Aphid biology: Expressed genes from alate Toxoptera citricida, the brown citrus aphid

Authors: Hunter, W. B., Dang, P. M., Bausher, M. G., Chaparro, J. X., McKendree, W., et al.

Source: Journal of Insect Science, 3(23) : 1-7

Published By: Entomological Society of America

URL: https://doi.org/10.1673/031.003.2301

Hunter WB, Dang PM, Bausher MG, Chaparro JX, McKendree W, Shatters RG, Jr., McKenzie CL, Sinisterra XH. 2003. Aphid biology: Expressed genes from alate *Toxoptera citricida*, the brown citrus aphid. 7pp. *Journal of Insect Science*, 3:23, Available online: insectscience.org/3.23

Journal of Insect Science

insectscience.org

Aphid biology: Expressed genes from alate *Toxoptera citricida*, the brown citrus aphid

W. B. Hunter, P.M. Dang, M.G. Bausher, J. X. Chaparro, W. McKendree, R.G. Shatters, Jr., C.L. McKenzie and X.H. Sinisterra

United States Department of Agriculture, Agricultural Research Service, U.S. Horticultural Research Laboratory, Fort Pierce, Florida, U.S.A.
whunter@ushrl.ars.usda.gov

Received 3 February 2003, Accepted 12 July 2003, Published 31 July 2003

Abstract

The brown citrus aphid, *Toxoptera citricida* (Kirkaldy), is considered the primary vector of citrus tristeza virus, a severe pathogen which causes losses to citrus industries worldwide. The alate (winged) form of this aphid can readily fly long distances with the wind, thus spreading citrus tristeza virus in citrus growing regions. To better understand the biology of the brown citrus aphid and the emergence of genes expressed during wing development, we undertook a large-scale 5' end sequencing project of cDNA clones from alate aphids. Similar large-scale expressed sequence tag (EST) sequencing projects from other insects have provided a vehicle for answering biological questions relating to development and physiology. Although there is a growing database in GenBank of ESTs from insects, most are from *Drosophila melanogaster* and *Anopheles gambiae*, with relatively few specifically derived from aphids. However, important morphogenetic processes are exclusively associated with piercing-sucking insect development and sap feeding insect metabolism. In this paper, we describe the first public data set of ESTs from the brown citrus aphid, *T. citricida*. The cDNA library was derived from alate adults due to their significance in spreading viruses (e.g., citrus tristeza virus). Over 5180 cDNA clones were sequenced, resulting in 4263 high-quality ESTs. Contig alignment of these ESTs resulted in 2124 total assembled sequences, including both contiguous sequences and singlets. Approximately 33% of the ESTs currently have no significant match in either the non-redundant protein or nucleic acid databases. Sequences returning matches with an *E*-value of ≤ -10 using BLASTX, BLASTN, or TBLASTX were annotated based on their putative molecular function and biological process using the Gene Ontology classification system. These data will aid research efforts in the identification of important genes within insects, specifically aphids and other sap feeding insects within the Order Hemiptera.

The sequence data described in this paper have been submitted to Genbank's dbEST under the following accession numbers.: CB814527-CB814982, CB832665-CB833296, CB854878-CB855147, CB909714-CB910020, CB936196-CB936346, CD449954-CD450759.

Keywords: Aphididae, cDNA, EST, Gene expression, Hemiptera, Development, Toxoptera

Abbreviation:
EST expressed sequence tag

Introduction

The brown citrus aphid, *Toxoptera citricida* (Kirkaldy), is one of the most devastating pests of citrus, causing extensive crop losses worldwide. Feeding by this aphid alone can cause severe damage to citrus. However, it poses an even greater threat to citrus because of its efficient transmission of citrus tristeza closterovirus (Fasulo and Halbert, 1993).

