

Research article

Open Access

## Differential gene expression in abdomens of the malaria vector mosquito, *Anopheles gambiae*, after sugar feeding, blood feeding and *Plasmodium berghei* infection

Ali N Dana<sup>1</sup>, Maureen E Hillenmeyer, Neil F Lobo<sup>1</sup>, Marcia K Kern<sup>1</sup>, Patricia A Romans<sup>2</sup> and Frank H Collins\*<sup>1</sup>

Address: <sup>1</sup>Center for Tropical Disease Research and Training, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA and <sup>2</sup>Department of Zoology, University of Toronto, Toronto, ON M5S 3G5, Canada

Email: Ali N Dana - [adana@iupui.edu](mailto:adana@iupui.edu); Maureen E Hillenmeyer - [maureen@stanford.edu](mailto:maureen@stanford.edu); Neil F Lobo - [nlobo@nd.edu](mailto:nlobo@nd.edu); Marcia K Kern - [mkern@nd.edu](mailto:mkern@nd.edu); Patricia A Romans - [victorie@zoo.utoronto.ca](mailto:victorie@zoo.utoronto.ca); Frank H Collins\* - [frank@nd.edu](mailto:frank@nd.edu)

\* Corresponding author

Published: 19 May 2006

Received: 17 October 2005

BMC Genomics 2006, 7:119 doi:10.1186/1471-2164-7-119

Accepted: 19 May 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/119>

© 2006 Dana et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Large scale sequencing of cDNA libraries can provide profiles of genes expressed in an organism under defined biological and environmental circumstances. We have analyzed sequences of 4541 Expressed Sequence Tags (ESTs) from 3 different cDNA libraries created from abdomens from *Plasmodium* infection-susceptible adult female *Anopheles gambiae*. These libraries were made from sugar fed (S), rat blood fed (RB), and *P. berghei*-infected (IRB) mosquitoes at 30 hours after the blood meal, when most parasites would be transforming ookinetes or very early oocysts.

**Results:** The S, RB and IRB libraries contained 1727, 1145 and 1669 high quality ESTs, respectively, averaging 455 nucleotides (nt) in length. They assembled into 1975 consensus sequences – 567 contigs and 1408 singletons. Functional annotation was performed to annotate probable molecular functions of the gene products and the biological processes in which they function. Genes represented at high frequency in one or more of the libraries were subjected to digital Northern analysis and results on expression of 5 verified by qRT-PCR.

**Conclusion:** 13% of the 1965 ESTs showing identity to the *A. gambiae* genome sequence represent novel genes. These, together with untranslated regions (UTR) present on many of the ESTs, will inform further genome annotation. We have identified 23 genes encoding products likely to be involved in regulating the cellular oxidative environment and 25 insect immunity genes. We also identified 25 genes as being up or down regulated following blood feeding and/or feeding with *P. berghei* infected blood relative to their expression levels in sugar fed females.

### Background

Sequencing of the *Anopheles gambiae* genome was completed in 2002 [1]. Annotation and gene prediction have been ongoing. Although more than 14,700 genes and

more than 16,100 transcripts have now been predicted, the functions of approximately 40% of the gene products remain unknown and *in silico* annotations of many others still require verification [1,2]. Information about the

**Table 1: cDNA LIBRARY INSERT AND EST SIZES\***

| Clone Source Library  | N   | Mean length (bp)                  | ± SE | Max. (bp) | Min. (bp) |
|-----------------------|-----|-----------------------------------|------|-----------|-----------|
| <b>S</b>              | 243 | 1003                              | 33   | 3526      | 322       |
| <b>RB</b>             | 189 | 1156                              | 45   | 3738      | 152       |
| <b>IRB</b>            | 272 | 1044                              | 28   | 2789      | 317       |
| <b>Average</b>        |     | 1068                              | 35   | 3351      | 264       |
| EST Source            | N   | Mean EST length (nt) <sup>†</sup> | ± SE | Max. (nt) | Min. (nt) |
| <b>S singletons</b>   | 590 | 404                               | 8    | 820       | 100       |
| <b>RB singletons</b>  | 381 | 394                               | 8    | 748       | 100       |
| <b>IRB singletons</b> | 437 | 377                               | 8    | 785       | 100       |
| <b>Contigs</b>        | 567 | 595                               | 11   | 1585      | 104       |
| <b>Average</b>        |     | 455                               | 9    | 1039      | 101       |

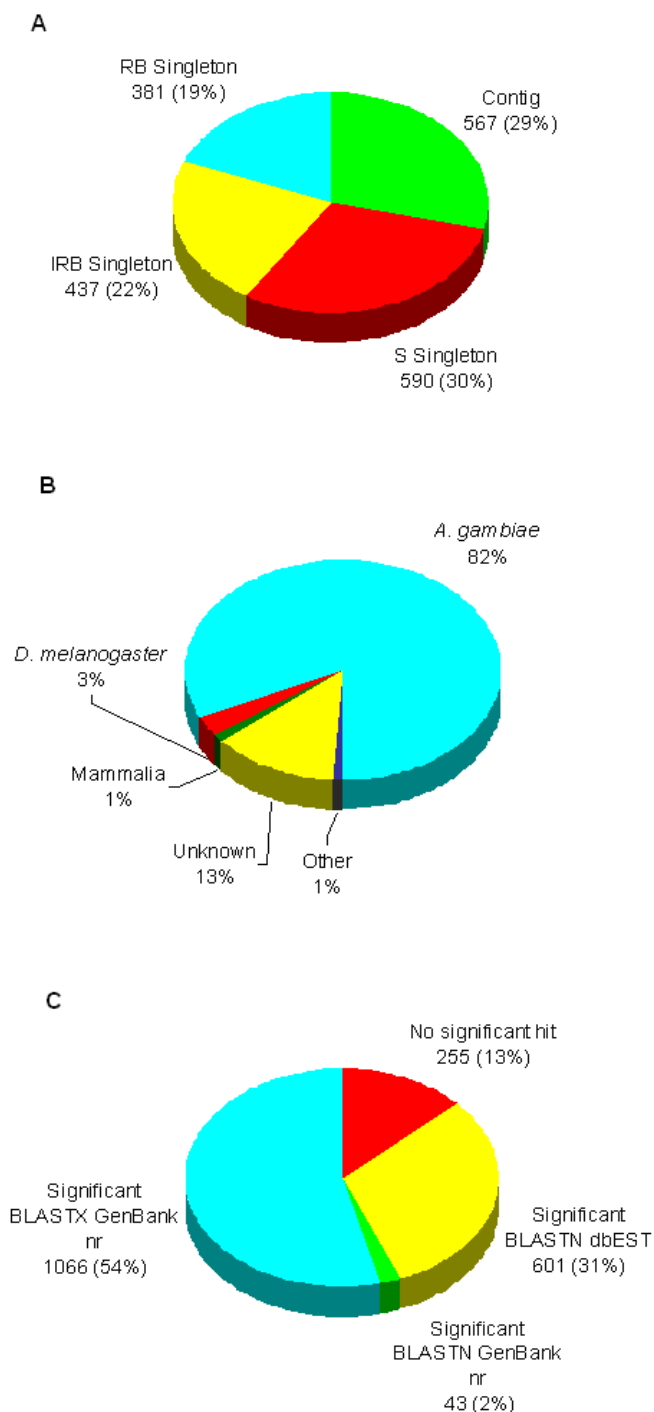
\*Library abbreviations are defined in the text. N = number of clone inserts, ESTs sequenced <sup>†</sup>The mean EST length in nucleotides (nt) is the length following end-trimming, short sequence filtering and EST contig assembly.  
SE = standard error of the mean.

structure, annotation and expression of these genes is necessary for understanding how they are regulated spatially and temporally, and for determining how they function in the mosquito. Large-scale sequencing of cDNA libraries, captures expressed gene products, creating a "molecular snapshot" of the transcriptome. A single sequence read corresponds to the transcript from which it was derived and generates an EST for the underlying gene. Genes can be identified putatively by comparing the derived ESTs with sequences of known annotated genes and gene products.

Large-scale EST sequencing of different cDNA populations provide opportunities for exploration of gene expression under defined biological and environmental conditions. All cells are complex molecular environments regulated by the information in their genes encoding thousands of proteins involved in a multitude of processes. However, only a subset of these genes is actively transcribed at any one time, and in eukaryotes, in any one organ, tissue and cell type. The "digital Northern", an *in silico* form of transcript profiling, can be used to study gene expression by comparing ESTs from clones randomly picked from two or more cDNA libraries created from non-normalized mRNA populations [3-5]. The frequency of any specific sequence should reflect the relative expression level or abundance of that transcript in the libraries [6]. Genes are identified as being differentially expressed using a number of statistical methods [4,7,8]. Finally, the ontology of a gene, the molecular function and biological process in which its product is involved, provide information about the system in which it is expressed.

Infection of the adult female anopheline mosquito with malaria parasites elicits both local and systemic responses from a range of vector organs and tissues. *Plasmodium* infection is also coincident with the ingestion of a blood meal which sets in motion a complex sets of events including digestion and egg production [9-11]. These events involve extensive changes in gene expression in multiple organs, three of which are found in the abdomen, midgut, fat body and ovaries [12-16]. Normal patterns of gene expression in these organs are often significantly further altered in parasitized mosquitoes [1,17,18].

