INFECTION AND IMMUNITY, July 2004, p. 4114–4126 Vol. 72, No. 7
0019-9567/04/$08.00+0 DOI: 10.1128/IAI.72.7.4114–4126.2004

Description of the Transcriptomes of Immune Response-Activated Hemocytes from the Mosquito Vectors *Aedes aegypti* and *Armigeres subalbatus*

Lyric C. Bartholomay,[1]† Wen-Long Cho,[2]† Thomas A. Rocheleau,[1] Jon P. Boyle,[1] Eric T. Beck,[1] Jeremy F. Fuchs,[1] Paul Liss,[1] Michael Rusch,[1] Katherine M. Butler,[1] Roy Chen-Chih Wu,[2] Shih-Pei Lin,[2] Hang-Yen Kuo,[2] I-Yu Tsao,[2] Chiung-Yin Huang,[2] Tze-Tze Liu,[3] Kwang-Jen Hsiao,[3] Shih-Feng Tsai,[4] Ueng-Cheng Yang,[5] Anthony J. Nappi,[1] Nicole T. Perna,[1] Chen-Cheng Chen,[2]* and Bruce M. Christensen[1]*

Department of Animal Health and Biomedical Sciences, University of Wisconsin—Madison, Madison, Wisconsin 53706,[1] and Department of Parasitology,[2] Genome Research Center and Institute of Genetics,[3] and Institute of Bioinformatics,[5] National Yang-Ming University, Shih-Pai, Taipei 112, and National Health Research Institutes, Nankang, Taipei 115,[4] Taiwan

Received 27 January 2004/Returned for modification 5 March 2004/Accepted 21 March 2004

Mosquito-borne diseases, including dengue, malaria, and lymphatic filariasis, exact a devastating toll on global health and economics, killing or debilitating millions every year (54). Mosquito innate immune responses are at the forefront of concerted research efforts aimed at defining potential target genes that could be manipulated to engineer pathogen resistance in vector populations. We aimed to describe the pivotal role that circulating blood cells (called hemocytes) play in immunity by generating a total of 11,952 *Aedes aegypti* and 12,790 *Armigeres subalbatus* expressed sequence tag (EST) sequences from immune response-activated hemocyte libraries. These ESTs collapsed into 2,686 and 2,107 EST clusters, respectively. The clusters were used to adapt the web-based interface for annotating bacterial genomes called A Systematic Annotation Package for Community Analysis of Genomes (ASAP) for analysis of ESTs. Each cluster was categorically characterized and annotated in ASAP based on sequence similarity to five sequence databases. The sequence data and annotations can be viewed in ASAP at https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm. The data presented here represent the results of the first high-throughput in vivo analysis of the transcriptome of immunocytes from an invertebrate. Among the sequences are those for numerous immunity-related genes, many of which parallel those employed in vertebrate innate immunity, that have never been described for these mosquitoes.

The maintenance of mosquito-borne disease transmission cycles is dependent upon the compatibility, as dictated by the genomes, of both the pathogen and the host. The decreasing efficacy of traditional methods for vector-borne disease control has provided the impetus to explore host-pathogen interactions at the molecular level with the aim of designing novel control methods (13). Innate immune responsiveness is of particular interest in such explorations, because extensive research efforts have demonstrated that vector mosquito species and mosquito cell lines produce robust humoral and cellular immune responses against invading pathogens.

Experimental evidence for a pivotal role of mosquito hemocytes (blood cells) as initiators and mediators of mosquito innate immune responses has recently become available. But because mosquito hemocytes are difficult to collect from mosquitoes and cannot be cultured or even maintained for significant amounts of time in vitro, they have received little research attention. Hemocytes phagocytose and melanize large numbers of bacteria, fungi, and malaria parasites (15, 24, 25, 28, 29). Several subpopulations of mosquito hemocytes have been characterized, and immunocytochemical and cytochemical assays have demonstrated that these cells are differentially responsive to invading pathogens (24, 27). For example, cell types are distinguished by ultrastructural characteristics, phagocytic capacity, or the presence and activity of enzymes involved in the melanin biosynthetic pathway (24, 28, 29, 36). It is hypothesized that hemocytes also participate in pattern recognition and mediate the production of immune peptides from the fat body (6).

In conjunction with the structural characterization of hemocytes done in our laboratory, we undertook a high-throughput molecular approach to gene discovery and characterization with the aim of gaining insight into vector-pathogen interactions mediated by hemocytes. Directional cDNA libraries and expressed sequence tags (ESTs) were generated from the perfusate (which contains the hemocytes) of the mosquitoes *Aedes aegypti* and *Armigeres subalbatus*. *Aedes aegypti* is a well-studied

* Corresponding author. Mailing address for Bruce M. Christensen: Department of Animal Health and Biomedical Sciences, University of Wisconsin—Madison, 1656 Linden Dr., Madison, WI 53706. Phone: (608) 262-3850. Fax: (608) 262-7420. E-mail: christensen@svm.vetmed.wisc.edu. Mailing address for Chen-Cheng Chen: Department of Parasitology, National Yang-Ming University, Shih-Pai, Taipei 112, Taiwan. Phone: 886-2-2821-3892. Fax: 886-2-2821-4670. E-mail: mosquito@ym.edu.tw.

† L.C.B. and W.-L.C. contributed equally to the present work.

model mosquito for studies of *Plasmodium* and filarial worm vector competence, as well as a principal vector for yellow fever and dengue viruses. A genome sequencing project has been proposed for this species (37). *Armigeres subalbatus* is a model organism for studies of inherent resistance to filarial worms, which cause lymphatic filariasis in 120 million people annually (7). Material was collected from mosquitoes that were immune response activated by inoculation with the bacteria *Escherichia coli* and *Micrococcus luteus*. These species of bacteria have been used extensively to study insect innate immunity, particularly immune peptide production (35, 41). It has been shown that each bacterium elicits a distinct effector arm of the mosquito innate immune response; following inoculation of either *Aedes aegypti* or *Armigeres subalbatus*, the primary response to *E. coli* is phagocytosis, and that to *M. luteus* is melanization (28, 29). Most recently, it was shown that melanization and/or phagocytosis is elicited by a variety of bacterial species, independent of the Gram type (30).

Sequences from each mosquito species were assembled into contigs to reduce redundancy. Based on sequence similarity, consensus sequences for EST clusters were annotated using a controlled vocabulary in the web-based interface called A Systematic Annotation Package for Community Analysis of Genomes (ASAP) (21). Each cluster set was compared to the *Drosophila melanogaster* transcriptome and proteome sequences, the GenPept database, predicted peptides from the *Anopheles gambiae* genome sequence, the Pfam protein families database, and the other cluster set by use of BLAST analyses. Annotators manually analyzed these data and, whenever possible, attached information to the sequences, including a predicted gene product, protein domains, and indication of biological function, by use of a controlled vocabulary. Each entry is supported by the evidence used to make the annotation and typically includes intact hyperlinks to relevant databases. Consensus sequences from EST clusters and their annotations were submitted to GenBank. However, the data in ASAP are subject to change because the ASAP platform readily supports the addition of new sequence data, sequence comparisons, and experimental data. To view and download sequences and annotations, readers are invited to enter the ASAP website at https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm.

EST clusters from mosquito hemocytes not only represent diverse cellular processes but also include numerous gene products related to immunity, many of which parallel those employed in vertebrate defense responses. These data constitute the first high-throughput, comprehensive analysis of immunity genes from these two mosquito species and are supplemented with corresponding information from the *Anopheles gambiae* genome sequencing project. The sequences and their annotations should prove to be an important resource for comparative genomic studies of mosquito species and for the prediction of gene products from and annotation of mosquito genome sequences. Additionally, because of the source of the library, these data provide valuable insight into potential transcript localization and abundance for the represented gene products.

MATERIALS AND METHODS

Mosquito maintenance. *Aedes aegypti* (Liverpool-strain) larvae were hatched from dry oviposition papers in deoxygenated, deionized water. *Armigeres subalbatus* larvae were hatched in oviposition dishes and transferred to enamel rearing pans. Approximately 300 *Aedes aegypti* or 250 *Armigeres subalbatus* larvae per pan were placed in deionized water. Female pupae were mechanically separated, and 100 *Aedes aegypti* or 80 *Armigeres subalbatus* pupae were placed into 0.473-liter ice cream cartons covered with fine-mesh marquisette. Adults were maintained as previously described for *Culex pipiens* adults and used for experimentation within 3 days of eclosion (3).

Bacterial inoculations. *E. coli* K-12 and *M. luteus* were used for bacterial inoculations as described previously (42). Cold-immobilized mosquitoes were held in place with a vacuum saddle, and a stainless steel probe that had been dipped into a bacterial pellet was inserted into the cervical membrane. The mosquitoes were returned to the ice cream cartons and placed in our insectary.