Since the brown citrus aphid genomic sequence is not available, expressed sequence tags (ESTs) derived from single-pass sequencing of cDNA clones prepared from the brown citrus aphid provide an invaluable resource for the identification of genes associated with the biology of the alate adult life stage. In the past, cloning of genes encoding enzymes of specific biochemical pathways by single-pass sequencing of cDNA clones has been a very successful strategy, particularly when the cDNA libraries have been prepared from tissues with high activity for the respective enzymes (Coyle-Thompson and Banerjee 1993; Newman *et al.*, 1994; Blaxter *et al.*,

1996; Cooke *et al.*, 1996; Rounsley *et al.*, 1996). This enables investigators to isolate genes derived from specific tissues and/or life stages for more detailed study, which may include developing efficient biocontrol methods.

Additionally, ESTs and their accompanying cDNAs, provide the means to construct glass or nylon based arrays that can be used for transcript profiling on a genome-wide scale (DeRisi *et al.*, 1997; Ruan *et al.*, 1998; Egger *et al.*, 2002). A careful bioinformatic analysis identifying life stage-specific ESTs is a prerequisite in order to obtain a comprehensive and representative set of cDNAs for gene expression studies by arrays (Loftus *et al.*, 1999). Given that there are only a small number of insect ESTs in public databases it was essential to build a life-stage specific library derived from aphids so that analysis of metabolism and development on a genome-wide scale could be accomplished. Even without subsequent array analysis, a relatively large number of ESTs from a specific life stage can provide clues toward the expression of specific genes important to the functions expressly connected with that life stage (Rafalski *et al.*, 1998; Arbeitman *et al.*, 2002). In most cases and within statistical limitations, the abundance of a specific cDNA in the EST collection is a measure of gene expression (Audic and Claverie, 1997). This technique, referred to as a"digital or electronic northern", has been utilized in several similar studies to gauge relative gene expression in various tissues. The data sets are available at GenBank, dbEST under the following accession numbers.: CB814527-CB814982, CB832665-CB833296, CB854878-CB855147, CB909714-CB910020, CB936196-CB936346, CD449954-CD450759.

Materials and Methods

Aphid rearing and collection

Alate brown citrus aphids, *Toxoptera citricida,* were obtained from a healthy colony maintained by WB Hunter at the USDA, ARS, U.S. Horticultural Research Laboratory, Ft. Pierce, FL. The founders were collected from a single collection site in Orlando, Florida. The colony was reared under continuous asexual reproduction for a period of 3 years on sweet orange, *Madam vinous,* seedlings in screen cages contained in an insectary, and held at 25° C, 16 L: 8 D. Plants free of insecticide and bearing new flush were cycled into cages on a weekly basis. Aphids and their host plants were surveyed biweekly for any incidence of contaminating insect species (e.g., mites, parasitoids, fungus gnats, shore flies, etc.). High-density aphid populations produced alate aphids that were collected by aspiration within two days of emergence. All alates were collected from the top of the cage so as to avoid sample contamination with other developmental forms or host plant tissue. Upon collection, alates were immediately submerged into liquid nitrogen prior to total RNA isolation. Approximately 50-100 alates were placed into 95% ethanol and stored at –80°C to be used as voucher specimens.

cDNA library construction

Approximately 4500 1-2 day old alate aphids were used in the construction of an expression library. Whole aphids were ground in liquid nitrogen and total RNA extracted using guanidinium salt-phenol-chloroform procedure as described by Strommer *et al.* (1993). Poly(A)+ RNA was purified using two rounds of selection on oligo dT magnetic beads according to the manufacturer's instructions (Dynal, www.dynal.no). A directional cDNA library was constructed in Lambda Uni-ZAP® XR Vector using Stratagene's ZAP-cDNA Synthesis Kit (Stratagene, www.stratagene.com). The resulting DNA was packaged into lambda particles using Gigapack® III Gold Packaging Extract (Stratagene). An amplified library was generated with a titer of 1.0 x 10^9 plaque-forming units per mL. Mass excision of the amplified library was carried out using Ex-Assist® helper phage (Stratagene). An aliquot of the excised, amplified library was used for infecting XL1-Blue MRF' cells and subsequently plated on LB agar containing 100 µg/mL ampicillin. Bacterial clones containing excised pBluescript SK(+) phagemids were recovered by random colony selection.

