Lyases	Carbon-Oxygen Lyases	4
4.3	Lyases	Carbon-Nitrogen Lyases	1
4.4	Lyases	Carbon-Sulfur Lyases	1
4.6	Lyases	Phosphorus-Oxygen Lyases	1
5.1	Isomerases	Racemases and Epimerases	1
5.2	Isomerases	*cis-trans*-Isomerases	2
5.3	Isomerases	Intramolecular Oxidoreductases	4
6.1	Ligases	Forming Carbon-Oxygen Bonds	4
6.3	Ligases	Forming Carbon-Nitrogen Bonds	6
6.4	Ligases	Forming Carbon-Carbon Bonds	1
Total			**258**

Unassembled EST's by hierarchical classification. Functional assignments of *Psyllid sequences* described are inferred from electronic evidence using top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. Definition according to The International Union of Biochemistry and Molecular Biology's Enzyme classification system.

diate information on psyllid biology and provide sequences for further functional genomics [16, 17, 33]. While the number of functional genes is not correlated to genome size [34] http://genomesize.com/results/) the psyllid genome was estimated to be (~0.35 pg) about a third the size of the whitefly *Bemisia argentifolii* (~1.1 pg [33], three and a half times smaller than the glassy-winged sharpshooter (~1.24 pg, Hunter unpublished results) and slightly smaller than the pea aphid genome at 464 Mb (http://www.hgsc.bcm.tmc.edu/projects/aphid/). Comparison of the psyllid dataset to five genomes, nematode, fruit fly, human, honey bee, and mosquito, demonstrated that EST annotation results in similar distributions across other organisms, most likely due to the nature of these datasets each identifying common genes early

in their production and each relying on strong comparisons among 'finished' genomes. Thus they are dependent upon a common annotation knowledge approach. As the annotations increase within these and other genome databases so will the ability to better compare and annotate new sequences through *in silico* analyses. Furthermore, the high incidence of 'no significant' and 'unknown protein' matches in EST datasets also demonstrates the need for further efforts on insects and arthropods to produce and to characterize more sequences from emerging EST and genome projects. As more genes, transcripts, and proteins are characterized the ability for rapid and accurate *in silico* annotations increases. While these advances are the cornerstone of the ability to compare sequences and functions through *in silico* analyses, the benefits from this and similar studies are the increased genomic data from psyllids which elucidates the genetic basis of psyllid biology. Continued work will rapidly advance the use and application of this information into the development of novel management strategies against these agricultural pests to reduce the spread of HLB disease in citrus trees.

Targets of Interest within the D. citri cDNA Database

While this study has focused on adult psyllids which disperse and spread HLB, there were other transcripts identified which provide important guidance for future research on psyllid biology. These include proteins that are critical for function and development of psyllid brain, nerve and synaptic transmission, muscle formation, reproduction of sperm and eggs, and the development of insecticide resistance (refer to Tables **2**, **3**). The Aspartic protease transcript identified is similar to vertebrate Cathepsin D, the function of which is thought to mediate the processing of yolk proteins in the oocyte. Aspartic proteases are reported to have important functions in the production of viable eggs [35, 36], and may prove useful in expanding our understanding of egg development in psyllids. Additionally the identification of transcriptional members within specific metabolic pathways broadens our knowledge of how *D. citri* evolved to feed specifically on citrus and its' near relatives. The ability of *D. citri* to detoxify citrus allelochemicals, along with the potential for the rapid development of insecticide resistance is of interest in the development of new management practices. Several enzymes identified are known to function in the metabolic breakdown of toxins. Different organisms use a variety of reactions to metabolize an amazing variety of organic molecules [37-40]. These reactions usually render toxic compounds more hydrophilic, thus generally less toxic and more easily excreted from the body [41]. The ability of insects to detoxify phytotoxins in plants probably evolved as many insects broadened their host range and encountered these defensive chemistries. However, insects continue to demonstrate their ability to metabolize xenobiotics, 'chemical substances that are foreign to the biological system' (def: http://medical.webends.com/kw/Xenobiotics). Examples of this are studies of P450-based insecticide resistance which have been shown to occur in most insect pests [38, 42, 43]. Examples of these transcripts reported from other insects which were also identified in *D. citri* include: Glutathione S-transferase, and several cytochrome P450 monooxygenases which include a transcript from the *Cyp4* gene family. [44-49]. Genes within the *Cyp4* family play a role in the development of insecticide resistance in Lepidoptera, *Helicoverpa armigera* [46], Coleoptera, *Diabrotica virgifera* [48], and Diptera, *Anopheles gambiae* [47, 50] and may similarly be examined for their role in the development of insecticide resistance in psyllids.