We have investigated genes that are up and down regulated following blood feeding and *Plasmodium berghei* infection of *A. gambiae* females using a direct sequencing approach. Three cDNA libraries were created from the abdomens of sugar-fed, naïve blood-fed, and *P. berghei*-infected females. These whole abdomens contain a multitude of organs, tissues and cell types, and provide an inventory of genes expressed during blood digestion, vitellogenesis and *Plasmodium* infection. ESTs were obtained and their frequencies compared among the 3 libraries to create transcript profiles. EST annotation using existing databases, BLAST tools and gene ontology classifications yielded information on the most dramatic transcriptional responses of these mosquitoes to blood feeding and parasitism. This catalog of abdominal gene expression will contribute to a more global understanding of anopheline physiology and immunity. It will also provide a resource for improving annotation of the *A. gambiae* genome, thus making it more useful for vector biologists



**Figure 1**

Distributions of ESTs among the S, RB and IRB libraries. A total of 1408 singletons, unique transcripts, and 567 multi-EST contigs was examined. A. Distribution of singletons among libraries relative to total contigs. The % of 1975 singletons and contigs is shown in brackets following the number of singletons within each library. B. Taxonomic distribution of all 1975 assembled sequence homologies following BLASTX and BLASTN searches of GenBank nr and a BLASTN search of dbEST. C. Identification of 1965 transcripts sharing *A. gambiae* genomic identity,  $E < 10^{-4}$ .

and scientists studying homologous genes in other organisms. Most importantly, increased understanding of anopheline biology at the molecular level may open new avenues for intervention against malaria transmission.

### Results and discussion

Three unidirectionally cloned cDNA libraries were constructed from mRNA isolated from abdomens of *A. gambiae* females that had been fed on 20% sucrose (S library), on rat blood (RB), or on rat blood infected with *P. berghei* (IRB), and then maintained at 19–20°C for 30 h. To determine the staging of *P. berghei* infections in the susceptible 4arr strain, we chose to count melanized transforming ookinetes and early oocysts in the L-35 refractory strain using the rationale that transforming ookinetes/early oocysts are the infection stages being examined in this study whereas counts of later stage oocysts at 5 or 6 d post infection would be likely to underestimate infection intensity at the experimental times points. We acknowledge that this approach assumes that the L35 and 4arr strains experience similar infection time kinetics and similar invasion rates of the midgut by ookinetes, infection attributes that have not been formally established. At 30 h post infection (PI), the majority of parasites in infected L-35 strain mosquitoes were ookinetes still traversing the midgut or transforming to early oocysts on the basal surface of the midgut: the mean number of parasites per *A. gambiae* midgut at 30 h PBM was  $10 \pm 11$  (standard error of the mean, SEM), rising to a peak of  $24 \pm 23$  by 36 h PBM. A *post hoc* Tukey test showed a significant difference in oocyst numbers between these time points,  $P < 0.05$ . These values for *A. gambiae* infection by *P. berghei* were similar to infection intensities published elsewhere [19,20]. The unamplified S, RB and IRB libraries contained a total of  $3.09 \times 10^6$ ,  $1.22 \times 10^7$ , and  $1.13 \times 10^6$  pfu/ml, respectively. The average insert size of 704 clones randomly picked from all three libraries was calculated to be  $1068 \pm 35$  bp (Table 1).

cDNA inserts from 3264, 1920, and 3456 white plaques randomly picked from the S, RB, and IRB libraries, respectively, were amplified by PCR and sequenced from their 5' ends. The resulting ESTs were filtered based on sequence trace file quality, screened for mitochondrial contamination and assembled into contigs using SeqMan II. Following the initial SeqMan assembly, only high quality non-mitochondrial sequences >100 bp, corresponding to 1736, 1216, and 1772 ESTs from each of the S, RB and IRB libraries, respectively, were analyzed further. Their average length was  $455 \pm 8$  nt (Table 1). These 4724 high quality ESTs assembled into a total of 1989 contigs and singletons. Sequence identity searches showed that 4 contigs and 10 singletons were of non-mosquito origin: these sequences shared no identity with the *A. gambiae* genome [21] but did with proteins such as rat alpha and beta

hemoglobin chains. When these contaminants had been removed, a total of 4541 high quality ESTs remained for further analysis, 1727 from S, 1145 from RB, and 1669 from IRB. They assembled into 1975 consensus sequences, 567 contigs and 1408 singletons (Table 1, Figure 1A). Each was given a unique assembled sequence (AS) number. These EST sequences can be accessed through the NCBI EST database, dbEST [22].

The EST assembly may have estimated the number of unique genes in the libraries inaccurately due to sequencing errors, sequence polymorphisms, alternate splicing of transcripts, and lack of overlap of 3' and 5' sequences representing the same gene product [22,23]. To evaluate this inaccuracy, we compared sequences sharing the same top BLAST hit by aligning them with the nucleotide sequences of their predicted genes using CLUSTAL W [24,25] and its default parameters. Of the 974 assembled sequences that shared identity with predicted *A. gambiae* proteins, 65 aligned with a protein already represented in the data set. When assembled sequences identified as being from the same gene failed to align, they failed to do so mainly because their sequences did not overlap, i.e., they were from different parts of a gene. Alternative splicing of transcripts and sequence dissimilarity in excess of contig assembly thresholds also contributed to these alignment failures. We estimate that the total number of assembled sequences may have overestimated the total number of unique genes by approximately 4%.

### Sequence identity searches

The 1975 assembled *A. gambiae* EST contig and singleton sequences were distributed among the three libraries as shown in Figure 1A. They were searched against nucleotide and protein databases for identity to known genes and proteins using BLAST algorithms. 82% of them showed highest identity to predicted or identified *A. gambiae* genes, 3% to *Drosophila melanogaster* genes, 1% to mammalian genes, 1% to other taxonomic groups, while 13% remained unknown after this analysis (Figure 1B) BLASTN alignments identified 1965 consensus sequences with significant identity ( $<E^{-4}$ ) to the *A. gambiae* genome assembly and 10 that did not (Figure 1C). The 1965 consensus sequences identified as being derived from *A. gambiae* are composed of 1710 sequences which shared identity with publicly available nucleotide and/or amino acid sequences from other organisms and 255 sequences which showed no identity to any sequence in the GenBank nr and dbEST databases. 95% (1620/1710) of the *A. gambiae*-derived consensus sequences exhibited identity with *A. gambiae* protein sequences predicted *in silico* from the genome sequence and sequences submitted to databases prior to genome sequencing (data not shown). These included 974 sequences with identity to predicted proteins, 48 with identity to protein sequences derived

from other submitted gene sequences, and 598 with identity to ESTs. It is surprising that no *Plasmodium* ESTs were identified.

Only 49.6% (974/1965) of assembled sequences sharing identity with the *A. gambiae* genome also shared identity with predicted *A. gambiae* proteins. The remaining 991 sequences may be derived from 5' or 3' untranslated regions (UTRs), may lack an ORF due to frame shift errors occurring during cloning or sequencing, may be 5' truncated, or may be just too short to be identified, despite the 100 nt cutoff. They may also represent novel genes. Holt *et al.* [1] concluded that more than 1300 genes might have escaped prediction in the first annotation of the *A. gambiae* genome. The current gene number prediction is 14,707 [2], an increase of more than 1000 from the original estimate. Computational techniques may inaccurately predict genes by missing exons derived solely from promoters, or that are due to alternative splicing of transcripts or to use of non-canonical splice sites, alternative translational initiation and/or polyadenylation sites [26]. Thus, the majority of ESTs generated in this study may enhance gene prediction in the *A. gambiae* genome through refinement of existing gene models and providing evidence for new ones.

By definition, ESTs are generally short sequences of approximately 300–500 nt derived from transcripts [27]. The sequences assembled in this study had an average length of 455 nt and many could have consisted mainly of 5'- or 3'-UTRs. The average lengths of the 5'- and 3'-UTRs in Release 3 of the *Drosophila* genome were 265 and 442 nt respectively, and the average ratio of their length/coding sequence was 0.75 [28]. Accordingly, BLASTN sequence identity searches were performed against dbEST to annotate assembled sequences that might have lacked a predicted ORF because they consisted mainly of either a long 5'-UTR, or of 3'-UTR resulting from 5' truncation of transcripts during cDNA library construction. dbEST contains more than 100,000 *A. gambiae* ESTs, the majority of which were generated from large-scale sequencing of two non-normalized cDNA libraries constructed from non blood-fed and blood-fed whole adult females [1]. 601, or 66.9% of the 898 unique gene products sharing identity with the genome but not identified by BLASTX search of GenBank Nr were identified by this BLASTN search of dbEST. 99.5% of these shared sequence identity with at least one *A. gambiae* EST. After completion of all identity searches, 255 assembled sequences still failed to show significant identity with the ESTs in dbEST or with the predicted proteins in GenBank Nr, though they did with the *A. gambiae* genome. These appear to be truly novel.