Sequencing of clones

pBluescript SK(+) phagemids were grown overnight at 37° C and 240 rpm in 96-deep well culture plates containing 1.7 mL of LB broth, supplemented with 100 µg/mL ampicillin. Archived stocks were prepared from the cell cultures using 75 µl of a LB-amp, glycerol mixture and 75 µl of cells. These archived stocks are held at the Horticultural Research Laboratory where they are kept in an ultra low temperature freezer set at –80° C. Plasmid DNA was extracted using the Qiagen 9600 liquid handling robot and the QIAprep 96 Turbo miniprep kit according to the recommended protocol (QIAGEN, www.quigen.com).

Sequencing reactions were performed using the ABI PRISM® BigDye™ Primer Cycle Sequencing Kit (Applied Biosystems, home.appliedbiosystems.com) along with a universal T3 primer. Reactions were prepared in 96-well format using the Biomek2000™ liquid handling robot (Beckman Coulter, www.beckman.com). Sequencing reaction products were precipitated with 70% isopropanol, resuspended in 15 µL sterile water and loaded onto an ABI 3700 DNA Analyzer (Applied Biosystems).

Computer analysis

Base confidence scores were designated using TraceTuner® (Paracel, www.paracel.com). Low-quality bases (confidence score <20) were trimmed from both ends of sequences. Quality trimming, vector trimming and sequence fragment alignments were executed using Sequencher® software (Gene Codes, www.genecodes.com). Contaminating sequences such as rRNA and mitochondrial DNA were identified using BLASTN and were excluded from analysis along with sequences less than 100 nucleotides in length after both vector and quality trimming. Additional ESTs that corresponded to vector contaminants were removed from the dataset. To estimate the number of genes represented in the library and the redundancy of specific genes, ESTs were assembled into contigs using Sequencher®. Contig assembly parameters that were set using a minimum overlap of 50 bases and 95% identity match.

Functional annotation of ESTs

Putative sequence identity was determined based on BLAST similarity searches using the National Center for Biotechnology Information (NCBI) BLAST server (http://www.ncbi.nlm.nih.gov) with comparisons made to both non redundant nucleic acid (BLASTN) and protein (BLASTX) databases. ESTs that had no significant similarity to any publicly available sequences using

BLASTN and BLASTX were then screened individually using TBLASTX.

The top 5 hits for each assembled sequence were then formatted using an in-house parsing program that allowed for direct import into a Microsoft Excel® spreadsheet for further analysis. Sequence matches with *E*-value scores ≤ -10 were considered significant and were categorized according to the Gene Ontology (GO) classification system based on annotation of the 5 'best hit' matches in BLASTX searches. All *D. melanogaster* matches were cataloged using FlyBase (www.flybase.org). Those sequences without a *D. melanogaster* hit were annotated using AmiGO (www.geneontology.org).

Results and Discussion

Generation and assembly of adult alate ESTs

An initial 5180 clones were sequenced from the 5' end. These sequences were trimmed of vector and low-quality sequence and filtered for minimum length (100 bp), producing 4267 high-quality ESTs of 481 bp average length. These ESTs were analyzed with the Sequencher® assembly program to identify those that represent redundant transcripts. ESTs were assembled into 468 contiguous sequences (contigs) with 1656 ESTs remaining as singlets, suggesting a 61% redundancy. Thus, the combined set of contigs and singlets included 2124 sequences (hereafter referred to as 'assembled sequences'), putatively representing different transcripts. Only 22 contig sequences contained more than 10 ESTs.

EST quality analysis and sequence survey

Of the 2124 assembled sequences analyzed, 993 (representing 2132 ESTs) were similar to known protein sequences in the non-redundant protein database (BLASTX; $E \leq -10$). Seven of these assembled sequences, representing 13 ESTs, were identified by BLASTX as contaminating vector sequences and were removed from the dataset.