Table 6. Classification of E.C. Subclass 3.6 Hydrolases, Psyllid Unassembled EST's by Hierarchical Classification

Subclass	Function	# Sequences
3.6.1	In Phosphorus-Containing Anhydrides	
3.6.1.1	In Phosphorus-Containing Anhydrides, inorganic diphosphatase	1
3.6.1.15	In Phosphorus-Containing Anhydrides, nucleoside-triphosphatase	5
3.6.1.5	In Phosphorus-Containing Anhydrides, apyrase	2
	Acting on acid anhydrides	
3.6.3.12	Acting on acid anhydrides; catalysing transmembrane movement of substances, K+-transporting ATPase	1
3.6.3.14	Acting on acid anhydrides; catalysing transmembrane movement of substances, H+-transporting two-sector ATPase	7
3.6.3.6	Acting on acid anhydrides; catalysing transmembrane movement of substances, H^+-exporting ATPase	1
	Acting on GTP	
3.6.5.1	Acting on GTP; involved in cellular and subcellular movement, heterotrimeric G-protein GTPase	5
3.6.5.3	Acting on GTP; involved in cellular and subcellular movement, protein-synthesizing GTPase	104
3.6.5.4	Acting on GTP; involved in cellular and subcellular movement, signal-recognition-particle GTPase	1
Total		**127**

Functional assignments of *Psyllid sequences* described are inferred from electronic evidence of top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. Definitions according to The International Union of Biochemistry and Molecular Biology's Enzyme classification system.

CONCLUSIONS

We provide the first functional genomics project for psyllids (Hemiptera: Psyllidae). The set of sequences developed in this study makes available a cDNA sequence dataset for an important disease vector, the Asian citrus psyllid, *D. citri*, which is rapidly spreading Huanglongbing disease across the USA. The availability of these sequences opens the door for further investigations into important questions regarding *D. citri* biology, development, insecticide resistance, and disease interactions.

ACKNOWLEDGMENTS

We thank P.M. Dang, Genomics lab, ARS, U.S. Horticultural Research Laboratory, Ft. Pierce, FL, for sequencing, L.E. Hunnicutt, Biological Science Technician, U.S. Horticultural Research Laboratory, Ft. Pierce, FL for cDNA library construction and data analysis, and K. Moulton, T. Panvil, Biological Science Technicians, U.S. Horticultural Research Laboratory, Ft. Pierce, FL for sample collections and preparations.

REFERENCES

[1] Halbert SE, Manjunath KL. Asian citrus psyllid (Sternorrhyncha: Psyllidae) and greening disease of citrus: a literature review and assessment of risk in Florida. Florida Entomol 2004; 87: 330-53.

[2] Bastianel C, Garnier-Semancik M, Renaudin J, Bové JM, Eveillard S. Diversity of "*Candidatus* Liberibacter asiaticus," based on the *omp* gene sequence. Appl Environ Microbiol 2005; 71: 6473-8.

[3] Garnier M, Jagoueix-Eveillard S, Cronje PR, Le Roux GF, Bové JM. Genomic characterization of a Liberibacter present in an ornamental rutaceous tree, *Calodendrum capense*, in the Western Cape province of South Africa. Proposal of '*Candidatus Liberibacter africanus* subsp. *capensis*'. Int J Syst Evol Microbiol 2000; 50: 2119-25.

[4] Halbert SE, Niblitt CL, Manjunath KL, Lee RF, Brown LG. Proceedings of the International Society of Citriculture IX Congress; Establishment of two new vectors of citrus pathogens in Florida 1016-1017, 2000 Dec 3-7; Orlando Florida, USA. 2003.

[5] Li W, Hartung JS, Levy L. Quantitative real-time PCR for detection and identification of *Candidatus Liberibacter* species associated with citrus Huanglongbing. J Microbiol Methods 2006; 66: 104-15.

[6] Bové JM. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. J Plant Pathol 2006; 88: 7-37.