Tissue collection. A volume displacement technique was used to collect hemocytes. A tear was made above the penultimate abdominal segment of the cold-immobilized mosquitoes, which were then placed on a vacuum saddle. A pulled glass capillary needle, attached to a syringe containing an RNA extraction solution, was inserted into the cervical membrane between the head and the thorax. The RNA extraction solution was injected, and the perfusate was immediately placed on dry ice and then stored at −80°C. For each mosquito, the perfusion takes only seconds. Hemocyte material was collected within minutes (<1 h) and at 3, 6, 12, and 24 h postinoculation.

cDNA library construction and EST production. RNA extractions were performed by the single-step acid guanidinium thiocyanate-phenol-chloroform extraction method (11). A directional cDNA library was generated from each mosquito species. RNA was electrophoresed to confirm its integrity, and then RNA from all time points was pooled. Total RNA was used as the template for constructing libraries by use of the Long-Distance PCR protocol of the SMART cDNA library construction kit (Clontech, Palo Alto, Calif.). Total RNA, as opposed to poly(A)-selected RNA, was used for the library construction because initial attempts to isolate selected RNA were unsuccessful, even when perfusate from approximately 500 mosquitoes was used.

Visual assessments of the quality of the libraries demonstrated that more than 80% of the phage out of 4×10^6 to 5×10^6 independent clones were recombinant. At the University of Wisconsin—Madison, sequences from recombinant plaques were generated from the resulting primary library as previously described (3). Sequencing efforts were primarily directed toward randomly picked clones, with the exception of a single batch of sequences generated at the University of Wisconsin—Madison. These clones were selected based on size such that approximately half of the sequences were more than 500 bp in length and half were more than 1,000 bp in length. PCR products in the desired size range, electrophoresed and stained with ethidium bromide, were chosen for sequencing. At National Yang-Ming University, primary and secondary libraries were first converted from phage to plasmids by use of the excision protocol provided by the manufacturer (Clontech).

Sequence compilation. Raw sequences obtained from the University of Wisconsin—Madison included sequences from the vector and primer used in library construction. Those obtained from National Yang-Ming University were subjected to Phred analysis for base calling, and then the vector sequence was trimmed by using Vector Strips (EMBOSS). The sequences were assembled to reduce redundancy by using Seqman (DNASTAR Inc., Madison, Wis.) set to trim sequence ends, scan for pTriplEx vector, and optimize sequence assembly order. A minimum match length of 18 nucleotides was used in vector-searching parameters. End trimming was executed by using the quality low-stringency parameters. All other assembly parameters were default settings. Low-stringency end-trimming parameters were used in order to keep sequences of lower quality in the assembly because these sequences could yield hits in BLAST analyses. Because of the parameters used, each contig was manually examined following assembly. Any vector sequence not recognized by the defined parameters was removed. Contigs containing what appeared to be different genes were manually split into appropriate separate contigs. Contigs containing fewer than 50 bp were removed.

Sequence similarity searches. In order to predict the gene products and associated biological functions represented in mosquito hemocyte ESTs, sequences were compared to those of the GenBank nonredundant database, the *D. melanogaster* and *Anopheles gambiae* genomes, and the reciprocal mosquito EST cluster sets created during this project. A FASTA file generated with Seqman (DNASTAR Inc.) was used for BLAST analyses. Arguments were used to produce the output in HTML format, and an expectation (*e*) value threshold was set at 0.01. Nucleotide sequences were translated and compared to the most recent version of the GenPept protein database (available at ftp://ftp.ncbi.nih.gov/GenBank/) by blastx analysis using stand-alone BLAST software available at ftp://ftp.ncbi.nih.gov/BLAST/executables/. Both a standard BLAST (blastn) and a translated BLAST (blastx) were used to compare our sequences to the *D.*

melanogaster genome (release 3.1) available through FlyBase at ftp://flybase.org/flybase-data/ (19). A blastx analysis also was done for 16,112 known and novel predicted peptides from *Anopheles gambiae* downloaded from Ensembl at ftp://ftp.ensembl.org/pub/current_mosquito/data/ (version 14.2, updated on 2 June 2003). Lastly, reciprocal BLAST (tblastx) analyses were done (with the BLAST parameters described above) for the two mosquito species to define orthologous sequences (i.e., for each species, a database was created against which sequences from the opposing species were compared). In order to qualify as orthologs, similarity between sequences had to be observed in reciprocal BLAST analyses and then those similar sequences had to have equivalent hits in the GenPept and *Drosophila* databases. Hyperlinks to orthologous sequences were then built into the database.

To obtain information about protein motifs, domains, and families, clusters were compared to the Pfam database of alignments and hidden Markov models (5). Sequences were translated in six reading frames by using a program written in Perl code. The translations were subjected to analysis against the Pfam_ls (version 9.0) library global alignment models (downloaded from http://pfam.wustl.edu) by using profile hidden Markov models software (hmmer-2.2g.bin.dos-cygwin) from the HMMER site at Washington University (http://hmmer.wustl.edu/). Annotations were made, by use of a controlled vocabulary, for Pfam hits with positive scores and expectation (*E*) values of less than 1.

Sequence annotation. Each EST cluster and its corresponding sequence similarity data were uploaded into ASAP, a relational database and web interface that was designed to facilitate the annotation of enterobacterial genomes. By using this platform, a community of annotators analyzed and attached information to the sequence data; that information was reviewed by a curator before being made available for public viewing. During the course of this project, ASAP has been adapted to accommodate EST data. Each cluster is identified by two unique numbers for two feature types: "EST_cluster" and "source." The source feature contains qualifiers that define the origin of the sequence, such as organism and library information. The EST_cluster feature type encompasses the sequence data, multiple sequence similarity searches, and annotations. Under the EST_cluster category in ASAP, annotators assessed sequence alignments and followed intact hyperlinks to the National Center for Biotechnology Information, the Wellcome Trust Sanger Institute (Ensembl and Pfam), FlyBase, and orthologous sequences within ASAP. For each cluster sequence, a predicted gene product and/or function was noted in the ASAP annotation tables. Supporting evidence for each annotation, including products and notes, is typically in the form of a hyperlink to a database. Annotations can be viewed in ASAP by following links to sequences for the mosquito species of interest that lead to the "Query Genome Annotations" page. Within this interface, users can search for genes of interest by entering an ASAP feature identification number (as listed in Table 5, for example) or by querying with a keyword.

Within ASAP, products were defined by using the "/product" qualifier, and additional information was annotated within "/note" qualifiers that are compatible with legal GenBank vocabulary. Whenever possible, products were annotated according to *D. melanogaster* (FlyBase) because the genome is well characterized and annotated. Products with homology to sequences in other databases that have no attributed function were annotated as "conserved unknown." Those that did not yield convincing hits in any database were annotated as "unknown." Annotations were preceded by additional qualifiers to categorize the quality of BLAST alignments to existing described gene products. The "questionable:" qualifier was used if the alignment was poor (i.e., low score, high *e* value, or short alignment), or "putative:" was used if the annotated product was likely according to the alignment (i.e., score close to 100 or low *e* value). ESTs without a preceding qualifier aligned very well to the sequence referred to in the annotation (i.e., high score, low *e* value, or alignment along the majority of the sequence length).

A controlled vocabulary was used to insert additional information related to the biological function of proposed gene products within the "/note" category. Broad categories of interest were defined, some of which reflect previously published mosquito genomics information, including cytoskeletal, antioxidant-related, stress response, detoxification, apoptosis-related, nuclear regulation, signal transduction, iron metabolism, and immunity-related categories. Within the immunity-related category, clusters were further described as being serine proteases or serpins or antimicrobial peptides or as being related to melanization, pattern recognition, antioxidant formation, or signal transduction; any of these can be used as a keyword search term in ASAP.

Pfam annotations were also added as notes for EST clusters with hits to the Pfam database that scored greater than 1 and had *E* values of less than 0.01. In order to provide preliminary information about the abundance of transcripts for a specific product, a note was added to describe the number of sequences that were assembled to produce each EST cluster, and an additional note ("/homolog") was created to support the annotation of sequences with homology (threshold set as the sum of identical and positive residues equal to approximately 120 or more) to *Anopheles gambiae*. This feature appears in the GenBank flat files as a note with the corresponding Ensembl identifier embedded in its content.

Sequence annotations were ongoing while we continued to collect the additional sequences that make up version 2 of our mosquito databases. ASAP provides a mechanism for tracking features across versions of a sequencing project. Should we choose to collect more EST data, ASAP provides the facility to integrate new data with existing data while tracking the history of features in all versions. We have already found it necessary to continuously revise our annotations to address the lack of this functionality in other database resources for ongoing sequencing projects.

Nucleotide sequence accession numbers. The sequences and annotations presented in this paper have been submitted to GenBank under accession numbers AY431103 to AY433788 (*Aedes aegypti*) and AY439334 toAY441440 (*Armigeres subalbatus*). More detailed information is available at https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm.