Because some genes encode RNAs rather than proteins, it was necessary to run BLASTN against our dataset. Eight assembled sequences were identified as ribosomal and 2 were identified as mitochondrial DNA, representing 582 and 65 ESTs respectively, and were removed from the dataset. Although the number of ribosomal sequences appears inflated, it has been shown that several non-coding RNAs, such as rRNA, have mRNA-like modifications, such as polyadenylation and splicing. Because this EST dataset was derived from a cDNA library that was enriched for poly(A+) RNA, it is reasonable to assume that some non-coding RNAs should be present (MacIntosh *et al.*, 2001). An additional 76 ESTs were identified as either rRNA or mitochondrial using TBLASTX, leaving 2031 assembled sequences used in subsequent functional analyses.

Of the initial 2124 assembled sequences (representing 4267 ESTs), 1045 (representing 1412 ESTs) showed no significant similarity (E>-10) to any publicly available sequence using BLASTX, BLASTN, or TBLASTX. This result suggests that a large percentage (~33%) of the ESTs sequenced here are novel. However, this estimation of potential unique sequences within the cDNA library is most likely to be an overestimation due to several factors, such as computer alignment parameters and low quality internal sequences (White *et al.*, 2000). Moreover, assembled sequences may have lacked an open reading frame because they were too short causing ESTs to consist mostly or entirely of a noncoding region (e.g., 3' untranslated region) (Whitfield *et al.*, 2002).

Functional annotation of ESTs

Each *Toxoptera citricida* assembled sequence was tentatively assigned Gene Ontology classification based on annotation of the top 5 "best hit" matches ($E \leq -10$) using BLASTX. Nearly all of these were characterized with respect to the functionally annotated genes in *D. melanogaster* using FlyBase. Of the 993 sequences demonstrating similarity to known protein sequences, 332 (33%) of these were of unknown molecular function and 685 (69%) were of unknown biological process. Tables 1 and 2 summarize assignments of *Toxoptera* sequences to major molecular functions and biological processes, respectively.

Genes of interest within the EST dataset

The BLASTX results provide useful information regarding the homology of proteins that may be critical for insect cellular communication and development. Table 3 lists sequences of the brown citrus aphid that match to *D. melanogaster* genes implicated in signal transduction, cell differentiation, cell fate commitment, embryonic and larval development, morphogenesis, reproduction and cuticle biosynthesis. Typically, genes involved in early development would not be present in cDNA libraries derived from adult tissues. However, many aphid species are composed entirely of viviparous parthenogenetic females. These insects telescopic generations as embryogenesis occurs in un-born daughters, producing up to three generations developing within an adult individual (Sabater *et al.*, 2001). Therefore, genes involved in the development of several life stages may be represented simultaneously in this analysis.

For the purposes of this paper, brown citrus aphid sequences were grouped into distinct gene ontology classifications. However, it is important to recognize that many of these gene products act in concert with one another to control cell fate determination which, in turn, drives morphological changes such as eye, leg, and wing development (e.g., the *Notch* pathway) (Coyle-Thompson and Banerjee, 1993; Baonza *et al.*, 2000).

Conclusions

We have provided a large data set of ESTs from the alate brown citrus aphid and have begun to analyze this valuable resource. The analysis of this data set is continually evolving and some of the conclusions may have to be revised as more advanced bioinformatic tools become available. Being the first EST data set for the brown citrus aphid, its preliminary examination clearly shows that it is substantially different from the aphid EST data set currently available to the public. For the most part, there is considerable congruence between conventional biochemistry regarding insect metabolism and the number of ESTs encoding metabolic enzymes. This data set provides the first experimental access to these genes and the basis for more in-depth molecular and genomic analysis. Moreover, it identifies genes that are critical in the physiology, reproduction, development, and wing morphogenesis of aphids. Genetic information is crucial to advancing our understanding of aphid