Since *ab initio* gene prediction programs used to analyze genomes can only identify open reading frames (ORFs),

cDNA sequences provide an essential tool for properly validating gene identification and annotation. Misra *et al.* [28] reported that reannotation of the *Drosophila* genome following Release 3 resulted in much-improved prediction of alternatively spliced transcripts and annotation of UTRs due to the increased number of ESTs and cDNAs available. The reannotation resulted in changes to 85% of gene models, including major changes in 40% of predicted proteins, without significantly affecting the number of genes predicted. Since the untranslated regions of transcripts contain sequences influencing transcript fates including subcellular localization and mRNA turnover, as well as *cis*-regulatory information, a database of nucleotide sequences corresponding to predicted transcripts that includes UTRs may provide a better tool for EST and genome annotation than a database of predicted proteins. Most importantly, however, accurate identification of UTRs and alternative patterns of intron splicing in the *A. gambiae* genomic sequence that can be obtained through EST projects such as this one are necessary for ultimately understanding gene regulation at the post-transcriptional level.

#### Functional annotation of ESTs

The adult female mosquito abdomen contains several complex organs and tissues including the midgut, the ovaries and the fat body. These function in the normal processes of blood meal digestion and egg production, as well as in responses to infection by and defense against pathogens. Functional annotation of the abdominal genes represented in the three cDNA libraries was performed to gain insight into the physiological events required for reproduction and the pathological ones induced by infection with *Plasmodium*. Molecular function and biological process were assigned to the consensus sequences based on sequence similarity to known genes and proteins and to the existing gene indices for *A. gambiae* and *D. melanogaster*, TIGR Gene Index and GadFly, respectively.

The 1975 gene products predicted after EST clustering were categorized into 8 major biological processes with 34 subdivisions. A 9<sup>th</sup> category, Unknown, represents gene products with no ascribable function. The library-specific results of these functional assignments are in Table 2. More detailed annotation of all 1975 gene products are provided [see Additional file 1]. For all three libraries taken together, the largest category, representing 1329 gene products or 67.3% of the total, remains the "Unknown". The three next most numerous categories are "Metabolism", 441 unique transcripts, 23.3% of the total; "Protein Synthesis", 418, 21.2%; and "Egg Production", 81, 4.1%. The largest represented subdivision is "Translation". Perhaps unexpectedly, the genes likely to be involved in egg production are represented in similar proportions among the three libraries. While it is possible

**Table 2: BIOLOGICAL PROCESSES OF GENES REPRESENTED IN LIBRARIES\***

|  | Total S          | Total RB         | Total IRB        |
|--|------------------|------------------|------------------|
| Metabolism   |                  |                  |                  |
| Simple/Complex Carbohydrate Metabolism and Transport                   | 21 (28)†         | 15 (17)          | 13 (15)          |
| Oxidative Phosphorylation  | 41 (64)          | 38 (61)          | 30 (69)          |
| Lysosomal Enzymatic Digestion  | 1 (2)            | 0                | 4 (4)            |
| Protein Digestion  | 13 (63)          | 7 (49)           | 11 (63)          |
| Protein Modification, Metabolism, Transport and Localization           | 42 (54)          | 38 (49)          | 39 (63)          |
| Amino Acid and Derivative Metabolism and Transport                     | 10 (13)          | 6 (10)           | 16 (20)          |
| Nucleobase/Nucleoside/Nucleotide/Nucleic acid Metabolism and Transport | 9 (21)           | 10 (12)          | 13 (19)          |
| Fatty Acid/Lipid Metabolism and Transport                              | 9 (11)           | 8 (8)            | 5 (6)            |
| Vitamin/Vitamin Derivative/Cofactor Metabolism and Transport           | 2 (3)            | 2 (3)            | 2 (2)            |
| Xenobiotic Metabolism and Transport                                    | 5 (5)            | 7 (8)            | 5 (6)            |
| Pigment Synthesis and Transport  | 1 (1)            | 2 (2)            | 4 (5)            |
| Other  | 3 (4)            | 6 (6)            | 3 (3)            |
| <b>Total</b>   | <b>157 (269)</b> | <b>139 (225)</b> | <b>145 (275)</b> |
| Transport  |                  |                  |                  |
| Ion Transport  | 10 (11)          | 14 (16)          | 12 (14)          |
| Receptor-mediated Endocytosis  | 7 (10)           | 4 (4)            | 6 (11)           |
| <b>Total</b>   | <b>17 (21)</b>   | <b>18 (20)</b>   | <b>18 (25)</b>   |
| Protein Synthesis  |                  |                  |                  |
| Transcription and mRNA Processing                                      | 17 (23)          | 19 (20)          | 19 (21)          |
| Translation  | 108 (436)        | 97 (325)         | 115 (482)        |
| Protein Folding  | 14 (24)          | 19 (21)          | 10 (13)          |
| <b>Total</b>   | <b>139 (483)</b> | <b>135 (366)</b> | <b>144 (516)</b> |
| Cellular Processes   |                  |                  |                  |
| Cell Cycle   | 14 (17)          | 10 (13)          | 6 (11)           |
| Cellular Proliferation   | 4 (4)            | 2 (2)            | 3 (3)            |
| Chromatin Assembly/Disassembly   | 4 (6)            | 5 (7)            | 5 (10)           |
| Apoptosis  | 2 (2)            | 2 (2)            | 5 (7)            |
| Senescence   | 2 (3)            | 0                | 1 (2)            |
| Viral Life Cycle   | 1 (1)            | 1 (1)            | 0                |
| <b>Total</b>   | <b>27 (33)</b>   | <b>20(25)</b>    | <b>20 (33)</b>   |
| Egg Production   |                  |                  |                  |
| Vitellogenesis/Oogenesis/Embryogenesis                                 | 29 (33)          | 24 (40)          | 27 (60)          |
| Melanization   | 0                | 1 (1)            | 0                |
| <b>Total</b>   | <b>29 (33)</b>   | <b>25 (41)</b>   | <b>27 (60)</b>   |

**Table 2: BIOLOGICAL PROCESSES OF GENES REPRESENTED IN LIBRARIES\*** (Continued)

|  |                  |                  |                  |
|--|------------------|------------------|------------------|
| Cellular Communication                   |                  |                  |                  |
| Signal Transduction                      | 11 (14)          | 5 (7)            | 6 (12)           |
| Cell-cell Signaling                      | 4 (7)            | 4 (4)            | 2 (3)            |
| <b>Total</b>                             | <b>15 (21)</b>   | <b>9 (11)</b>    | <b>8 (15)</b>    |
| Intra-/Extra-cellular Architecture       |                  |                  |                  |
| Maintenance                              |                  |                  |                  |
| Structural                               | 8 (35)           | 11 (24)          | 6 (30)           |
| Muscle-related                           | 6 (6)            | 4 (5)            | 7 (8)            |
| Cell Adhesion                            | 3 (3)            | 3 (3)            | 1 (1)            |
| Cytoskeleton Organization and Biogenesis | 4 (6)            | 12 (17)          | 12 (18)          |
| <b>Total</b>                             | <b>21 (50)</b>   | <b>30 (49)</b>   | <b>26 (57)</b>   |
| Response to Stress/External Stimulus     |                  |                  |                  |
| Response to Oxidative Stress             | 7 (9)            | 8 (11)           | 14 (29)          |
| Immune/Defense Response                  | 11 (29)          | 9 (15)           | 14 (20)          |
| Chemosensory Perception                  | 3 (3)            | 1 (1)            | 1 (1)            |
| <b>Total</b>                             | <b>21 (41)</b>   | <b>18 (27)</b>   | <b>29 (50)</b>   |
| Unknown                                  |                  |                  |                  |
| <b>Total</b>                             | <b>573 (776)</b> | <b>326 (415)</b> | <b>430 (663)</b> |

\*This analysis does not take into account that some genes are represented in more than 1 library or that more than one biological process may have been ascribed to particular genes.

†Number of Contigs containing one or more ESTs (total ESTs/genes).

that this result is an artifact created by timing the mRNA samples used for library creation to 30 hr PBM at 19°C, a time close to the peak of vitellogenic activity at this temperature, the gene products involved in oogenesis may also play other roles in the life cycle of the mosquito. For instance, some of the gene products in the Toll pathway, a signaling cascade that controls dorsal-ventral patterning of the *Drosophila* embryo during development [29], are also important in the induction of several immune-related peptides [30-33].

The first large scale studies to identify genes involved in mosquito responses to *Plasmodium* infection relied on cDNA libraries prepared from bacteria-challenged mosquito tissues [34,35]. Three more recent studies have reported direct screens for *Anopheles* genes responding to *Plasmodium* infection [17,36,37]. Our IRB library has provided an additional opportunity to study transcripts whose abundance may be regulated by infection with *P. berghei*. It showed an increase in the proportion of gene products present in biological processes likely to be responses to parasite infection, including responses to oxidative stress and immunity-related defenses. Some transcript profiles looked at here are based on the comparison of ESTs that vary by only one unit. Though this may be considered to be of limiting value, it is important to note that most profiles are consistent with previous stud-

ies and are discussed individually. In addition, the tissues used here are composed of whole abdomens, which contain gut, blood cells, fat body, epidermis, ovaries and other cell and tissue types. Differences of gene expression patterns between this and previous studies may relate to the somewhat different tissues that were assayed.