TABLE 1. Summary of EST assembly data from the mosquito species *A. aegypti* and *A. subalbatus*

Species	Total no. of sequences read	Total no. of sequences included in assemblies	Total no. of EST clusters assembled	Avg. cluster length after trimming (bp)
A. aegypti	11,952	10,373	2,686	575
A. subalbatus	12,790	11,296	2,107	539

RESULTS AND DISCUSSION

Sequence compilation. Raw sequences, generated from multiple rounds of sequencing, were assembled to construct EST clusters that were uploaded into ASAP. A total of 11,952 *Aedes aegypti* and 12,790 *Armigeres subalbatus* sequences yielded 2,686 and 2,107 EST clusters, respectively. Of the clusters, 66% of those from *Aedes aegypti* and 62% of those from *Armigeres subalbatus* are composed of a single sequence (singletons). The hemocyte libraries were not normalized, so this yield is not surprising. Those clusters that are not singletons are composed of an average of 9.4 *Aedes aegypti* or 12.3 *Armigeres subalbatus* sequences; because these clusters may reflect transcriptionally up-regulated genes, the number of ESTs that were assembled in each cluster has been incorporated into the annotations and is presented here for genes of interest. Summaries of sequence assembly results and cluster composition data are presented in Tables 1 and 2.

Sequence annotation. EST clusters were compared to the *D. melanogaster* transcriptome and proteome sequences, the GenPept database, predicted peptides from the *Anopheles gambiae* genome sequence, and to one another via BLAST analyses. The number of sequences that yielded hits in these databases is summarized in Table 3, as is the number of products annotated according to those databases. A brief summary of the

TABLE 2. Composition of EST clusters from the mosquito species *A. aegypti* and *A. subalbatus*

Species	No. of clusters (% of total no. of clusters) with the indicated no. of sequences per cluster:					
	1	2–10	11–20	21–50	51–100	>100
A. aegypti	1,775 (66)	762 (28)	75 (3)	47 (2)	16 (0.6)	11 (0.4)
A. subalbatus	1,297 (61.5)	632 (30)	71 (3.4)	67 (3.2)	24 (1.1)	16 (0.8)

TABLE 3. Results from multiple sequence similarity searches of *A. aegypti* and *A. subalbatus* EST clusters against five publicly available databases and number of annotations made using evidence from each database

Database	Search program	No. of hits yielded (% of total no. of EST clusters)		No. of annotations made using evidence from five databases	
		A. aegypti	*A. subalbatus*	*A. aegypti*	*A. subalbatus*
D. melanogaster peptide (version 3.1)	blastx	1,537 (57)	1,318 (63)	1,346	1,084
D. melanogaster RNA (version 3.1)	blastn	1,076 (40)	864 (41)	28	14
Genpept (version 135)	blastx	1,615 (60)	1,365 (65)	590	522
Pfam_ls (version 9.0)	HMMER	703 (26)	639 (30)	716	639
A. gambiae peptides (version 14.2)	blastx	1,557 (58)	1,332 (63)	1,534	1,321
A. aegypti, *A. subalbatus*	tblastx	484 (18)	484 (23)		

results from the sequence annotations appears in Table 4. Although the libraries were constructed from total RNA, fewer than 10% of the EST clusters encode ribosomal genes or gene products. Over 50% of the sequences from both mosquito species were annotated as "unknown" or "conserved unknown," meaning that they have no similarity to sequences in other databases (unknown) or are similar to those of gene products with unknown function from other organisms.

In the interest of brevity, we have restricted the discussion and tables in this paper primarily to clusters that were annotated with confidence to known gene products, with a particular emphasis on immunity-related products. The evolutionary conserved nature of innate immune responsiveness to parasites makes this information pertinent to the communities of research groups interested in the biology of vector-pathogen-vertebrate host relationships (2, 31). The text will not be replete with references. Readers interested in particular EST clusters are encouraged to visit ASAP at https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm, where the evidence for each annotation provides hyperlinks to GenBank, FlyBase, Ensembl, and the Pfam database and, therefore, links to relevant literature.

The libraries from which the ESTs were generated probably contain clones from three sources: hemocytes, the bacteria used to activate the mosquitoes' immune responses, and the fat body. Although the source libraries were not generated from poly(A)-selected RNA, only four *Aedes aegypti* clusters (composed of 33 sequences) and two *Armigeres subalbatus* clusters (composed of 5 sequences) were annotated as bacterial contaminants, perhaps owing to the fact that these libraries were not normalized and therefore did not select for less abundant sequences. Fat body contamination is inherent in the perfusion process employed to collect hemocytes (27). The inevitable presence of ESTs from the fat body in these data sets does not diminish their value, because this tissue is essential for immune responsiveness (26, 52). Molecules that are likely produced in the fat body are noted in the text.

TABLE 4. Sequence annotations in general categories of interest

Sequence category	No. of annotations per category (% of total no. of EST clusters)	
	A. aegypti	*A. subalbatus*
Immunity-related genes	169 (6)	103 (4.8)
Homologs in *A. gambiae*	980 (36)	776 (37)
Ribosomal	115 (4)	137 (7)
Unknown	981 (36)	711 (34)
Conserved unknown	598 (22)	429 (20)

Immunity-related gene products. Hemocytes function in pattern recognition, phagocytosis, melanization, and signaling cascades that initiate varied cytotoxic effector responses, and they are the central element mediating systemic mosquito innate immune responses (6). As a framework for examining hemocyte sequences, we defined the following as subcategories of immunity relatedness: general (no subcategory), pattern recognition molecules, Toll signaling pathway related, antimicrobial peptides (AMPs), melanization related, antioxidant related, serine proteases, and serpins, all of which are summarized in Table 5. Using these criteria, 169 *Aedes aegypti* EST clusters were identified (based on sequence similarity alone) as immunity related, as were 103 from *Armigeres subalbatus*. These numbers do not reflect a differential analysis of transcription between mosquito species. This is, however, a substantial number of genes, considering that only 38 were identified in a previous study of 2,380 clone clusters from a normalized *Anopheles gambiae* hemocyte-like cell line library (16), and it underscores the importance of in vivo evaluations of these cells. In comparison, the genomes of *Anopheles gambiae* and *D. melanogaster* contain 242 and 185 immunity-related genes, respectively, from 18 families (12). Of those families, 16 are represented in this study (and are indicated in Table 5), with only STATs and galectins not being represented.

Abundant sequences. Eleven EST clusters in *Aedes aegypti* and 16 in *Armigeres subalbatus* were categorized as abundant because they are composed of more than 100 sequences (Table 5). Transcriptional up-regulation of these genes will have to be confirmed in further assays; however, transcript abundance alone is a valuable and important measure used to identify genes of interest in EST projects (see reference 43). We speculate that the sequences in this category are illustrative of the concerted investment that hemocytes make in immune responsiveness, employing multiple effector arms to combat pathogens, including phagocytosis via cytoskeletal rearrangements, apoptosis, and the production of effector molecules and serine proteases. One intriguing example is an abundantly represented sensory appendage protein with similarity to odorant or pheromone binding proteins that was noted among the *Aedes aegypti* EST clusters. Two such molecules recently have been reported as having a role in tissue remodeling and are inducible by viral and bacterial infection in *Drosophila* (49).

Products that are necessary for basic cellular function also were noted in this category. The three large protein complexes

TABLE 5. Abundant and immunity-related EST clusters observed from *A. aegypti* and *A. subalbatus* immune response-activated hemocyte libraries