Table 1. Molecular Function

Gene Ontology Term[a]	Number of ESTs	% of total ESTs represented[b]	Number of contigs	Number of singlets
[p] Antioxidant	1	0.06%	0	1
[p] Apoptosis	1	0.06%	0	1
[p] Binding				
[c] General binding	8	0.50%	2	3
[c] Calcium ion binding	16	1.00%	3	5
[c] Carbohydrate binding	2	0.12%	1	0
[c] Drug binding	10	0.62%	1	1
[c] Heavy-metal binding	10	0.62%	3	0
[c] Hormone binding	1	0.06%	0	1
[c] Isoprenoid binding	7	0.44%	1	1
[c] Lipid binding	12	0.75%	3	2
[c] Neurotransmitter binding	2	0.12%	0	2
[c] Nucleic acid binding				
[i] DNA binding	29	1.81%	4	9
[ii] Transcription factor	8	0.50%	1	6
[i] RNA binding	48	2.99%	9	18
[i] Nuclease activity	1	0.06%	0	1
[i] Transcription factor	6	0.37%	1	4
[i] Translation factor	27	1.68%	1	5
[c] Nucleotide binding	22	1.37%	3	10
[c] Odorant binding	1	0.06%	0	1
[c] Oxygen binding	14	0.87%	3	1
[c] Protein binding	60	3.74%	12	20
[c] Steroid binding	1	0.06%	0	1
[p] Cell Adhesion Molecule	1	0.06%	0	1
[p] Chaperone	26	1.62%	6	12
[p] Enzyme				
[c] Helicase	2	0.12%	0	2
[c] Histone deactylase	1	0.06%	0	1
[c] Hydrolase				
[i] General hydrolase	12	0.75%	3	6
[i] Acting on acid anhydrides	98	6.11%	21	21
[i] Acting on ester bonds	22	1.37%	4	14
[i] Acting on ether bonds	2	0.12%	0	2
[i] Acting on glycosyl bonds	1	0.06%	0	1
[i] Peptidase	51	3.18%	11	24
[c] Isomerase	15	0.94%	5	4
[c] Kinase	23	1.43%	5	11
[c] Ligase	19	1.19%	5	5
[c] Lyase	27	1.68%	6	9
[c] Molybdopterin cofactor sulfarase	1	0.06%	0	1
[c] Oxidoreductase	283	17.65%	31	51
[c] Small protein activating enzyme	1	0.06%	0	1
[c] Small protein conjugating enzyme	5	0.31%	1	3
[c] Transferase	66	4.12%	10	36
[p] Enzyme regulator	21	1.31%	3	11
[p] Protein degradation tagging	18	1.12%	4	8
[p] Signal transducer	13	0.81%	2	7
[p] Structural molecule				
[c] Cuticular protein	84	5.24%	14	6
[c] Cytoskeleton protein	42	2.62%	3	2
[c] Muscle fiber	2	0.12%	1	0
[c] Ribosomal protein	337	21.02%	65	29
[p] Transcription regulation	16	1.00%	3	10
[p] Translation regulation	13	0.81%	5	1
[p] Transporter	114	7.11%	10	23
Totals	1603		266	395

[a]Classification is hierarchial: idented terms are children [c] of parent terms [p] listed above. All functional assignments of *Toxoptera citricida* ESTs described here are the "inferred from electronic evidence" (IEA) using the top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. The definition term associated with each sequence was entered into both FlyBase and AmiGO where the it was given a molecular function according to The Gene Ontology Consortium.

[b]% of total ESTs represented was calculated using only those ESTs with a BLASTX hit with an *E*-value of ≤-10 and of known protein function.