[7] Mead FW. Florida Department of Agriculture and Consumer Services, Division of Plant Industry. Originally published FDACS/DPI Entomol Cir No. 180. Pub Number: EENY-33. Copyright 1998-2005 University of Florida; USA. 2005.

[8] Newman T, de Bruijn FJ, Green P, *et al.* Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol 1994; 106: 1241-55.

[9] Cooke R, Raynal M, Laudie M, *et al.* Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5,000 non-redundant ESTs. Plant J 1996; 9: 101-24.

[10] Rounsley SD, Glodek A, Sutton G, *et al.* The construction of *Arabidopsis* expressed sequence tag assemblies: a new resource to facilitate gene identification. Plant Physiol 1996; 112: 1177-83.

[11] Marutani-Hert M, Hunter WB, Katsar CS, Sinisterra XH, Hall DG, Powell CA. Reovirus-like sequences isolated from adult Asian citrus psyllid, (Hemiptera: Psyllidae: *Diaphorina citri*). Florida Entomol 2009; 92: 314-20.

[12] Valles SM, Strong CA, Dang PM, *et al.* A picorna-like virus from the red imported fire ant, *Solenopsis invicta*: initial discovery, genome sequence, and characterization. Virology 2004; 328: 151-7.

[13] Valles SM, Strong CA, Hunter WB, *et al.* Expressed sequence tags from the red imported fire ant, *Solenopsis invicta*: annotation and utilization for discovery of viruses. J Invertebr Pathol 2008; 99: 74-81.

[14] Hunter WB, Katsar CS, Chaparro JX. (2006) Molecular analysis of capsid protein of *Homalodisca coagulata virus -1*. A new leafhopper-infecting virus from the glassy-winged sharpshooter. J Insect Sci 2008; Available from: http://www.insectscience.org/6.28/

[15] Hunnicutt LE, Hunter WB, Cave RD, Powell CA, Mozoruk JJ. Genome sequence and molecular characterization of *Homalodisca coagulata virus-1*, a novel virus discovered in the glassy-winged sharpshooter (Hemiptera: Cicadellidae). Virology 2006; 350: 67-78.

[16] Hunter WB, Dang PM, Bausher MG, *et al.* Aphid biology: Expressed genes from alate *Toxoptera citricida*, the brown citrus aphid. J Insect Sci 2003; Available from: http://www.insectscience.org/3.23/

[17] Tagu D, Prunier-Leterme N, Legeai F, *et al.* Annotated expressed sequence tags for studies of the regulation of reproductive modes in aphids. Insect Biochem Mol Biol 2004; 34: 809-22.

[18] Nakabachi A, Shigenobu S, Sakazume N, *et al.* Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. Proc Natl Acad Sci USA 2005; 102: 5477-82.

[19] Wilson ACC, Dunbar HE, Davis GK, Hunter WB, Stern DL, Moran NA. A duel-genome microarray for the pea aphid, *Acyrthosiphon pisum*, and its obligate bacterial symbiont, *Buchnera aphidicola*. BMC Genomics 2006; 7: 50, doi:10.1186/1471-2164-7-50.

[20] DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997; 278: 680-6.

[21] Loftus SK, Chen Y, Gooden G, *et al.* Informatic selection of a neural crestmelanocyte cDNA set for microarray analysis. Proc Natl Acad Sci USA 1999; 96: 9277-80.

[22] Ruan Y, Gilmore J, Conner T. Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1,400 genes using cDNA microarrays. Plant J 1998; 15: 821-33.

[23] Sabater-Muñoz B, Legeai R, Rispe C, *et al.* Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). Genome Biol 2006; 7: R21, doi:10.1186/gb-2006-7-3-r21.

[24] Rafalski JA, Hanafey M, Miao GH, *et al.* New experimental and computational approaches to the analysis of gene expression. Acta Biochim Polonica 1998; 45: 929-34.

[25] Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 1999; 9: 950-9.

[26] Zhu Q, Arakane Y, Beeman RW, Kramer KJ, Muthukrishnan S. Characterization of recombinant chitinase-like proteins of *Drosophila melanogaster* and *Tribolium castaneum*. Insect Biochem Mol Biol 2008; 38: 467-77.

[27] Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res 1997; 7: 986-95.

[28] Strommer JN, Gregerson R, Vayda M. Methods in plant molecular biology and biotechnology. Boca Raton, Florida, USA: CRC Press 1993.