Reactive oxygen species may be generated through the activities of nitric oxide synthase (NOS) and peroxidases [38,39]. Transcripts encoding number of enzymes involved in regulating the cellular oxidative environment were identified in all three libraries. These include multiple glutathione S-transferases (GSTs), peroxidases, and peroxiredoxin (Table 3), but not nitric oxide synthase (NOS). Failure to identify inducible NOS transcripts at least in the IRB library may relate to the abundance and or timing of its transcription though *P. berghei* invasion induced NOS both systemically and locally in the midgut in *Anopheles stephensi* 24–48 hours post infection (PI) [40]. In parasite-damaged midgut cells, the increase in NOS levels was concurrent with other morphological changes associated with apoptosis [41]. NOS is known to be activated transcriptionally in *A. gambiae* within 22–24 hr PI with *P. berghei* [42,43]. Activation is mainly in the midgut, as expected for this time period.

**Table 3: OXIDATION AND STRESS RESPONSE GENES**

| AS   | # S ESTs | # RB ESTs | # IRB ESTs | Total # ESTs | Genome Scaffold | Start   | End     | E value | Blast Type | Accession | Blast Hit  | E value   | Putative Identity                         |
|------|----------|-----------|------------|--------------|-----------------|---------|---------|---------|------------|-----------|--|-----------|---|
| 34   | 2        | 1         | 5          | 8            | 8898            | 1665469 | 1665984 | 0       | X          | EAA09273  | agCPI4153  | 1.00E-111 | glutathione S-transferase D11             |
| 35   | 0        | 4         | 3          | 7            | 8880            | 1407780 | 1408378 | 0       | X          | Q93113    | glutathione S-transferase 1-6                    | 1.00E-110 | glutathione S-transferase 1-6             |
| 172  | 2        | 0         | 2          | 4            | 8816            | 4859270 | 4859540 | 1E-130  | X          | EAA05108  | agCP6896   | 7E-65     | Glutaredoxin (GRX1)                       |
| 307  | 0        | 1         | 0          | 1            | 8986            | 5275996 | 5276383 | 1E-173  | X          | EAA00516  | agCP9336   | 8E-46     | copper ion transporter                    |
| 620  | 0        | 0         | 2          | 2            | 8984            | 9812533 | 9812822 | 1E-161  | X          | AAL58538  | glutathioneS-transferase E3                      | 1E-73     | glutathione S-transferase E3              |
| 622  | 0        | 0         | 1          | 1            | 8980            | 538441  | 538745  | 1E-172  | X          | AAF68382  | thioredoxin I                                    | 5E-49     | thioredoxin I                             |
| 704  | 0        | 0         | 1          | 1            | 8880            | 1406678 | 1407026 | 0       | X          | EAA08605  | agCP2490   | 4E-36     | glutathione S-transferase 1-6             |
| 909  | 0        | 1         | 0          | 1            | 8984            | 9810450 | 9810635 | 1E-101  | X          | AAG45163  | glutathioneS-transferase E1                      | 2E-20     | glutathione S-transferase 3-1             |
| 910  | 0        | 0         | 1          | 1            | 8984            | 9809004 | 9809209 | 1E-108  | X          | AAG45164  | glutathioneS-transferase E2                      | 6E-28     | glutathione S-transferase E2              |
| 931  | 0        | 1         | 0          | 1            | 8933            | 621123  | 621383  | 1E-145  | X          | EAA09899  | agCPI1759  | 1E-73     | Manganese Superoxide dismutase I          |
| 1012 | 0        | 0         | 4          | 4            | 8849            | 2536282 | 2536794 | 0       | X          | EAA07169  | agCPI0692  | 6E-80     | Superoxide dismutase 3-D (Cu,Zn)          |
| 1041 | 1        | 0         | 2          | 3            | 8849            | 1674668 | 1675054 | 0       | X          | EAA07207  | agCPI0713  | 1E-115    | glutathione S-transferase S1-2            |
| 1042 | 0        | 1         | 2          | 3            | 8986            | 8102109 | 8102313 | 1E-105  | X          | EAA00332  | agCP9864   | 2E-90     | l-cys peroxiredoxin TPX4                  |
| 1065 | 0        | 0         | 3          | 3            | 8804            | 83740   | 83869   | 6E-45   | N          | AJ284424  | 4A3B-AAW-E-09-F A. gambiae immune competent 4A3B | 2E-51     | 2-Cys thioredoxin peroxidase TPX2         |
| 1482 | 0        | 0         | 1          | 1            | 8984            | 9812211 | 9812365 | 2E-74   | X          | AAL58538  | glutathioneS-transferase E3                      | 8E-06     | glutathione S-transferase E3 2-Cys        |
| 1684 | 0        | 1         | 0          | 1            | 8804            | 84431   | 84661   | 1E-117  | X          | EAA03855  | agCPI990   | 2E-35     | thioredoxin peroxidase TPX2 thioredoxin I |
| 1909 | 1        | 0         | 0          | 1            | 8980            | 539114  | 539324  | 1E-109  | N          | AJ283949  | 4A3B-AAH-C-12-F A. gambiae immune competent 4A3B | 1E-107    |   |
| 2033 | 1        | 0         | 0          | 1            | 8898            | 1672566 | 1673171 | 0       | X          | EAA09147  | agCPI4131  | 1E-120    | glutathione S-transferase D3              |
| 2078 | 1        | 0         | 0          | 1            | 8984            | 9807310 | 9807705 | 0       | X          | AAL59653  | glutathioneS-transferase E4                      | 4E-71     | glutathione S-transferase E4              |
| 2127 | 0        | 1         | 0          | 1            | 8986            | 8102593 | 8102854 | 1E-13i6 | X          | EAA00332  | agCP9864   | 6E-31     | peroxiredoxin TPX4                        |
| 2156 | 0        | 0         | 1          | 1            | 8880            | 3392447 | 3392665 | 1E-120  | X          | EAA08586  | agCP2356   | 2E-43     | thioredoxin                               |
| 2296 | 0        | 0         | 1          | 1            | 8807            | 1924679 | 1924904 | 1E-110  | X          | EAA03983  | agCP3166   | 5E-75     | thioredoxin peroxidase TPX3               |
| 2334 | 1        | 0         | 0          | 1            | 8880            | 3614672 | 3614912 | 1E-124  | X          | EAA08535  | agCP2389   | 3E-58     | glutathione peroxidase 2-A, 2-B           |



Although GSTs are most often associated with the detoxification of xenobiotics they are also involved in a number of cellular processes including protection from oxidative stress and apoptosis [44]. In addition to regulating thioredoxin, GSTs regulate the redox state of pro-apoptotic proteins [29,45]. These redox-modulating enzymes may be inducing oxidative stress either in response to *Plasmodium*-induced inflammation or to apoptosis of mosquito tissues. Thus, the generation and modulation of reactive oxygen species by multiple enzymes induced later in *A. gambiae* infection, which we now observe, may regulate or limit oocyst development.

Transcripts of 25 genes involved in insect immune responses were identified (Table 4). Previously characterized *A. gambiae* immune-related peptides included AS 1286, gram negative [bacteria] binding protein, GNBp; AS 2178, cecropinA, CecA; and AS 1197, cecropinB, CecB. Two ficolins, AS 1364 and AS 1922, and four lysozymes AS 32, AS 221, AS 659 and AS 2030, were also identified. GNBp, has been linked with the immune response to *Plasmodium* infection, had already been found to be induced in the 20–30 h following *A. gambiae* infection with *P. berghei* [35,46].

Cecropins are small, basic peptides which cause lysis of gram negative and some gram positive bacteria by forming pores in their inner cell membranes (see [47] for review). They have been found in a wide variety of insects, including many vectors of parasitic diseases. The *A. gambiae* genome contains either 3 or 4 cecropin genes [48,49]. The *CecA* gene product, Cecropin A, is induced by *Plasmodium* during the early stages of infection [48,50]. The divergently transcribed *CecA* and *CecB* genes are both up-regulated in an *A. gambiae* cell line after challenge with lipopolysaccharide and heat-inactivated bacteria [49]. This induction is regulated by a currently uncharacterized NF B-class transcription factor. Since these two *Cec* genes were identified only in the IRB library, they may be involved specifically in anti-*Plasmodium* responses.

Ficolins are carbohydrate-binding proteins related to collectins, a class of innate immunity lectins involved in the phagocytic ingestion of apoptotic cells [51–53]. In vertebrate innate immune responses, ficolins initiate the lectin pathway of complement activation [54]. The two ficolins we identified showed highest amino acid identity to the *Drosophila* Ficolin 2 precursor, but they correspond to different *A. gambiae* gene products located on different chromosomes *in silico*. Since they were found only in the IRB library, it is likely that they are involved in *Plasmodium* recognition prior to immune activation unless recognition, which appears to involve multiple pattern recognition proteins, is progressive.