Table 2. Biological Process

Gene Ontology Term[a]	Number of ESTs	% of total ESTs represented[b]	Number of contigs	Number of singlets
[p] Behavior	7	0.81%	1	4
[p] Cell Communication				
[c] Cell adhesion	2	0.23%	0	2
[c] Cell-cell signaling	3	0.35%	0	3
[c] Response to external stimulus	19	2.20%	5	7
[c] Signal transduction	8	0.93%	0	8
[p] Cell Growth and/or maintenance				
[c] General cell growh and/or maintenance	1	0.12%	0	1
[c] Cell cycle	12	1.39%	0	12
[c] Cell motility	10	1.16%	2	0
[c] Cell organization and biogenesis	60	6.94%	6	13
[c] Homeostasis	16	1.85%	4	2
[c] Membrane fusion	1	0.12%	0	1
[c] Metabolism				
[i] Alcohol metabolism	18	2.08%	7	0
[i] Amine metabolism	2	0.23%	1	0
[i] Aromatic compound metabolism	3	0.35%	1	1
[i] Biosynthesis	18	2.08%	4	7
[i] Carbohydrate metabolism	4	0.46%	0	4
[i] Catabolism	4	0.46%	1	0
[i] Energy pathways	2	0.23%	1	0
[i] Lipid metabolism	1	0.12%	0	1
[i] Nucleobase, nucleoside, nucleotide and nucleic acid metabolism	36	4.17%	5	26
[i] Phosporous metabolism	126	14.58%	4	11
[i] Protein metabolism	393	45.49%	76	31
[i] Sulfur metabolism	2	0.23%	0	2
[c] Response to stress	1	0.12%	0	1
[c] Transport	56	6.48%	13	11
[p] Development				
[c] Cell differentiation	8	0.93%	2	0
[c] Cell fate commitment	1	0.12%	0	1
[c] Embryonic development	2	0.23%	0	2
[c] Larval development	2	0.23%	1	0
[c] Morphogenesis				
[i] Morphogenesis of an epithelium	1	0.12%	0	1
[i] Organogenesis				
[ii] Histogenesis				
[iii] Ectoderm development	2	0.23%	1	0
[ii] Imaginal disc development	2	0.23%	0	2
[ii] Muscle development	7	0.81%	3	0
[ii] Neurogenesis	23	2.66%	4	4
[ii] Trachael system development	1	0.12%	0	1
[c] Reproduction	9	1.04%	1	5
[p] Physiological processes				
[c] Cuticle biosynthesis	1	0.12%	0	1
Totals	864		143	165

[a]Classification is hierarchial: idented terms are children [c] of parent terms [p] listed above. All functional assignments of *Toxoptera citricida* ESTs described here are the "inferred from electronic evidence" (IEA) using the top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. The definition term associated with each sequence was entered into both FlyBase and AmiGO where the it was given a molecular function according to The Gene Ontology Consortium.

[b]% of total ESTs represented was calculated using only those ESTs with a BLASTX hit with an *E*-value of ≤ -10 and of known protein function.

biology, and will play a major role in the development of future non-chemical, gene-based control strategies against these insect pests.

Acknowledgements

We thank J. Mozoruk for helpful discussions and review of the manuscript. Special thanks to L.E. Hunnicutt for data analysis, annotation of data, and assistance in manuscript construction and review.

References

Arbeitman MN, Furlong EEM, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* 297: 2270-2275.

Audic S, Claverie JM. 1997. The significance of digital gene expression profiles. *Genome Research* 7: 986-995.

Blaxter ML, Raghavan N, Ghosh I, Guiliano D, Lu W, Williams SA, Slatko B, Scott AL. 1996. Genes expressed in *Brugia malayi* infective third stage larvae. *Molecular and Biochemical Parasitology* 77: 77-93.

Baonza A, de Celis JF, García-Bellido A. 2000. Relationships between *extramacrochaetae* and *Notch* signaling in *Drosophila* wing development. *Development* 127: 2383-2393.