EST-cluster characteristic or function[f]	Sequence product	*A. aegypti*		*A. subalbatus*		*A. gambiae* homolog Ensembl reference no.[c]
		ASAP reference no.[a]	No. of ESTs[b]	ASAP reference no.[a]	No. of ESTs[b]	
Abundance	Adenosine triphosphatase subunit 6			43759	212	
	Apolipophorin			43855	141	ENSANGP00000016631
	Cytochrome *c* oxidase	42809	115	39409	42	ENSANGP00000010310
	Cytochrome *c* oxidase subunit I	34503	152	39115	253	
	Cytochrome *c* oxidase subunit III	33949	248			
	Cytochrome *c* oxidase subunit III			38859	240	
	Cytosolic large ribosomal subunit L41	42807	510	38851	176	ENSANGP00000023895
	Defensin A1 (*DEF*)[d]	34041	195	42475	34	
		33951	118			ENSANGP00000015621
		42891	102			
	Defensin C1	35269	2	43873	138	ENSANGP00000015621
	Gelsolin	35243	3	38857	148	ENSANGP00000020539
	Large ribosomal subunit	34835	400	43761	348	
	Lysozyme	34403	16	43875	123	ENSANGP00000022875
	NADH dehydrogenase subunit 2			43757	266	
	NADH dehydrogenase subunit 6			43861	172	
	Sensory appendage protein	33953	107			ENSANGP00000011659
	Serine protease	35051	35	38861	107	ENSANGP00000020166
	Ubiquinol-cytochrome *c* reductase	42953	125			ENSANGP00000021887
	Ubiquinol-cytochrome *c* reductase			43763	104	
	Unknown	38327	50	43749	349	
	Unknown			43859	118	
	Unknown			43839	122	
	Unknown	42931	125			
Pattern recognition[e]	C-type lectins (*CTL*)[d]	34507	23	42651	4	
		42943	19	39171	40	
				43831	5	
		34019	10	39609	2	
				39397	3	
		42881	21	40377	7	
		35465	2	39337	7	
		34509	17	39599	3	
				39389	8	
				39397	3	
	Gram-negative binding protein (*GNBP*)[d]	37203	1			
	Ficolin (*FGN*)[d]	36031	5			
		36029	2			ENSANGP00000002335
		34957	20			ENSANGP00000011248
				43795	4	ENSANGP00000021318
	Lipid recognition	42885	19	39887	10	ENSANGP00000020083
		43228	1			
				38973	5	
	Peptidoglycan recognition (*PGRP*)[d] [PGRP-LB]			43805	7	ENSANGP00000013948
		34523	15	39497	1	ENSANGP00000017320
	Scavenger receptor (*SCR*)[d]			41117	1	ENSANGP00000012656
		34805	6			
	Sugar binding	42865	12			
		35397	5			
		42899	10			
	Tachylectin-5B	34435	6			ENSANGP00000002335
	TEP III (*TEP*)[e]	42801	8			ENSANGP00000019522
Signal transduction[e]	Toll [*Toll-6*] (*Toll*)[d]	36151	4			
	Toll	36639	1	44039	1	
	kappaB kinase			44041	1	
	Defense response transmembrane receptor [Tollo]	43130	3			
	ECSIT (evolutionarily conserved signaling intermediate in Toll pathways)			41549	1	ENSANGP00000019534

Continued on following page

TABLE 5—*Continued*

EST-cluster characteristic or function[f]	Sequence product	*A. aegypti*		*A. subalbatus*		*A. gambiae* homolog Ensembl[c] reference no.
		ASAP reference no.[a]	No. of ESTs[b]	ASAP reference no.[a]	No. of ESTs[b]	
	Relish isoform (*REL*)[d]	36957	1			
	Carboxylesterase (putative: signal transduction/juvenile hormone [JH] esterase)			41831	1	ENSANGP00000014256
	Conserved unknown (putative: signal transduction/RAS)	36059	1	41057	1	ENSANGP00000011918
	Conserved unknown (putative: GTPase)	34341	5	40543	2	ENSANGP00000012225
	Excitatory extracellular ligand-gated ion channel	36207	3			ENSANGP00000021279
	Fibroblast growth factor binding	37533	1			ENSANGP00000021723
	Ferredoxin-NADP reductase	37795	1			ENSANGP00000014217
	Granulocyte-macrophage colony-stimulating factor signalling molecule			40903	2	ENSANGP00000018753
	GTP binding			40949	1	ENSANGP00000019083
	Histidine triad nucleotide binding protein (signal transduction/ protein kinase C inhibitor)			40237	2	ENSANGP00000012999
	Myo-inositol-1 (or 4)-monophosphatase	36195	5			ENSANGP00000011437
	MAP kinase phosphatase			40515	3	ENSANGP00000013201
	Phosphatidylinositol transporter	36311	1			ENSANGP00000017008
	Presenilin enhancer (signal transduction/Notch pathway)			41183	1	ENSANGP00000010559
	Protein kinase			42191	1	ENSANGP00000011850
	Protein kinase C inhibitor (14-3-3 zeta)	34013	15	42581	6	ENSANGP00000009311
	Protein serine/threonine kinase			41249	1	ENSANGP00000010242
	RAB interactor	36963	1			ENSANGP00000020253*
	Rab-protein 7	43000	1			ENSANGP00000018151
	RAB small monomeric GTPase	36085	5			ENSANGP00000018202
	RAS-related small monomeric GTPase			40009	4	ENSANGP00000023894
	RAS small monomeric GTPase	37835	1			ENSANGP00000013477
	RAS small monomeric GTPase	37479	3	39057	2	ENSANGP00000020422
	RAS opposite	35959	1			ENSANGP00000017600
	Receptor for activated protein kinase C (RACK)			38955	3	ENSANGP00000012560
	Receptor for activated protein kinase C (RACK)	34655	7			ENSANGP00000012560
	Rho GTPase activator			39197	1	ENSANGP00000002091
	Rho small monomeric GTPase			40997	2	ENSANGP00000015684
	Rho small monomeric GTPase			41801	1	ENSANGP00000011746
	Rho small monomeric GTPase	34121	1			ENSANGP00000013799
	Rho small monomeric GTPase	37457	2			ENSANGP00000020445
	Rho small monomeric GTPase	38789	1			ENSANGP00000015684
	Rho small monomeric GTPase	38791	1			ENSANGP00000015684
	SH3/SH2 adaptor	43208	1	39667	1	ENSANGP00000018687
	Suppressor of cytokine signaling (JAK-STAT)	36821	1			ENSANGP00000019768
	Wnt receptor signaling pathway			41855	1	ENSANGP00000010034
Cytoskeletal	Actin	42959	3			ENSANGP00000019055
	Actin	38629	1	42385	4	ENSANGP00000022307
	Actin	38631	11			ENSANGP00000019055
	Actin	38633	1			ENSANGP00000022308
				40125	1	
	Actin			42383	6	ENSANGP0000009996
	Actin			44119	1	ENSANGP00000022306
				38935	6	
	Actin-binding	37431	1			ENSANGP00000020957
	Actin-binding	43301	1			ENSANGP00000001283
	Actin-binding	37667	1			ENSANGP00000021592
	Actin-binding			40691	2	ENSANGP00000008511
	Actin-binding	36245	3	42459	1	ENSANGP00000012542

Continued on following page

TABLE 5—*Continued*

EST-cluster characteristic or function[f]	Sequence product	*A. aegypti*		*A. subalbatus*		*A. gambiae* homolog Ensembll[c] reference no.
		ASAP reference no.[a]	No. of ESTs[b]	ASAP reference no.[a]	No. of ESTs[b]	
	Actin depolymerizing factor	42893	6	39707	6	ENSANGP00000012938
	Alpha actinin			41917	1	ENSANGP00000011796
	Alpha tubulin	36071	3	40413	5	ENSANGP00000002667
	Annexin B11	35171	17			
	Annexin X	37145	1			ENSANGP00000015318
	Beta-tubulin			39059	1	ENSANGP00000024132
	Calcium ion binding	43023	1			ENSANGP00000025334
	Centrosomin	43049	5			
	Clathrin adaptor complex subunit	37429	2	42667	2	ENSANGP00000023452
				42669	2	ENSANGP00000013513
	Coronin			41145	1	ENSANGP00000009406
	Cytoskeleton-associated protein (CAP)			43867	3	ENSANGP00000019876
	Dynein ATPase	36703	1			ENSANGP00000017519
	Dynein ATPase	37177	1			ENSANGP00000015395
	Dynein ATPase	42978	1			ENSANGP00000018736
	Dynamitin			42261	2	ENSANGP00000017909
	GABA-A receptor associated	34983	13			ENSANGP00000023684
	Gelsolin	35243	3	38857	148	ENSANGP00000020539
	Interaptin	34021	14			
	Kinesin motor	43156	1			ENSANGP00000014236
	Myosin light chain kinase			41249	1	ENSANGP00000010242
	Profilin			41677	1	
	Retinoid and fatty acid binding protein	36129	4			ENSANGP00000018348
	Tropomyosin 2			44106	1	ENSANGP00000024231
	Troponin C	35889	1			ENSANGP00000018434
Antioxidant related[e]	Catalase	37745	1			ENSANGP00000021298
		43143	1			
	Glutathione peroxidase	35507	6	42545	4	ENSANGP00000024750
				42547	5	
		35505	14			ENSANGP00000013962
	Glutathione *S*-transferase			42047	1	ENSANGP00000016648
	Glutathione *S*-transferase	43298	1			ENSANGP00000024041
	Glutathione *S*-transferase	35183	2			ENSANGP00000018735
	Peroxiredoxin	36381	2	42607	4	ENSANGP00000020201
	Peroxiredoxin	36499	2	41933	1	ENSANGP00000009997
	Thioredoxin peroxidase			41829	1	ENSANGP00000010951
	Thioredoxin reductase	42805	4			ENSANGP00000017329
	Superoxide dismutase	36733	1	40427	3	ENSANGP00000015824
	Superoxide dismutase			40431	1	ENSANGP00000016164
	Superoxide dismutase			40429	1	ENSANGP00000016164
	Superoxide dismutase	42863	14	43825	30	ENSANGP00000020588
Apoptosis related[e]	Angiopoieitin	36065	6			
	Annexin IX			43939	1	ENSANGP00000015300
	Apoptosis inhibitor (*IAP*)[d]	37367	1			ENSANGP00000009540
	Apoptosis inhibitor	37785	1			ENSANGP00000018745
	Apoptosis-related/cell death Regulatory protein GRIM19			42567	16	ENSANGP00000010318
	Caspase (*CASP*)[d]			39965	1	
	Calcium ion binding	43247	1	39141	1	ENSANGP00000021244
	Conserved unknown			42003	1	
	Conserved unknown (ASPP1 protein)	37303	1			ENSANGP00000012127
	Cyclin A-CDK2 kinase complex p19(Skp1) subunit	35013	5	40587	4	ENSANGP00000011120
	Death-associated protein DAP-1			42609	10	
	Death-associated LIM-only protein DALP			41293	1	
	Death-associated protein	43348	1	42613	1	
	Death-associated protein (DAP-1)	35555	6			
	Nitrilase			40847	3	ENSANGP00000011026