The lysosome contributes to cellular maintenance through involvement in autophagy and to immunity through protease-mediated degradation of phagocytosed substances and apoptosis-like programmed cell death [55]. Since all lysosomal enzymes identified in this study except for AS 32 were found solely in the IRB library, the increase in lysosomal proteases in the IRB library may be indicative of phagocytic, inflammatory, and/or apoptotic responses to *Plasmodium* infection. We identified transcripts of four different lysozyme genes among our ESTs. AS 32, for which we had 9 ESTs distributed in all three libraries, showed clear evidence of alternative splicing. This gene corresponds to the previously characterized basic lysozyme gene, ENSANGG00000019898 [56]. They had concluded that this gene was expressed much more abundantly in sugar-fed than in blood-fed *A. gambiae*. However, the primers they used for their RT-PCR would also have amplified transcripts from a more recently identified lysozyme gene, ENSANGG00000015399, agCP3675, (AD, unpublished), thus potentially confounding their results. AS 221, AS 659 and AS 2030, the three other lysozyme EST sequences corresponding to ENSANGG00000015399 (agCP3675), ENSANGG00000015906 (agCP3967), and ENSANGG00000015950 (agCP3164) were all found at one EST each in the IRB library. Thus, among abdominal cDNAs expressed following an IRB, we have identified transcripts representing 4 of the 5 lysozyme genes and 5 of the 7 potential lysozyme transcripts encoded in the *A. gambiae* genome. It is interesting that all of the *A. gambiae* lysozyme genes are located close together in the same region of chromosome 2L and that their promoters all contain potential binding sites for NFB-like transcription factors, as would be expected for pathogen-induced transcription. Since AS 221 has also been identified as being induced more than two-fold at 48 h after an uninfected blood meal in our microarray study [12], it is also possible that different lysozymes or combinations of lysozymes may act as antibacterial agents in *A. gambiae* following sugar and blood feeding.

Lysozymes were not identified as potential immunity-related proteins by Christophides *et al.* [48]. However, Hultmark [57] suggested that lysozyme may well be an immune protein that acts synergistically with cecropin to release microbial surface components, since its activity increases in concert with cecropin activity in bacteria-challenged silk worm larvae [58]. Therefore it is possible that multiple *A. gambiae* lysozymes may act together with cecropin(s) in an anti-*Plasmodium* response. The overall complexity of lysozyme genes, transcripts, and potential induction patterns in *A. gambiae* suggest that their roles as immune mediators deserve additional study in this mosquito.

**Table 4: IMMUNE-RELATED/DEFENSE RESPONSE GENES**

| AS   | # S ESTs | # RB ESTs | # IRB ESTs | Total # ESTs | Genome Scaffold | Start    | End      | E value | Blast Type | Accession | Blast Hit  | E value | Putative Identity                                |
|------|----------|-----------|------------|--------------|-----------------|----------|----------|---------|------------|-----------|--|---------|--|
| 28   | 17       | 3         | 0          | 20           | 8987            | 14943052 | 14943642 | 0       | X          | EAA01687  | agCPI1956  | 1E-142  | serine protease                                  |
| 32   | 2        | 3         | 4          | 9            | 8807            | 3383097  | 3383307  | 1E-116  | X          | Q17005    | Lysozyme precursor (1,4-beta-N-acetylmuramidase) | 5E-67   | lysozyme precursor (1,4-beta-N-acetylmuramidase) |
| 68   | 1        | 0         | 0          | 1            | 8794            | 217313   | 217497   | 1E-100  | X          | EAA02509  | agCPI1665  | 3E-13   | serine protease 14D2                             |
| 126  | 1        | 0         | 0          | 1            | 8964            | 2322618  | 2322792  | 2E-86   | X          | EAA12171  | agCPI0937  | 1E-25   | TEP3   |
| 153  | 2        | 0         | 0          | 2            | 8986            | 8953596  | 8953709  | 4E-45   | N          | BM621296  | 17000687446469<br>A.Gam.ad.c DNA1                | 2E-63   | fat-spondin                                      |
| 180  | 1        | 0         | 1          | 2            | 8859            | 3227518  | 3227913  | 0       | X          | EAA07758  | agCPI1111  | 4E-70   | signal transducer                                |
| 221  | 0        | 0         | 1          | 1            | 8807            | 3388356  | 3388647  | 1E-158  | X          | EAA04406  | agCP3675   | 8E-79   | lysozyme c-8                                     |
| 418  | 1        | 0         | 3          | 4            | 8944            | 1827235  | 1827637  | 0       | X          | EAA10153  | agCPI5402  | 3E-64   | gambicin   |
| 659  | 0        | 0         | 1          | 1            | 8807            | 3140140  | 3140290  | 5E-80   | X          | EAA04458  | agCP3967   | 3E-56   | lysozyme c-4                                     |
| 804  | 0        | 1         | 0          | 1            | 8960            | 17845546 | 17845795 | 1E-139  | X          | CAA09389  | ICHIT protein                                    | 5E-31   | ICHIT  |
| 832  | 1        | 0         | 1          | 2            | 8811            | 1718220  | 1718532  | 1E-151  | N          | BM635649  | 17000687559053<br>A.Gam.ad.c DNA1                | 3E-64   | AgToll   |
| 995  | 0        | 2         | 2          | 4            | 8960            | 23616    | 23997    | 0       | X          | EAA11001  | agCP5701   | 1E-30   | serine protease                                  |
| 1049 | 0        | 2         | 0          | 2            | 8944            | 1995046  | 1995300  | 1E-138  | X          | EAA10138  | agCPI5205  | 1E-78   | TEP12  |
| 1095 | 1        | 1         | 0          | 2            | 8960            | 16595875 | 16596267 | 0       | X          | EAA11334  | agCP6381   | 7E-30   | signal transducer                                |
| 1120 | 0        | 1         | 1          | 2            | 8986            | 8676449  | 8676732  | 1E-159  | X          | EAA00414  | agCP9557   | 5E-31   | serine protease                                  |
| 1197 | 0        | 0         | 1          | 1            | 8847            | 1123621  | 1123927  | 1E-173  | X          | EAA06859  | agCP7366   | 3E-25   | Cecropin B                                       |
| 1286 | 0        | 0         | 1          | 1            | 8898            | 2844394  | 2844678  | 1E-149  | X          | EAA09116  | agCPI4093  | 4E-28   | GNBPB1   |
| 1364 | 0        | 0         | 1          | 1            | 8948            | 918161   | 918491   | 1E-166  | X          | EAA10406  | agCP2049   | 2E-54   | ficolin  |
| 1616 | 0        | 1         | 0          | 1            | 8980            | 7957050  | 7957284  | 1E-128  | X          | CAB90818  | serine protease                                  | 7E-79   | serine protease                                  |
| 1701 | 0        | 1         | 0          | 1            | 8794            | 212210   | 212531   | 0       | X          | AAB62929  | serine protease 14D                              | 2E-56   | serine protease 14D                              |
| 1922 | 0        | 0         | 1          | 1            | 8816            | 1069890  | 1070396  | 0       | X          | EAA05160  | agCP6864   | 2E-55   | ficolin  |
| 1997 | 1        | 0         | 0          | 1            | 8859            | 9665976  | 9666551  | 0       | N          | AF444782  | AgToll9  | 4E-06   | AgToll9  |
| 2030 | 0        | 0         | 1          | 1            | 8807            | 3137120  | 3137454  | 0       | X          | EAA04667  | agCP3164   | 3E-82   | Lysozyme c-7                                     |
| 2038 | 1        | 0         | 0          | 1            | 8975            | 72415    | 72691    | 1E-155  | N          | AJ420785  | spi21F gene                                      | 1E-71   | serpin spi21F                                    |
| 2178 | 0        | 0         | 1          | 1            | 8847            | 1121909  | 1122047  | 6E-73   | X          | EAA06858  | agCP7503   | 6E-21   | Cecropin A                                       |

AS 418, Gambicin, transcripts were found in both the S and IRB libraries, insignificantly more in the latter (Table 4). The gambicin gene encodes a 6.8 kDa antimicrobial peptide unique to *A. gambiae* [59]. Gambicin transcripts were found primarily, but not exclusively in the anterior midguts of both sugar-fed and blood-fed adult female mosquitoes. RT-PCR suggested that its transcription is induced slightly above basal levels at 30 h PBM and by about 3-fold by 24 h post infection, at which point *P. berghei* ookinetes are invading the midgut epithelium.

The association of AS804, ICHIT, a galectin with chitin-like domains, only with the RB library was unexpected. This is because ICHIT transcripts were found to be abundant in the midguts of sugar-fed adult female mosquitoes and only weakly induced in *A. gambiae* midguts 24 h PI with *P. berghei* [42].

Two Toll receptor gene transcripts were identified (Table 4). The Toll signal transduction pathway is involved both in insect immune responses, and, in *Drosophila*, at least, in specification of the dorsal-ventral body axis during embryogenesis. While there are 11 known Toll receptor genes in *A. gambiae* [48], the expression of only 4 has been characterized [60]. One, AS 832, corresponding to *AgToll*, was found only in the S and IRB libraries. Luna *et al.* [60] demonstrated that this gene is abundantly expressed in ovaries, and not at all in midgut. They also found that this gene is only weakly induced by bacterial challenge. This gene is actually duplicated in the *A. gambiae* genome as *AgToll1* and *AgToll1B*, both of which are closely related to *AgToll5A* and *AgToll5B*, as well as to *D. melanogaster Toll*, *DmToll1*, encoding the receptor mediating body axis formation, and to *DmToll5* [32].

Consequently, despite the fact that AS 832 was not found in the RB library, its primary function is more likely to be in embryogenesis than in defense. AS 1997, *AgToll 9*, was found only in the S library. It is abundantly and specifically expressed in midgut during multiple developmental stages [60], and is weakly induced in larvae by bacterial challenge, though it has not been tested for induction in *Plasmodium*-infected mosquitoes. It and its *Drosophila* ortholog, *DmToll9*, are most closely related to mammalian TLR genes, and may well be involved in immunity functions.