Cooke R, Raynal M, Laudie M, Grellet F, Delseny M, Morris PC,

Table 3. Genes of interest in the Alate BrCA EST dataset

Gene Ontology[a]	Sequence Identifier	Accession Number[b]	NCBI Descriptor	Source Organism	*E*- value
[p] Cellular Communication					
[c] Signal transduction	WHWTC-30_G08	NP_476884	14-3-3zeta; D14 3 3 protein; leonardo	*Drosophila melanogaster*	6.00E-78
	WHWTC-44_E04	NP_511144	strawberry notch; glossy-like	*Drosophila melanogaster*	5.00E-55
	WHWTC-50_D08	AAF57785	CG5036	*Drosophila melanogaster*	2.00E-52
	WHWTC-51_F11	AAF54188	CG7918	*Drosophila melanogaster*	9.00E-44
	WHWTC-52_H11	NP_477130	corkscrew CG3954	*Drosophila melanogaster*	1.00E-37
	WHWTC-27_F03	O43541	Mothers against decapentaplegic homolog 6	*Homo sapiens*	2.00E-19
	WHWTC-29_H02	AAF56349	CG7012 gene product	*Drosophila melanogaster*	2.00E-19
	WHWTC-49_B08	NP_511064	deltex	*Drosophila melanogaster*	5.00E-15
[p] Development					
[i] Cell differentiation	ContigWTC[0908]	NP_477122	Muscle LIM protein at 84B CG1019	*Drosophila melanogaster*	9.00E-50
	ContigWTC[0147]	AAF58567	guf gene product [alt 1]	*Drosophila melanogaster*	4.00E-17
[i] Cell fate commitment	WHWTC-09_D08	NP_523455	anterior open CG3166	*Drosophila melanogaster*	9.00E-51
[i] Embryonic development	WHWTC-26_C07	2115375A	snr1 gene	*Drosophila melanogaster*	2.00E-38
	WHWTC-01_F03	AAF49547	CG5891 gene product	*Drosophila melanogaster*	5.00E-15
[i] Larval development	ContigWTC[0852]	AAF53765	CG10691	*Drosophila melanogaster*	2.00E-72
[i] Morphogenesis					
[ii] Morphogenesis of an epithelium	WHWTC-44_E08	NP_477342	discs lost CG12021	*Drosophila melanogaster*	3.00E-27
[ii] Organogenesis					
[iii] Histogenesis					
[iv] Ectoderm development	ContigWTC[0527]	AAF50606	CG8624	*Drosophila melanogaster*	3.00E-12
[iii] Imaginal disc development	WHWTC-39_G11	NP_477444	COP9 complex homolog subunit 4 CG8725	*Drosophila melanogaster*	3.00E-10
	WHWTC-41_C08	A56922	transcription factor shn	*Drosophila melanogaster*	4.00E-22
[iii] Muscle development	ContigWTC[1198]	NP_477098	CG8416	*Drosophila melanogaster*	3.00E-98
	ContigWTC[0705]	A38594	troponin I	*Drosophila melanogaster*	6.00E-53
	ContigWTC[1037]	AAF47158	Mlp60A gene product	*Drosophila melanogaster*	6.00E-24
[iii] Neurogenesis	WHWTC-53_E07	AAL76026	putative calreticulin	*Aedes aegypti*	5.00E-75
	ContigWTC[0118]	NP_523792	FK506-binding protein 2 CG11001	*Drosophila melanogaster*	4.00E-40
	WHWTC-28_C11	AAD03559	failed axon connections protein	*Drosophila virilis*	4.00E-39
	ContigWTC[0174]	P58375	60S ribosomal protein L30	*Spodoptera frugiperda*	3.00E-35
	ContigWTC[1050]	AAF45520	CG7727	*Drosophila melanogaster*	4.00E-33
	WHWTC-42_F10	AAD43793	CDC42 protein	*Drosophila melanogaster*	4.00E-28
	ContigWTC[0008]	AAF47413	CG1007	*Drosophila melanogaster*	1.00E-12
	WHWTC-52_E04	AAB60619	neuralized protein	*Drosophila virilis*	3.00E-11
[iii] Trachael system development	WHWTC-24_D12	AAF50772	CG10624	*Drosophila melanogaster*	7.00E-12
[i] Reproduction	WHWTC-26_F08	XP_079633	CG5395 gene product	*Drosophila melanogaster*	5.00E-60
	WHWTC-03_C02	AAB34841	syntaxin 1, Dsynt1	*Drosophila sp.*	3.00E-39
	WHWTC-51_A03	AAF49765	CG6451	*Drosophila melanogaster*	2.00E-26
	WHWTC-33_F11	NP_477016	chickadee CG9553	*Drosophila melanogaster*	3.00E-24
	ContigWTC[0744]	NP_477016	chickadee CG9553	*Drosophila melanogaster*	5.00E-18
	WHWTC-51_H09	AAF56175	CG10367	*Drosophila melanogaster*	2.00E-11
[p] Physiological processes					
[c] Cuticle biosynthesis	WHWTC-04_G08	CAC34734	Yellow protein	*Drosophila ananassae*	3.00E-14