[29] White JA, Todd J, Newman T, *et al.* A new set of *Arabidopsis* expressed sequence tags from developing seeds: the metabolic pathway from carbohydrates to seed oil. Plant Physiol 2000; 124: 1582-94.

[30] Bausher M, Shatters R, Jr, Chaparro J, Dang P, Hunter W, Niedz R. An expressed sequenced tag (EST) dataset from *Citrus sinensis* L. Osbeck whole seedlings and the implications of further perennial source investigations. Plant Sci 2003: 165: 415-22.

[31] Whitfield CW, Band MR, Bonaldo MF, *et al.* Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Res 2002; 12: 555-66.

[32] Hardie DC, Gregory TR, Hebert PDN. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. J Histochem Cytochem 2002; 50: 735-49.

[33] Leshkowitz D, Gazit S, Reuveni E, *et al.* Whitefly (*Bemisia tabaci*) genome project: analysis of sequenced clones from egg, instar, and adult (viruliferous and non-viruliferous) cDNA libraries. BMC Genomics 2006; 7: 79, doi:10.1186/1471-2164-7-79.

[34] Gregory TR. The evolution of the genome. San Diego, USA: Elsevier 2005.

[35] Brooks S, Tyler CR, Carnevali O, Coward K, Sumpter JP. Molecular characterisation of ovarian cathepsin D in the rainbow trout, *Oncorhynchus mykiss*. Gene 1997; 201: 45-54.

[36] De Stasio R, Borrelli L, Kille P, Parisi E, Filosa S. Isolation, characterization and molecular cloning of cathepsin D from lizard

ovary: changed in enzyme activity and mRNA expression throughout ovarian cycle. Mol Reprod Dev 1999; 52: 126-34.
[37] Feyereisen R. Molecular biology of insecticide resistance. Toxicol Lett 1995; 82: 83-90.
[38] Feyereisen R. Insect P450 enzymes. Annu Rev Entomol 1999; 44: 507-33.
[39] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999; 29: 757-77.
[40] Scott JG, Wen Z. Cytochromes P450 of insects: the tip of the iceberg. Pest Manag Sci 2001; 57: 958-67.
[41] Coon MJ, Vaz AD, Bestervelt LL. Peroxidative reactions of diversozymes. Fed Am Soc Exp Biol J 1996; 10: 428-34.
[42] Berge JB, Feyereisen R, Amichot M. Cytochrome P450 monooxygenases and insecticide resistance in insects. Philos Trans R Soc B Biol Sci 1998; 353: 1701-5.
[43] Hodgson E. Comprehensive insect physiology, biochemistry, and pharmacology. UK: Oxford: Pergamon Press 1985; Vol. 11.
[44] Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Mol Biol 2005; 14: 3-8.
[45] Francis F, Vanhaelen N, Haubruge E. Glutathione S-transferases in the adaptation to plant secondary metabolites in the *Myzus persicae* aphid. Arch Insect Biochem Physiol 2005; 58: 166-74.
[46] Pittendrigh B, Aronstein K, Zinkovski E, *et al*. Cytochrome P450 genes from *Helicoverpa armigera*: expression in a pyrethroid-susceptible and -resistant strain. Insect Biochem Mol Biol 1997; 27: 507-12.
[47] Ranson H, Nikou D, Hutchinson M, *et al*. Molecular analysis of multiple cytochrome P450 genes from the malaria vector, *Anopheles gambiae*. Insect Mol Biol 2002; 11: 409-18.
[48] Scharf ME, Parimi S, Meinke LJ, Chandler LD, Siegfried BD. Expression and induction of three family 4 cytochrome P450 (CYP4)* genes identified from insecticide-resistant and susceptible western corn rootworms, *Diabrotica virgifera virgifera*. Insect Mol Biol 2001; 10: 139-46.
[49] Singh SP, Coronella JA, Benes H, Cochrane BJ, Zimniak P. Catalytic function of *Drosophila melanogaster* glutathione S-transferase DmGSTS1-1 (GST2) in conjugation of lipid peroxidation end products. Eur J Biochem 2001; 268: 2912-23.
[50] Amichot M, Brun A, Cuany A, *et al*. Induction of cytochrome P450 activities in *Drosophila melanogaster* strains susceptible or resistant to insecticides. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 1998; 121: 311-9.

Received: July 03, 2008 Revised: February 02, 2009 Accepted: February 11, 2009