Serine proteases (SPs) and serine protease inhibitors (serpins, SRPNs) the inhibitors of SPs, function in multiple processes, of which immunity is only one. AS 68, Sp 14D2; AS 28, SP G13; AS 153, fat-spondin; AS 2038, a serpin transcribed from the spi21F/SRPN10 locus, were all unique to the S library. The majority, 20/31, of SP transcripts identified, 17 of them derived from AS 28, SP G13, were from the S library (Table 4). Sp14D2 is abundantly

expressed in adult females at 4 days post eclosion and is induced only slightly following bacterial challenge and *P. berghei* infection [61,62]. It may not be involved in immune responses. The adult gut-specific non-trypsin SP G13 is found in both blood fed and non-blood fed females, and is also immune responsive in bacterially challenged larvae [34]. The SRPN10 gene encodes four alternatively spliced transcripts which are differentially expressed in the midgut during development and following microbial challenge [63]. We were unable to determine which one of these transcripts corresponds to AS 2038.

Four SPs were found in the RB library in addition to the previously described SP G13 (Table 4). These are AS 170, Sp14D1; AS 1616, CLIPB15; AS 1120, CLIPA7; and AS 995, agCP5701, previously predicted only computationally. Sp14D1 transcripts are expressed constitutively in multiple mosquito stages [62]. In adult females, they are expressed in the ovary and fat body, but not in the midgut. The Sp14D1 gene is induced by 24 h PBM and after a bacterial challenge. Therefore this serine protease may have roles both in development and in immunity. CLIPB15 showed significant induction following bacterial challenge and during malaria parasite invasion [48]. CLIPA7 and agCP5701 were found in both the RB and IRB libraries. CLIPA7 has not yet been characterized. However, the agCP5701 sequence shares identity with Ssp3, a serine protease recently identified in the hematophagous fly, *Stomoxys calcitrans* [64]. This serine protease colocalizes with defensin and is thought to activate it.

It is well known that insect immunity-effector genes are not necessarily induced by pathogen challenge, but instead, may be constitutively expressed in situations in which pathogens could be encountered. Certainly, defensin and lysozyme are constitutively expressed during blood feeding in ticks [65-68]. Thus several of the SP genes represented in our libraries may have immune functions, even though none is unambiguously upregulated following *P. berghei* infection.

Fat-spondin is a serine protease inhibitor of the Kunitz family that is down-regulated following septic injury. It is also regulated by Spaetzle, the activator of the Toll pathway in *Drosophila* [30,31]. Therefore mosquito fat-spondin may regulate immune responses during nectar feeding.

#### **Digital Northern and verification of selected gene expression patterns by qRT-PCR**

The complement of genes expressed in a cell or tissue represents its transcriptome [69]. Transcriptome analysis has been approached in several ways. Okubo *et al.* [5] demonstrated that a non-normalized, non-amplified cDNA

library can faithfully represent the mRNA population in a tissue, and that such a resource could be used to explore the diverse array of active genes and their mRNA abundances in a tissue. Thus, the "digital Northern" became one of the first genome scale analytical methods employed in the study of gene expression. Three classes of mRNA transcript abundance were recognized based on reassociation kinetics; high abundance (5–15 mRNA species at ~10,000 copies per cell), intermediate abundance (500 species at ~300 copies per cell) and low abundance (10,000 different species at 1–15 copies per cell) [70-72]. Lee *et al.* [73] showed that in a random sample of approximately 3000 ESTs from a single cDNA library, > 99% of highly abundant transcripts, 85% of intermediate abundance transcripts, and < 5% of low abundance transcripts should be represented at least once. Hwang *et al.* [4] demonstrated that differentially expressed genes could be identified digitally even in small data sets although the analysis would necessarily be restricted to more abundantly expressed transcripts, those that are expressed at a frequency of greater than 1 transcript in 800.

Since our three cDNA libraries were neither normalized nor amplified, the number of ESTs in a contig should represent the abundance of the corresponding transcript in the libraries. Therefore we used the digital Northern technique to identify genes differentially expressed in the abdomens of *A. gambiae* females in response to a sugar meal, to a blood meal and to a blood meal containing infective malaria parasites. Although our investigation is necessarily limited to the analysis of moderately and highly abundant mRNAs, it offers a unique opportunity to identify a diverse array of transcripts and to examine some of the greater fluctuations in transcript abundance between and among libraries.

Although large EST sequencing experiments are not repeated and consequently do not exhibit variation, the number of ESTs corresponding to particular transcribed genes approximates a Poisson distribution [74]. Consequently two statistics, the Audic and Claverie Statistic and the R statistic, both based on a Poisson distribution, are used to evaluate the results of these experiments. The Audic and Claverie statistic is based on the assumptions that identifying any specific cDNA in a library is a rare event that represents one possible outcome of many, and that the total number of possible outcomes is unknown [7]. Confidence intervals, both 1% and 5%, corresponding to the likelihood of selecting a specific species of cDNA among a subset of all possible cDNAs, are generated. The probability of selecting a specific clone is independent of sample size because the statistic accounts for differences in population size between libraries. This statistic becomes more reliable as the size of the sample analyzed increases. However, it is applicable only to pairwise

comparisons and cannot be used to identify transcripts differentially expressed in more than two libraries. Stekel *et al.* [8] proposed using the R statistic to analyze the abundances of cDNAs in multiple libraries. The R statistic is a log likelihood ratio and similarly to the  $C^2$  distribution, its distribution is asymptotic. This log likelihood ratio is constructed from the likelihood of seeing an observed event, or in this case, the probability of selecting a specific species of cDNA from multiple libraries. It represents the differences in observed EST counts among multiple libraries as differences in gene expression levels rather than as random sampling variability.

Consequently we made pairwise comparisons in gene expression frequency using the Audic and Claverie Statistic and identified genes as differentially expressed among the three libraries using the R statistic. Contig sizes varied significantly (Figure 2). Almost half of the contigs contained only 2 ESTs, although the maximum number was 90. As expected, the number of contigs containing larger numbers of ESTs decreased exponentially in frequency as the number of contained ESTs increased.

Housekeeping genes are usually constitutively expressed in virtually all cell types of a multicellular organism even under a wide range of physiological and experimental circumstances [5]. Typically, though there are exceptions, they are also abundantly expressed. Thus, contigs represented in all three libraries and containing higher numbers of ESTs are most likely to represent transcripts encoding proteins involved in housekeeping functions. Contigs composed of more than 20 ESTs are described in Table 5. Indeed, 18/29 ASs, approximately 62% of contigs containing more than 20 ESTs, encode structural constituents of ribosomes. Unexpectedly, other genes commonly considered to housekeeping genes, such as those involved in oxidative phosphorylation, are not represented among these contigs. In fact, the most frequently identified gene products were AS 3, the *A. gambiae* agCP10095, and AS 1, a homolog of *D. melanogaster* LP07070, with 89 and 90 ESTs respectively. Surprisingly for abundant transcripts likely to encode housekeeping functions, there are currently no clues in the literature as to their functions. One potential clue is that transcription of AS 1 and AS 3 may be induced by *P. berghei* infection. It is also noteworthy that three of the contigs containing more than 20 ESTs, ASs 28, 99, and 18, represent genes encoding serine-type endopeptidases. These enzymes all share high sequence identity with previously studied serine proteases and are differentially expressed among the three libraries (see below).

25 of the 149 contigs containing more than five ESTs represent cDNAs differentially expressed among the three libraries (Table 6). We detected three main patterns of

expression among these genes. These include 1) up-regulation in the IRB library relative to any other library, 2) up-regulation in both of the RB and IRB libraries relative to S, and 3) down-regulation in the RB and IRB libraries and/or up-regulation in the S library. Gene products up-regulated at 30 h PI with *P. berghei* included AS 996, agCP14019, an apparent cathepsin B; AS 230, Vitellogenin but not AS 447, the other vitellogenin gene 3' end contig (see below); AS 24, a probable vitelline membrane protein; AS 139, ribosomal protein L44; AS 475, mucin; AS 145, a high molecular weight (HMW) kininogen; AS 313, cytochrome c oxidase; and AS 270, which is similar to GenBank accession #BM600177, an unknown gene product sequenced in the Celera Genomics *A. gambiae* EST project [1].

AS 996, a cathepsin B, sharing marginally-significant sequence identity with the *Ae. aegypti* vitellogenic cathepsin B-like protease, VCB, [75], also appeared to be up-regulated in the IRB library. In *Ae. aegypti*, VCB is secreted maximally from the fat body at 24 h PBM and accumulated by developing oocytes. During embryogenesis VCB degrades vitellin, the stored form of vitellogenin. qRT-PCR of AS 996 showed that it is induced following a blood meal, and increased further in infected blood-fed mosquitoes in comparison with naïve blood-fed mosquitoes (Figure 3A). Ribeiro [15] also found using a digital Northern approach that this gene is up-regulated in whole adult female mosquitoes 24 h PBM. It is possible that the observed increase in this cathepsin B may be responsible for part of the decrease in vitellogenin protein observed late in the first gonotrophic cycle and during the second gonotrophic cycle following infection with *Plasmodium* [18]. It is also possible that this cathepsin B may have an immunity function.