Continued on following page

TABLE 5—*Continued*

EST-cluster characteristic or function[f]	Sequence product	*A. aegypti*		*A. subalbatus*		*A. gambiae* homolog Ensembll[c] reference no.
		ASAP reference no.[a]	No. of ESTs[b]	ASAP reference no.[a]	No. of ESTs[b]	
	Novel cell death-regulatory protein GRIM19	36005	8	42565	33	ENSANGP00000010318
	Oligosaccharyl transferase	42831	16	41913	1	ENSANGP00000019361
	Paxillin-derived LIM-only protein			41293	1	
	Pendulin	35971	1			ENSANGP00000015835
	Phospholipase A2	37471	1	43958	1	ENSANGP00000012556
	Phospholipid scramblase	37097	1			ENSANGP00000022984
	Programmed cell death gene 5	36505	4	39567	1	ENSANGP00000014166
	Signal recognition particle subunit			41619	1	ENSANGP00000018437
	26S proteasome regulatory ATPase subunit 10b	35087	1	40253	4	ENSANGP00000017473
	Viral inhibitor of apoptosis-associated factor 1 (viaf1)	37817	1	39067	1	ENSANGP00000023564
	Zinc RING finger SAG (sensitive to apoptosis gene)	37649	1	40813	2	ENSANGP00000007214
Melanization[e]	Beta-alanyl-dopamine synthase	36265	1			ENSANGP00000021728
	Calreticulin	36981	1	40361	8	ENSANGP00000012895
	Dopachrome conversion enzyme	35157	3			ENSANGP00000020299
	5′ nucleotidase	37217	1			ENSANGP00000008588
	Phenylalanine hydroxylase	36461	2	44137	1	ENSANGP00000016481
	Prophenoloxidase [PPO9] (*PPO*)[d]	37239	1			ENSANGP00000010740
	Prophenoloxidase [PPO2]	37573	1			
				41327	1	ENSANGP00000020648
		43235	1			
Stress response	ATP/ADP translocase	35589	2			
	Contains similarity to Pfam family BolA: BolA-like protein	34423	2			ENSANGP00000021299
	Contains similarity to Pfam family BolA: BolA-like protein	43102	1			ENSANGP00000010860
	Contains similarity to Pfam domain DnaJ: DnaJ domain	36485	2			ENSANGP00000010358
	Heat shock protein	36133	4	42445	9	ENSANGP00000018891
	Heat shock protein	36759	1			ENSANGP00000012893
	Heat shock protein	43056	1			ENSANGP00000016646
	Protein chaperone			40323	1	ENSANGP00000020237
	Heat shock cognate 70-3 (Hsc70-3)			40969	1	ENSANGP00000012893
	Heat shock			41111	1	ENSANGP00000019416
	Heat shock	35067	2			ENSANGP00000018254
	Heat shock protein			42443	11	ENSANGP00000018891
	Heat shock	37057	3			ENSANGP00000022995
	Heat shock	37503	1			ENSANGP00000016349
	Heat shock	37547	3			ENSANGP00000019412
	Hsp70/Hsp90 organizing protein	34131	1			ENSANGP00000012254
	Selenoprotein T	36405	4	42397	3	
	Ubiquitin-63E	35327	2			ENSANGP00000024710
Iron metabolism	Cysteine desulfhydrase	36807	1			ENSANGP00000016500
	Ferridoxin	35009	1	40679	2	ENSANGP00000013242
	Ferritin subunit			40495	3	
		42813	79			
		34033	13			
		42825	14			
		42839	20			ENSANGP00000022116
	Iron-sulfur cluster assembly	37069	1	40989	2	ENSANGP00000019230
		42819	50	39637	3	ENSANGP00000010440
	Transferrin			39807	3	
				43896	1	ENSANGP00000010836
		34773	5			
Serine proteases (*CLIP*)[e]	Clip domain serine protease	35481	1			
	Serine protease	34111	9			ENSANGP00000019999
	Serine protease	35069	1			
	Serine protease	35917	1			ENSANGP00000020166

Continued on following page

TABLE 5—*Continued*

EST-cluster characteristic or function[f]	Sequence product	*A. aegypti*		*A. subalbatus*		*A. gambiae* homolog Ensembl[c] reference no.
		ASAP reference no.[a]	No. of ESTs[b]	ASAP reference no.[a]	No. of ESTs[b]	
	Serine protease	34921	3			
	Serine protease			41821	1	
	Serine protease	42961	7			
	Serine protease	35597	3			
	Serine protease	35849	1	40523	3	
	Serine protease	34123	5			
	Serine protease [CLIPA4]	35035	15	40457	5	ENSANGP00000020196
	Serine protease [CLIPA5]	34915	20			
				39619	9	ENSANGP00000020259
		42835	10			
	Serine protease [CLIPA6]	35917	1			ENSANGP00000020166
	Serine protease [CLIPD1]	43074	2			ENSANGP00000014938
	Serine protease [CLIPB8]	35153	4	41235	2	ENSANGP00000024671
	Serine protease [CLIPB13]			43829		ENSANGP00000012642
	Serine protease [CLIPB15]	36015	10			ENSANGP00000015815
	Serine protease [*easter*]			40019	4	ENSANGP00000013542
		35675	4	38879	32	
				43821	8	
		34873	11			
				42497	3	
		36039	3	41959	1	
		38651	7	42495	4	
		38653	3			
	Serine protease 14A	35957	1	43797	23	
	Serine protease [14D]	35073	5	40373	8	ENSANGP00000023886
	Serine protease [14D2]			42015	2	ENSANGP00000011720
	Serine protease [14D2]			39437	4	ENSANGP00000010548
	Serine protease [SP24D]	36057	6	42187	1	ENSANGP00000025173
	Serine protease [*snake*]	34123	5			
Serpins (*SRPN*)[e]	Pacifastin light chain	35523	14			
	Serpin			40445	6	ENSANGP00000022846
	Serpin [SRPN2]	36055	5	43912	2	ENSANGP00000021812
	Serpin	36027	3	41629	1	
	Serpin	36053	5			
	Serpin [*Spn-27A*]			43962	2	ENSANGP00000007723
	Serine protease inhibitor 4	38597	5			ENSANGP00000023448
	Serine protease inhibitor	38599	2			ENSANGP00000015833
General[e]	Scribbled	35359	3			ENSANGP00000014905
	Fat spondin			41163	1	ENSANGP00000008856
	Conserved unknown (expressed in CD34+ hematopoietic stem/ progenitor cells)	43263	1	41251	1	ENSANGP00000010523
	Conserved unknown (T-cell activation protein phosphatase 2C; TA-PP2C)	36419	3	41749	1	ENSANGP00000019458
	Asparagine-tRNA ligase (autoimmunity-related)	36519	2			ENSANGP00000011058

[a] More information is available at https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm by using these reference numbers.
[b] Number of ESTs that were compiled into the EST cluster.
[c] Homologs identified in the *Anopheles gambiae* genome. More information is available at Ensembl (http://www.ensembl.org/Anopheles_gambiae/) by using these reference numbers.
[d] Abbreviations represent immunity-related gene families identified in the *A. gambiae* genome project.
[e] Group specifically defined as immunity related.
[f] The total number of EST clusters in *A. aegypti* and *A. subalbatus* respectively, with the indicated characteristics or functions were as follows: sequence abundance, 11 and 16; pattern recognition, 36 and 20; signal transduction, 43 and 34; cytoskeletal, 26 and 23; antioxidant related, 14 and 10; apoptosis related, 19 and 22; melanization, 8 and 3; stress response, 14 and 8; iron metabolism, 12 and 9; serine proteases, 57 and 26; serpins, 7 and 4.