[a]All functional assignments of *Toxoptera citricida* ESTs described here are the "inferred from electronic evidence" (IEA) using the top 5 BLASTX hits with an *E*-value of ≤-10 generated from NCBI's nr database. The definition term associated with each sequence was entered into both FlyBase and AmiGO where the it was given a molecular function according to The Gene Ontology Consortium.
[b]Accession numbers correspond to the "best hit" match to Genbank's nr protein database using BLASTX.

Guerrier D, Giraudat J, Quigley F, Clabault G, Li YF, Mache R, Krivitzky M, Gy IJ, Kreis M, Lecharny A, Parmentier Y, Marbach J, Fleck J, Clement B, Phillips G, Herve C, Bardet C, Tremousaygue D, Hofte H. 1996. Further progress towards a catalogue of all *Arabidopsis* genes: Analysis of a set of 5,000 non-redundant ESTs. *Plant Journal* 9: 101-124.

Coyle-Thompson CA, Banerjee U. 1993. The *strawberry notch* gene functions with *Notch* in common developmental pathways. *Development* 119: 377-395.

DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680-686.

Egger B, Leemans R, Loop R, Kammermeier L, Fan Y, Radimerski T, Strahm MC, Certa U, Reichert H. 2002. Gliogenesis in *Drosophila*: Genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. *Development* 129: 3295-3309.

Fasulo TR, Halbert SE. 1993. Aphid Pests of Citrus. ENY-811, *Entomology and Nematology Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida.*

Loftus SK, Chen Y, Gooden G, Ryan JF, Birznieks G, Hilliard M, Baxevanis AD, Bittner M, Meltzer P, Trent J, Pavan W. 1999. Informatic selection of a neural crestmelanocyte cDNA set for microarray analysis. *Proceeding of the National Academy of Sciences USA* 96: 9277-9280.

MacIntosh GC, Wilkerson C, Green PJ. 2001. Identification and Analysis of Arabidopsis Expressed Sequence Tags Characteristic on Non-Coding RNAs. *Plant Physiology* 127: 765-776.

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Sommerville S, Thomashow M. 1994. Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiology* 106: 1241-

1255.

Rafalski JA, Hanafey M, Miao GH, Ching A, Lee JM, Dolan M, Tingey S. 1998. New experimental and computational approaches to the analysis of gene expression. *Acta Biochimica Polonica*. 45: 929-934.

Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JC, Kerlavage AR. 1996. The construction of *Arabidopsis* expressed sequence tag assemblies: A new resource to facilitate gene identification. *Plant Physiology* 112: 1177-1183.

Ruan Y, Gilmore J, Conner T. 1998. Towards *Arabidopsis* genome analysis: Monitoring expression profiles of 1,400 genes using cDNA microarrays. *Plant Journal* 15: 821-833.

Sabater B, van Ham RCHJ, Martínez-Torres D, Silva F, Latorre A, Moya A. 2001. Molecular evolution of aphids and their primary (*Buchnera* sp.) and secondary endosymbionts: Implications for the role of symbiosis in insect evolution. *Interciencia* 26 (10): 508 – 512.

Strommer JN, Gregerson R, Vayda M. 1993. Isolation and characterization of plant mRNA. *Methods in Plant Molecular Biology and Biotechnology* 49-65.

White JA, Todd J, Newman T, Focks N, Girke T, de Ilarduya OM, Jaworski JG, Ohlrogge JB, Benning C. 2000. A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiology* 124: 1582-1594.

Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE. 2002. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Research* 12: 555-566.