Female mosquitoes synthesize large quantities of vitellogenin in the first day following a blood meal. Consequently, it was expected that the RB and IRB libraries would show increases in vitellogenin (Vg) gene transcript abundance above the S library. Both AS 230 and AS 447 share sequence identity with Vg but represent non-overlapping sequences from the 5' and 3' ends of the genes, respectively. When these two contigs were analyzed together, Vg appeared to be expressed at significantly higher abundance in the RB library than in S, as expected, and not to be repressed and/or degraded within 30 h of *P. berghei* infection (Table 6). This result is supported by qRT-PCR analysis (Figure 3B). Since Ahmed *et al.* [18] showed that *P. yoelii nigeriensis* infection of *A. gambiae* results in fecundity reduction, due, in part, to reduction in vitellogenin mRNA accumulation, we might have expected to see a slight, though statistically insignificant reduction in Vg mRNA abundance in the IRB library by 30 hr PBM, even at the lower temperature required for devel-

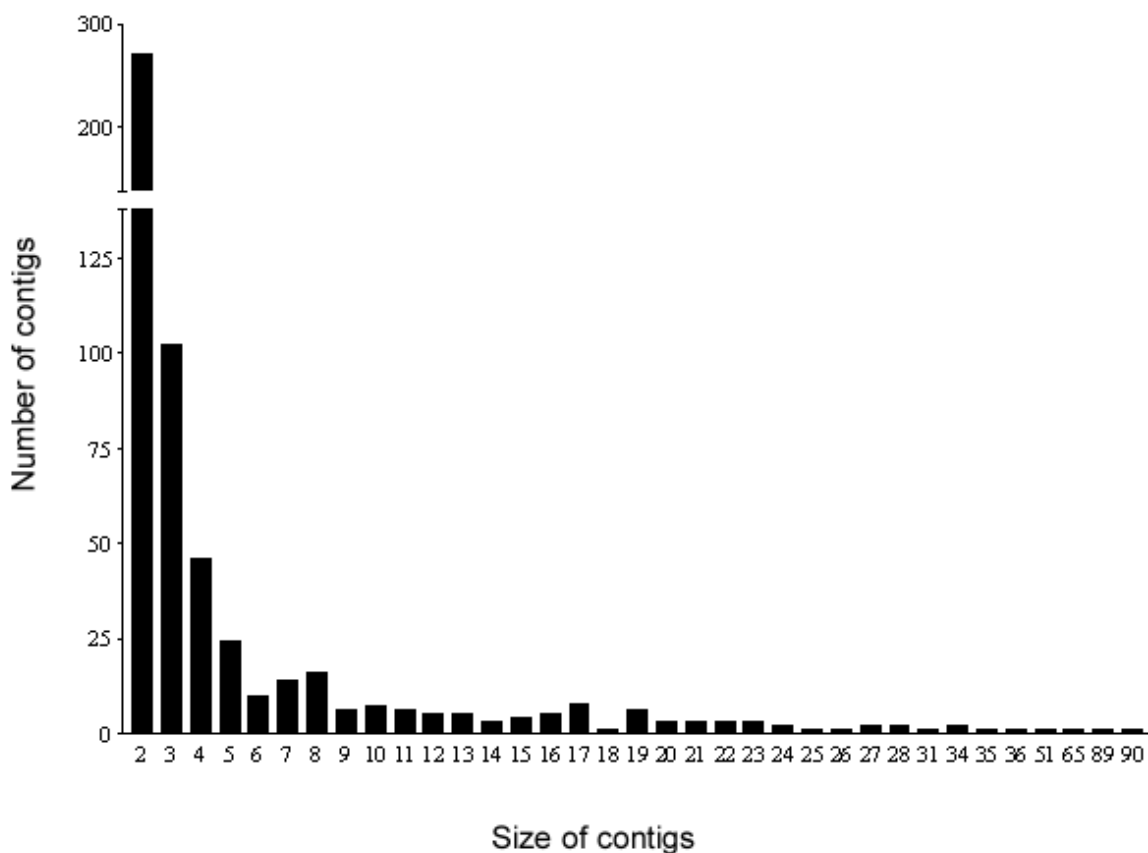
opment of *P. berghei* (19°C vs. the 24°C used by Ahmed *et al.*)<sup>2</sup> analysis showed that AS 230 is expressed in significantly higher abundance than AS 447 in both the RB and IRB libraries,  $0.05 > P > 0.025$  and  $P < 0.001$ , respectively. However a goodness of fit test showed that the under-representation of Vg mRNA 3' ends is exaggerated significantly in the IRB library,  $G = 26.3$ ,  $P < 0.01$  [76]. We can not currently explain this observation, except to suggest that Vg transcripts may be degraded from their 3' ends in response to *Plasmodium* infection.

S 24 also appears to be up-regulated in response to *Plasmodium* infection. BLASTN analysis showed that its nucleotide sequence is 96% identical to the *A. gambiae* transcript ENSANG00000021567, a gene product belonging to the vitelline membrane protein family. Insect orthologs of AS 24 include both the *Drosophila* Vm34Ca protein and the *Ae. aegypti* vitelline membrane protein 15A-1. In *Ae. aegypti*, 15A-1 mRNA is most abundant between 30 and 45 h PBM [77]. Our qRT-PCR analysis showed that this gene is induced following a blood meal (data not shown).

*P. berghei* invasion of *An. stephensi* midguts results in damage to invaded epithelial cells and their extrusion into the midgut lumen [41,78]. This may induce an inflammatory response. AS 475 is 100% identical at the amino acid level to the previously identified midgut-specific, membrane-bound mucin AgMuc1. Associated with the apical microvilli on the midgut, this mucin contains a putative GPI-anchor and two hydrophobic domains. This result suggested that there may be a link between this membrane mucin and signal transduction following damage to the epithelium [79]. Membrane mucins also act as physical barriers protecting the free surface of the cell [80]. Therefore, increased mucin gene expression may serve as a protective response to parasitic invasion. Ribeiro [15] found that this mucin is up-regulated in blood-fed mosquitoes 24 h PBM [15]. Our qRT-PCR analysis showed no difference in gene expression PBM or PI (Figure 3C).

AS 145, putatively identified as agCP6338, contains a HMW kininogen protein domain. This domain is a signature for a family of similar inflammatory response proteins in vertebrates. HMW kininogen is a component of the vertebrate kinin system, a pathway involved in inflammation and pain responses to cell damage [44]. Both digital Northern analysis and qRT-PCR (Figure 4D) identified AS 145 as being induced only after *Plasmodium* infection. Damage to the midgut epithelium caused by invasion and penetration of *P. berghei* ookinetes may have been responsible for inducing AS145.

In contrast to other gene products induced in mosquito abdomens by infection with *P. berghei*, AS 170 appears to



**Figure 2**

Transcript abundance distribution within the 567 multi-EST contigs of the S, RB and IRB libraries. The 1408 singletons are not shown. Contig size = the number of ESTs contained within it.

be repressed following a blood meal to levels below those found in sugar-fed females though it is well represented in the IRB library. This assembled sequence, which is derived completely from 5' UTR, not from protein-coding sequence, is highly identical to GenBank Accession # [BM653334](#), identified among their *A. gambiae* ESTs by Celera Genomics [1]. BM653334 was assembled into TC10892 in the TIGR gene indices, a tentative consensus sharing identity with the *Drosophila* 70-kDa heat shock protein, Hsp70. Hsp70s are molecular chaperones highly conserved in all organisms. While Hsp70s are induced in response to heat shock and other stresses, they also function in many normal cellular processes including protein translation, translocation, folding and quality control, as well as repression of cell growth and apoptosis through

specific protein-protein interactions [81]. Clearly, induction of this gene post IRB could be a stress response to *Plasmodium* infection. The interactions of mosquito Hsp70s with both mosquito and parasite proteins deserve further study.

Genes up-regulated following a blood meal could be divided into two groups, genes whose transcription is unaffected by *P. berghei* infection at 30 h PI, and genes whose transcription is apparently repressed. Gene products grouped into the first category include AS 98, hydrogen-transporting two-sector ATPase; AS 521, ribosomal protein L13; and AS 447, the second Vitellogenin 3' end contig. In contrast, gene products in the latter category include AS 35, glutathione S-transferase 1-6; AS 537, gly-

Table 5: MOLECULAR FUNCTIONS OF CONTIGS CONTAINING &gt; 20 ESTS.