(NADH dehydrogenase, cytochrome reductase, and cytochrome *c* oxidase) involved in the respiratory chain, in addition to ATPase (ATP synthase), are represented. These enzyme systems likely serve to provide requisite energy for various cellular processes, or their up-regulation may be indicative of impending apoptosis, as has been observed in various vertebrate cells (9). Notably, four highly abundant EST clusters could not be identified by comparison to sequence databases. It is interesting to speculate that orthologous EST clusters that were assembled from disparate numbers of sequences, or

abundantly represented EST clusters that do not have an orthologous sequence, are associated with observed experimental differences in immune responsiveness between these two mosquito species (7, 28).

Pattern recognition. Recognition of pathogen-associated molecular patterns initiates innate immune responsiveness in both vertebrates and invertebrates (44). The cascade of events begins when pathogen-associated molecular patterns are bound by pattern recognition receptors such as peptidoglycan recognition proteins (see reference 12). These and a number of additional molecules that likely function in pattern recognition, including scavenger receptors, lectins, gram-negative binding proteins, and products containing fibrinogen domains, were identified in hemocyte ESTs of both mosquito species (Table 5). Four ESTs from *Aedes aegypti* are similar to thioester-containing proteins shown to be immune responsive with complement-like opsonin properties (38, 40). Three EST clusters were identified as having high similarity (score, 116; e = 7.1e-32) to the Pfam model MD-2-related lipid-recognition domain (E1_DerP2_DerF2), a lipid-binding protein essential for mammalian recognition of lipopolysaccharide in cooperation with Toll-like receptor 4 (34).

Phagocytosis. Phagocytosis is a primary innate immune response and a mediator of subsequent responses in adaptive immunity. Like vertebrates, insects utilize subpopulations of professional phagocytic cells, sometimes referred to as macrophage-like, in response to invading pathogens. Evidence for phagocytic events is manifested in the large number of signal transduction, cytoskeletal rearrangement, and apoptotic elements identified (23) (Table 5). In agreement with the results from ultrastructural studies of *Aedes aegypti* and *Armigeres subalbatus* hemocytes (28, 29), the sequences presented here emphasize the importance of phagocytosis in the mosquito immune response.

Signal transduction. Considering the extensive cell-cell communication and collaborative interaction mandated by phagocytic activity, it is not surprising to see numerous signal transduction elements and pathways represented in ESTs from immune response-activated mosquito hemocytes. Emphasis has been placed on the Toll and immune deficiency (*imd*) signaling pathways as elicitors of antimicrobial responses (reviewed in reference 32), but the importance of alternative signaling pathways in insect cell phagocytosis and immunity is also becoming clear (8, 33). A report concerning *Anopheles gambiae* showed that *Anopheles gambiae* STAT is translocated to the nucleus in response to a bacterial challenge (1). *Drosophila* Janus kinase kinase hopscotch (*hop*) gain-of-function mutants constitutively express the complement-like protein Tep-1 (38). Molecules related to signaling pathways including Toll and *imd* are presented in Table 5 and include numerous Rho family GTPases that may be involved in actin reorganization during phagocytosis (23).

Cytoskeletal elements. Phagocytic cells in *Aedes aegypti* and *Armigeres subalbatus*, called granulocytes, have the capacity to engulf hundreds of bacteria, which undoubtedly requires extensive cytoskeletal remodeling (23, 29). In support of this observation, 31 EST clusters were identified as cytoskeletal elements, including actin, actin-binding and polymerizing factors, and alpha- and beta-tubulins (Table 5). Noteworthy among these is gelsolin, an actin-severing molecule represented by 148 sequences in a single EST from *Armigeres subalbatus*.

Cytotoxic molecule production and antioxidant-related immune response. Reactive oxygen intermediates function ubiquitously in phagocytic cells as cytotoxic effector molecules. Evidence for both reactive oxygen species production and restorative chemistry following an oxidative burst are evident in hemocyte EST clusters (Table 5). Reductases, catalase, and peroxidases remove excess reactive oxygen intermediates to return cells to homeostasis following oxidative bursts.

Apoptosis-related immunity. Apoptosis and phagocytic immune responses are intimately linked so that organisms can achieve homeostasis before and after an immune challenge (47). As do many immune response elements, this phenomenon transcends phyla (20, 31). The sequence of the *Anopheles gambiae* genome revealed both pro- and antiapoptotic regulators (12). To further demonstrate that mosquitoes employ these gene products in response to bacteria, ESTs from immune response-challenged mosquito hemocytes include 19 *Aedes aegypti* and 22 *Armigeres subalbatus* apoptosis-related molecules (Table 5).

Melanotic encapsulation. Melanin biosynthesis is a hemocyte-mediated immune response that involves a complex yet well-characterized (6) cascade of reactions beginning with tyrosine and ending with the polymerization of a capsule that surrounds an invading parasite (55), eventually killing the parasite (10, 45). This is an important mechanism for mosquito resistance to eukaryotic parasites, including malaria parasites and filarial worms (7, 14). Ultrastructural studies demonstrate that this response is also rapidly deployed against bacteria (24, 25, 28–30). In *Aedes aegypti* and *Armigeres subalbatus*, enzymes involved in melanization, including prophenoloxidase and phenylalanine hydroxylase, are produced primarily by oenocytoids (27, 28, 36). These enzymes, as well as dopachrome conversion enzyme and multiple sequences with similarity to prophenoloxidase activating factors, are represented in EST clusters from both species (Table 5).

AMPs. The AMPs—effector molecules that exhibit in vitro activities against bacteria, fungi, and protozoa—are considered a primary defense element in mosquito innate immunity. Transcriptional up-regulation of AMPs has been correlated not only with responses to bacteria and fungi but also with various stages of *Plasmodium* infection in *Anopheles gambiae* (17, 48). A number of novel and previously reported defensins, cecropins, and lysozymes, as well as gambicin, are represented in EST clusters from immune response-challenged hemocytes. The production of AMPs occurs predominantly in the fat body but has also been reported to occur in various tissues and in immune-responsive mosquito cell lines (see references 18 and 26). The extent to which the immune peptides discussed herein are produced by the fat body or hemocytes will need to be resolved with cytochemical assays.

Previously reported immune peptides from *Aedes aegypti* include three isoforms of defensin and a single isoform of cecropin (41). Immune peptide sequences from *Armigeres subalbatus* have not been reported. A number of sequences encoding immune peptides are represented in the EST clusters produced for the present project, several of which were compiled from a large number of sequences. For example, 3 of the 11 *Aedes aegypti* EST clusters that are composed of more than

TABLE 6. AMPs observed in immune response-activated *A. aegypti* and *A. subalbatus* EST clusters

AMP[e]	*A. aegypti*		*A. subalbatus*		*A. gambiae* homolog[c] Ensembl reference no.
	ASAP reference no.[a]	No. of ESTs[b]	ASAP reference no.[a]	No. of ESTs[b]	
Defensin A1 (*DEF*)[d]	33951	118	42475	34	ENSANGP00000015621
	42891	102			
	34041	195			
Defensin A2	42889	48	42477	48	ENSANGP00000015621
Defensin A3			38965	6	ENSANGP00000015621
Defensin B			43865	20	ENSANGP00000015621
Defensin C1	35269	2	43841	15	ENSANGP00000015621
	42923	76	43873	138	
	35023	1			
Defensin C2	35267	3			None
Defensin N	35455	2			ENSANGP00000015622
Cecropin A (*CEC*)[d]	33981	72	43847	31	ENSANGP00000011957
			43849	6	
Cecropin B	34401	9	39353	3	ENSANGP00000011957
Cecropin C1	33979	42	43851	13	ENSANGP00000011957
			43853	13	
Cecropin C2	34223	19	39621	1	ENSANGP00000011957
Cecropin C3	34293	24			ENSANGP00000011957
Cecropin D	37327	3			ENSANGP00000011957
Cecropin N	38409	29	39315	23	ENSANGP00000011957
	38411	4			
Gambicin	35613	4	42551	2	ENSANGP00000013255
	34077	2	42549	21	
Lysozyme A	34403	16	43875	123	ENSANGP00000022875
Lysozyme B	42917	3	40443	5	ENSANGP00000018439
Lysozyme C	36997	3			ENSANGP00000018395
Diptericin (putative)	37085	1			ENSANGP00000018302
Holotricin (questionable)	34447	2			None

[a] More information is available at https://asap.ahabs.wisc.edu/annotation/php/ASAP1.htm by using with these reference numbers.
[b] Number of ESTs that were compiled into the EST cluster.
[c] Homologs identified in the *A. gambiae* genome. More information is available at Ensembl (http://www.ensembl.org/Anopheles_gambiae/) by using these reference numbers.
[d] Abbreviations represent immunity-related gene families identified in the *A. gambiae* genome project.
[e] Total AMPs observed for *A. aegypti*, 30; total AMPs observed for *A. subalbatus*, 24.

100 sequences encode defensins; likewise, 1 of the 16 ESTs from *Armigeres subalbatus* that is most abundantly represented is a defensin. Another *Armigeres subalbatus* EST encoding lysozyme is abundantly represented with 123 sequences (Tables 5 and 6). Further sequence analysis was done to resolve the type and number of representatives of each immune peptide for both mosquito species.

Among the EST clusters from both mosquito species, four isoforms of defensin were observed. An *Armigeres subalbatus* sequence shows similarity to defensin genes but lacks the conserved cysteine residues necessary for the characteristic structure of the molecule. Five different cecropin isoforms were observed. Isoform N, not previously described, exhibits a unique amino acid sequence. Three isoforms of lysozyme were also among the sequences reported here, as was gambicin, a newly described immune peptide from *Anopheles gambiae* that has broad activity against bacteria, fungi, and *Plasmodium* parasites (53). Gambicin was not found in our BLAST analysis of the *Anopheles gambiae* predicted peptide sequences; however, the sequence was identified in a more recent version (version 17.2a.1) of the genome. Two *Aedes aegypti* sequences, categorized as questionable, have weak similarity to diptericin and holotricin. A tabular description of these sequences and this nomenclature is presented in Table 6.

The clear energetic investment made by these mosquitoes to produce redundant AMPs in abundance, when the capacity of hemocytes is likely sufficient to clear an infection (16, 28), suggests that these molecules might have additional roles in insect innate immunity beyond cytotoxicity (4). Signaling and chemotaxis may be functions of mosquito AMPs, as is the case in vertebrate immunity (50). An alternative view is that redundant AMPs in the *Anopheles gambiae* genome have alternate antimicrobial specificities (12).

Stress response. Bacterial infection and the robust cellular and humoral responses elicited by that infection undoubtedly serve as stressors for infected mosquitoes. Numerous stress-responsive molecules were detected among these ESTs, including heat shock proteins and ubiquitin-63 (Table 5). These data implicate a link between the innate immune and stress responses in mosquitoes and support a previous assessment of altered immune capacity in environmentally stressed *Anopheles gambiae* (51).

Iron metabolism. Iron transport and storage proteins have diverse roles in insect physiology. Of interest for innate immunity is the host's ability to sequester iron to hinder pathogen survival. ESTs from both species contain multiple molecules related to iron metabolism, including iron-sulfur cluster proteins, ferredoxin, and cysteine desulfhydrase (Table 5). Several reports have described up-regulation of transferrins in insects or insect cells challenged with bacteria (reviewed in reference 46), and the transferrin molecule is represented by three EST clusters in immune response-activated hemocytes. In contrast, ferritin, an iron storage protein, is represented by five clusters that contain numerous sequences. Two ferritin light chain-like clusters contain 92 *Aedes aegypti* sequences, suggesting the importance of this molecule in the immune response; whether this apparent transcriptional abundance is inducible will need to be examined further, because ferritin modulation has not previously been demonstrated in association with innate immune responses in insects (39).

Serine proteases and serpins. Diverse functions in insect innate immunity, including hemolymph coagulation, AMP production, and melanotic encapsulation, are modulated by various serine proteases and serpins (12, 22). The importance of these molecules is illustrated by the number of immunity-related serine proteases identified and their abundance (Table 5). One serine protease in *Armigeres subalbatus* is represented by 107 sequences, and its *Aedes aegypti* ortholog contains 35 sequences.

Mosquito-specific EST clusters. Following the addition of *Anopheles gambiae* BLAST results to *Aedes aegypti* and *Armigeres subalbatus* sequences in ASAP, 102 EST clusters that had no similarity to sequences in nonmosquito databases were noted (data not shown). These intriguing sequences, especially those that are represented multiple times, may represent mosquito-specific gene products related to immune responsiveness and merit further investigation.

Future prospects. The data presented here represent the first high-throughput analysis of the transcriptome of immunocytes from an invertebrate and provide a road map for future investigations that aim to elucidate interactions between the mosquito host and the pathogens it transmits. Host-parasite interactions represent coevolved adaptations of considerable complexity, and the compatibility of these relationships is dependent on the relative capacities of the host to recognize and respond to a foreign object and of the parasite to weaken or inhibit the immune system of the host. These dynamic interactions are unquestionably manifested in the transcriptomes of immune response-activated mosquito hemocytes. In anticipation of conducting in-depth studies of the dynamics of the hemocyte transcriptome in response to bacteria, malaria, and filarial worm parasites, microarrays are being generated. Having well-characterized ESTs provided a unique capacity to strategically design oligonucleotide-based microarrays.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI19769, AI053772, and AI46032 (to B. M. Christensen), National Science Council of Taiwan grant NSC 91-3112-B-010-001 (to C.-C. Chen), and NIAID graduate fellowship AI07414-11 (to L. C. Bartholomay).

REFERENCES

1. **Barillas-Mury, C., Y.-S. Han, D. Seeley, and F. C. Kafatos.** 1999. *Anopheles gambiae* Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. EMBO J. **18:**959–967.
2. **Barillas-Mury, C., B. Wizel, and Y. Soo Han.** 2000. Mosquito immune responses and malaria transmission: lessons from insect model systems and implications for vertebrate innate immunity and vaccine development. Insect Biochem. Mol. Biol. **30:**429–442.
3. **Bartholomay, L. C., H. A. Farid, R. Ramzy, and B. M. Christensen.** 2003. *Culex pipiens pipiens*: characterization of immune peptides and the influence of immune activation on development of *Wuchereria bancrofti*. Mol. Biochem. Parasitol. **130:**43–50.
4. **Bartholomay, L. C., J. F. Fuchs, L. L. Cheng, E. T. Beck, J. Vizioli, C. Lowenberger, and B. M. Christensen.** 2004. Reassessing the role of defensin in the innate immune response of the mosquito, *Aedes aegypti*. Insect Mol. Biol. **13:**125–132.
5. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. L. Sonnhammer.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276–280.
6. **Beerntsen, B. T., A. A. James, and B. M. Christensen.** 2000. Genetics of mosquito vector competence. Microbiol. Mol. Biol. Rev. **64:**115–137.
7. **Beerntsen, B. T., S. Luckhart, and B. M. Christensen.** 1989. *Brugia malayi* and *Brugia pahangi*: inherent difference in immune activation in the mosquitoes *Armigeres subalbatus* and *Aedes aegypti*. J. Parasitol. **75:**76–81.
8. **Boutros, M., H. Agaisse, and N. Perrimone.** 2002. Sequential activation of signaling pathways during innate immune responses in *Drosophila*. Dev. Cell **3:**711–722.
9. **Chandra, D., J.-W. Liu, and D. G. Tang.** 2002. Early mitochondrial activation and cytochrome c up-regulation during apoptosis. J. Biol. Chem. **277:**50842–50854.
10. **Chen, C. C., and C. S. Chen.** 1995. *Brugia pahangi*: effects of melanization on the uptake of nutrients by microfilariae in vitro. Exp. Parasitol. **81:**72–78.
11. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
12. **Christophides, G. K., E. Zdobnov, C. Barillas-Mury, E. Birney, S. Blandin, C. Blass, P. T. Brey, F. H. Collins, A. Danielli, G. Dimopoulos, C. Hetru, N. T. Hoa, J. A. Hoffmann, S. M. Kanzok, I. Letunic, E. A. Levashina, T. G. Loukeris, G. Lycett, S. Meister, K. Michel, L. F. Moita, H.-M. M. Muller, M. A. Osta, S. M. Paskewitz, J. M. Reichhart, A. Rzhetsky, L. Troxler, K. D. Vernick, D. Vlachou, J. Volz, C. von Mering, J. Xu, L. Zheng, P. Bork, and F. C. Kafatos.** 2002. Immunity-related genes and gene families in *Anopheles gambiae*. Science **298:**159–165.
13. **Collins, F. H., and A. A. James.** 1996. Genetic modification of mosquitoes. Sci. Med. **3:**52–61.
14. **Collins, F. H., R. K. Sakai, K. D. Vernick, S. Paskewitz, D. Seeley, L. H. Miller, W. E. Collins, C. C. Campbell, and R. W. Gwadz.** 1986. Genetic selection of a *Plasmodium*-refractory strain of the malaria vector *Anopheles gambiae*. Science **234:**607–610.
15. **Da Silva, J. B., C. M. D. Albuquerque, E. C. De Araujo, C. A. Peixoto, and H. Hurd.** 2000. Immune defense mechanisms of *Culex quinquefasciatus* (Diptera: Culicidae) against *Candida albicans* infection. J. Invertebr. Pathol. **76:**257–262.
16. **Dimopoulos, G., T. L. Casavant, S. Chang, T. Scheetz, C. Roberts, M. Donohue, J. Schultz, V. Benes, P. Bork, W. Ansorge, M. B. Soares, and F. C. Kafatos.** 2000. *Anopheles gambiae* pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. Proc. Natl. Acad. Sci. USA **97:**6619–6624.
17. **Dimopoulos, G., D. Seeley, A. Wolf, and F. C. Kafatos.** 1998. Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during the critical stages of the parasite life cycle. EMBO J. **17:**6115–6123.
18. **Fallon, A. M., and D. Sun.** 2001. Exploration of mosquito immunity using cells in culture. Insect Biochem. Mol. Biol. **31:**263–278.
19. **FlyBase Consortium.** 2002. The FlyBase database of the *Drosophila* genome projects and community literature. Nucleic Acids Res. **30:**106–108.
20. **Franc, N. C.** 2002. Phagocytosis of apoptotic cells in mammals, *Caenorhabditis elegans*, and *Drosophila melanogaster*: molecular mechanisms and physiological consequences. Front. Biosci. **7:**1298–1313.
21. **Glasner, J. D., P. Liss, G. Plunkett III, A. Darling, T. Prasad, M. Rusch, A. Byrnes, M. Gilson, B. Beihl, F. R. Blattner, and N. T. Perna.** 2003. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids Res. **31:**147–151.
22. **Gorman, M. J., and S. M. Paskewitz.** 2001. Serine proteases as mediators of mosquito immune responses. Insect Biochem. Mol. Biol. **31:**257–262.
23. **Greenberg, S., and S. Grinstein.** 2002. Phagocytosis and innate immunity. Curr. Opin. Immunol. **14:**136–145.
24. **Hernandez, S., H. Lanz, M. H. Rodriguez, J. A. Torres, A. Martinez-Palomo, and V. Tsutsumi.** 1999. Morphological and cytochemical characterization of female *Anopheles albimanus* (Diptera: Culicidae) hemocytes. J. Med. Entomol. **36:**426–434.

25. **Hernandez-Martinez, S., H. Lanz, M. H. Rodriguez, L. Gonzalex-Ceron, and V. Tsutsumi.** 2002. Cellular-mediated reactions to foreign organisms inoculated into the hemocoel of *Anopheles albimanus* (Diptera: Culicidae). J. Med. Entomol. **39:**61–69.
26. **Hetru, C., L. Troxler, and J. A. Hoffmann.** 2003. *Drosophila melanogaster* antimicrobial defense. J. Infect. Dis. **187:**S327–S334.
27. **Hillyer, J. F., and B. M. Christensen.** 2002. Characterization of hemocytes from the yellow fever mosquito, *Aedes aegypti*. Histochem. Cell Biol. **117:** 431–440.
28. **Hillyer, J. F., S. L. Schmidt, and B. M. Christensen.** 2003. Hemocyte-mediated phagocytosis and melanization in the mosquito *Armigeres subalbatus* following immune challenge by bacteria. Cell Tissue Res. **313:**117–127.
29. **Hillyer, J. F., S. Schmidt, and B. M. Christensen.** 2003. Rapid phagocytosis and melanization of bacteria and *Plasmodium* sporozoites by hemocytes of the mosquito *Aedes aegypti*. J. Parasitol. **89:**62–69.
30. **Hillyer, J. F., S. L. Schmidt, and B. M. Christensen.** 2004. The antibacterial innate immune response by the mosquito *Aedes aegypti* is mediated by hemocytes and independent of Gram type and pathogenicity. Microbes Infect. **6:**448–459.
31. **Hoffmann, J. A., F. C. Kafatos, C. A. Janeway, and R. A. Ezekowitz.** 1999. Phylogenetic perspectives in innate immunity. Science **284:**1313–1318.
32. **Hoffmann, J. A., and J. M. Reichhart.** 2002. *Drosophila* innate immunity: an evolutionary perspective. Nat. Immunol. **3:**121–126.
33. **Huang, S.-J., R. C.-C. Wu, M.-F. Shaio, P. S. Wang, and W. L. Cho.** 2003. An immune signalling kinase AaMEK3 from mosquitoes: cDNA cloning and characterization. Insect Mol. Biol. **12:**595–603.
34. **Inohara, N., and G. Nunez.** 2002. ML—a conserved domain involved in innate immunity and lipid metabolism. Trends Biochem. Sci. **27:**219–221.
35. **Irving, P., L. Troxler, T. S. Heuer, M. Belvin, C. Kopczynski, J.-M. Reichhart, J. A. Hoffmann, and C. Hetru.** 2001. A genome-wide analysis of immune responses in *Drosophila*. Proc. Natl. Acad. Sci. USA **98:**15119–15124.
36. **Johnson, J. K., T. A. Rocheleau, J. F. Hillyer, C. C. Chen, J. Li, and B. M. Christensen.** 2003. A potential role for phenylalanine hydroxylase in mosquito immune responses. Insect Biochem. Mol. Biol. **33:**345–354.
37. **Knudson, D. L., S. E. Brown, and D. W. Severson.** 2002. Culicine genomics. Insect Biochem. Mol. Biol. **32:**1193–1197.
38. **Lagueux, M., E. Perrodou, E. A. Levashina, M. Capovilla, and J. A. Hoffmann.** 2000. Constitutive expression of a complement-like protein in Toll and JAK gain-of-function mutants of *Drosophila*. Proc. Natl. Acad. Sci. USA **97:**11427–11432.
39. **Law, J. H.** 2002. Insects, oxygen, and iron. Biochem. Biophys. Res. Commun. **292:**1191–1195.
40. **Levashina, E. A., L. F. Moita, S. Blandin, G. Vriend, M. Lagueux, and F. C. Kafatos.** 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. Cell **104:**709–718.
41. **Lowenberger, C.** 2001. Innate immune response of *Aedes aegypti*. Insect Biochem. Mol. Biol. **31:**219–229.
42. **Lowenberger, C. A., C. T. Smartt, P. Bulet, M. T. Ferdig, D. W. Severson, J. A. Hoffmann, and B. M. Christensen.** 1999. Insect immunity: molecular cloning, expression, and characterization of cDNAs and genomic DNA encoding three isoforms of insect defensin in *Aedes aegypti*. Insect Mol. Biol. **8:**107–118.
43. **Maizels, R. M., M. L. Blaxter, and A. L. Scott.** 2001. Immunological genomics of *Brugia malayi*: filarial genes implicated in immune evasion and protective immunity. Parasite Immunol. **23:**327–344.
44. **Medzhitov, R., and C. A. Janeway.** 2002. Decoding the patterns of self and nonself by the innate immune system. Science **296:**298–300.
45. **Nappi, A. J., and E. Ottaviani.** 2000. Cytotoxicity and cytotoxic molecules in invertebrates. Bioessays **22:**469–480.
46. **Nichol, H., J. H. Law, and J. J. Winzerling.** 2002. Iron metabolism in insects. Annu. Rev. Entomol. **47:**535–559.
47. **Opferman, J. T., and S. J. Korsmeyer.** 2003. Apoptosis in the development and maintenance of the immune system. Nat. Immunol. **4:**410–415.
48. **Richman, A. M., G. Dimopoulos, D. Seeley, and F. C. Kafatos.** 1997. *Plasmodium* activates the innate immune response of *Anopheles gambiae* mosquitoes. EMBO J. **16:**6114–6119.
49. **Sabatier, L., E. Jouanguy, C. Dostert, D. Zachary, J.-L. Dimarq, P. Bulet, and J.-L. Imler.** 2003. Pherokine-2 and -3: two *Drosophila* molecules related to pheromone/odor-binding proteins induced by viral and bacterial infections. Eur. J. Biochem. **270:**3398–3407.
50. **Salzet, M.** 2002. Antimicrobial peptides are signaling molecules. Trends Immunol. **23:**283–284.
51. **Suwanchaichinda, C., and S. Paskewitz.** 1998. Effects of larval nutrition, adult body size, and adult temperature on the ability of *Anopheles gambiae* (Diptera: Culicidae) to melanize sephadex beads. J. Med. Entomol. **35:**157–161.
52. **Tzou, P., E. De Gregorio, and B. Lemaitre.** 2002. How *Drosophila* combats microbial infection: a model to study innate immunity and host-pathogen interactions. Curr. Opin. Microbiol. **5:**102–110.
53. **Vizioli, J., P. Bulet, J. A. Hoffmann, F. C. Kafatos, H.-M. Muller, and G. Dimopoulos.** 2001. Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA **98:**12630–12635.
54. **World Health Organization.** 2002. The world health report 2002: reducing risks, promoting healthy life. World Health Organization, Geneva, Switzerland.
55. **Zhao, X., M. T. Ferdig, J. Li, and B. M. Christensen.** 1995. Biochemical pathway of melanotic encapsulation of *Brugia malayi* in the mosquito, *Armigeres subalbatus*. Dev. Comp. Immunol. **19:**205–215.

Editor: W. A. Petri, Jr.