| Contig     | S ESTs    | RB ESTs   | IRB ESTs  | Total ESTs | Blast Hit            | Molecular Function                 | Organism                      |
|------------|-----------|-----------|-----------|------------|----------------------|------------------------------------|-------------------------------|
| 7          | 3         | 8         | 9         | 20         | agCPI1398            | structural constituent of ribosome | <i>A. gambiae</i>             |
| 14         | 11        | 6         | 3         | 20         | agCPI538             | structural constituent of ribosome | <i>A. gambiae</i>             |
| <b>28</b>  | <b>17</b> | <b>3</b>  | <b>0</b>  | <b>20</b>  | <b>agCPI1956</b>     | <b>serine-type peptidase</b>       | <b><i>A. gambiae</i></b>      |
| 140        | 5         | 6         | 10        | 21         | agCP7468             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 167        | 11        | 4         | 6         | 21         | RpS27A               | structural constituent of ribosome | <i>D. melanogaster</i>        |
| <b>516</b> | <b>8</b>  | <b>10</b> | <b>3</b>  | <b>21</b>  | <b>agCP3409</b>      | <b>peritrophin</b>                 | <b><i>A. gambiae</i></b>      |
| <b>5</b>   | <b>9</b>  | <b>4</b>  | <b>9</b>  | <b>22</b>  | <b>agCPI2023</b>     | <b>unknown</b>                     | <b><i>A. gambiae</i></b>      |
| 20         | 13        | 5         | 4         | 22         | agCPI0687            | structural constituent of ribosome | <i>A. gambiae</i>             |
| <b>201</b> | <b>7</b>  | <b>6</b>  | <b>9</b>  | <b>22</b>  | <b>agCP7935</b>      | <b>unknown</b>                     | <b><i>A. gambiae</i></b>      |
| 8          | 15        | 5         | 3         | 23         | agCP8133             | structural constituent of ribosome | <i>A. gambiae</i>             |
| <b>13</b>  | <b>16</b> | <b>2</b>  | <b>7</b>  | <b>23</b>  | <b>peritrophin I</b> | <b>peritrophin</b>                 | <b><i>A. gambiae</i></b>      |
| 532        | 7         | 6         | 10        | 23         | agCPI729             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 165        | 10        | 7         | 7         | 24         | agCP7766             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 199        | 9         | 5         | 10        | 24         | agCP4384             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 17         | 12        | 3         | 10        | 25         | agCP8340             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 528        | 6         | 7         | 13        | 26         | agCP4228             | structural constituent of ribosome | <i>A. gambiae</i>             |
| <b>99</b>  | <b>16</b> | <b>3</b>  | <b>8</b>  | <b>27</b>  | <b>agCP3123</b>      | <b>serine-type endopeptidase</b>   | <b><i>A. gambiae</i></b>      |
| <b>230</b> | <b>1</b>  | <b>7</b>  | <b>19</b> | <b>27</b>  | <b>agCP2518</b>      | <b>vitellogenin</b>                | <b><i>A. gambiae</i></b>      |
| 139        | 6         | 5         | 17        | 28         | ebiP415              | structural constituent of ribosome | <i>An. gambiae</i>            |
| 229        | 8         | 9         | 11        | 28         | agCP8207             | structural constituent of ribosome | <i>An. gambiae</i>            |
| <b>18</b>  | <b>25</b> | <b>2</b>  | <b>4</b>  | <b>31</b>  | <b>agCPI1264</b>     | <b>serine-type endopeptidase</b>   | <b><i>A. gambiae</i></b>      |
| <b>15</b>  | <b>19</b> | <b>5</b>  | <b>10</b> | <b>34</b>  | <b>Eflalpha48D</b>   | <b>elongation factor</b>           | <b><i>D. melanogaster</i></b> |
| 269        | 10        | 6         | 18        | 34         | agCP9994             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 4          | 17        | 5         | 13        | 35         | agCP9893             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 521        | 7         | 14        | 15        | 36         | agCP8317             | structural constituent of ribosome | <i>A. gambiae</i>             |
| 91         | 13        | 12        | 26        | 51         | agCPI1873            | structural constituent of ribosome | <i>A. gambiae</i>             |
| 228        | 19        | 16        | 30        | 65         | agCP9509             | structural constituent of ribosome | <i>A. gambiae</i>             |
| <b>3</b>   | <b>31</b> | <b>21</b> | <b>37</b> | <b>89</b>  | <b>agCPI0095</b>     | <b>unknown</b>                     | <b><i>A. gambiae</i></b>      |
| <b>1</b>   | <b>27</b> | <b>19</b> | <b>44</b> | <b>90</b>  | <b>LP07070</b>       | <b>unknown</b>                     | <b><i>D. melanogaster</i></b> |

cine hydroxymethyltransferase; AS 273, ATP synthase b; and AS 516, Peritrophin 1.

AS 516 is 98% identical at the amino acid level to Peritrophin 1, a midgut peritrophic matrix (PM) protein, and maps *in silico* to the same location on chromosome 2L as Peritrophin 1. Peritrophin 1 mRNA is present in sugar-fed females at 5 days post-eclosion, but is induced 12–24 h PBM [82]. Our digital Northern analysis showed that AS 516 is induced by at 30 h PBM, but repressed in response to *Plasmodium* infection. Ookinetes may partially inhibit PM formation, facilitating their penetration through it to gain access to the midgut epithelium. This is unsurprising, given that ookinetes produce locally-acting chitinases that are required for PM penetration and midgut invasion [83–85], and that they are targets for transmission blocking vaccines [86].

Almost half of the differentially expressed genes are down-regulated PBM. AS 113 and AS 408, two unknown gene products; AS 13, a peritrophin-like protein; AS 18, Trypsin 1; and AS 99, Chymotrypsin 2, are significantly less abundantly expressed in the mosquito abdomen at 30 h PBM than in S females and not affected further by ookinete invasion. qRT-PCR (Figure 3E, see below) verifies this expression pattern for Chymotrypsin 2. Genes repressed PBM may also include a AS 591, a gene product with no known function but which is part of the mucin 4, tracheobronchial mucin fragment protein family, and AS 208, an ATP dependent RNA helicase, although the differences in abundance of these transcripts between S and RB are not statistically significant.

AS13 shares greatest amino acid identity with Peritrophin 1. However, it shares greater nucleotide identity with ENSANGG00000020776, a gene located 4 kb downstream of Peritrophin 1, and is likely to have been derived from the latter. AS 18 and AS 99 share identity with two enzymes involved in digestion of the blood meal. AS 18 is 100% identical at the amino acid level to Trypsin 1. Northern blots showed that transcripts of Trypsin 1, P35035, are present in adult female mosquitoes by 4 h PBM, increase rapidly until 12 h, peak at 16 h, remain at this plateau until 24 h, then decrease steadily until 40 h, and drop to baseline levels by 48 h PBM [87,88]. Trypsin 1 is also expressed in S females until 5 days post eclosion, but not at the high levels exhibited by Trypsin 4 [87]. Therefore it is not entirely clear why we found more transcripts of Trypsin 1 than of Trypsin 4 in the S library, and more in the S than in the 30 h PBM library. AS 99, another trypsin-like SP, is 97% identical to agCP3123, previously identified as the Chymotrypsin 2 precursor, Anchym2 [89]. This gene is expressed in the midgut at 12 h PBM and remains abundant until 48 h. Anchym2 transcripts are undetectable in S females. Using qRT-PCR, we found that

AS 99 transcript abundances are reduced in both RB and IRB mosquitoes relative to S females (Figure 3E), though Ribeiro [15], also using transcript frequency analysis, found that both Trypsin 1 and Chymotrypsin 2 are up-regulated in whole adult female *A. gambiae* PBM. Since trypsin activity is known to be age-dependent in *A. gambiae* females [90], it is possible that these serine protease genes may be transcribed at higher levels in younger adult females than after the peak of proteolytic activity following a blood meal.

AS 28, agCP11956, SP G13, appears to be down-regulated PBM and further repressed following ookinete invasion. However it was shown not to be blood meal responsive in a previous study [34].

Housekeeping gene products involved in normal cellular maintenance may also be differentially expressed following septic injury. AS 28 and AS 170, the hsp70 homolog discussed above, are potential examples of two such proteins. Two assembled sequences corresponding to housekeeping genes up-regulated following *P. berghei* infection, AS 313 and AS 139, share sequence identity with a cytochrome c oxidase subunit and ribosomal protein L44, respectively. In contrast, the only housekeeping gene product down-regulated following *P. berghei* infection is AS 8 which shares amino acid identity with ribosomal protein L38e. The only gene identified as significantly more abundantly expressed in the mosquito abdomen PBM but unaffected by the presence of *Plasmodium* infection was AS 98, a hydrogen-transporting two-sector ATPase. Increased abundances of oxidative phosphorylation proteins may be linked to apoptosis or to increased metabolic demands placed on invertebrate hosts by invaders. The roles of these housekeeping gene products in immunity remain to be established.

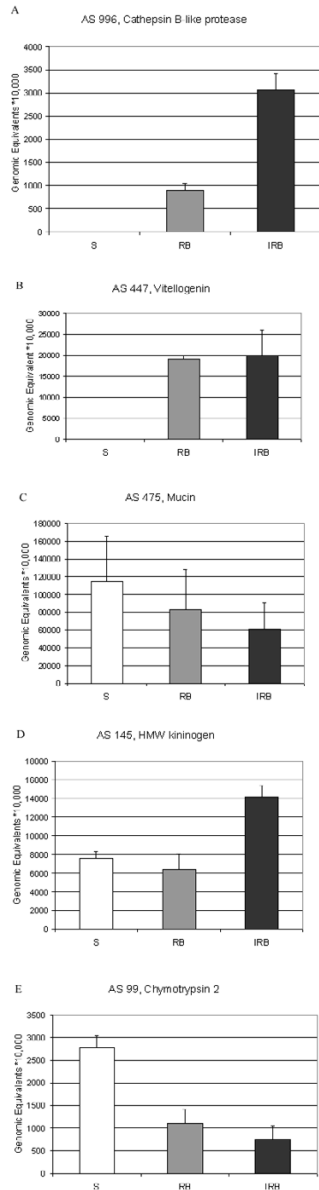
Several *A. gambiae* genes identified in this study as being differentially regulated (Table 6) have potentially also been studied in microarray experiments performed in *A. stephensi* under generally similar experimental conditions [17,37]. Abraham *et al.* studied *A. stephensi* genes primarily from a cDNA library enriched for genes expressed in mosquito midguts containing early oocysts of *P. berghei* [17] and identified 226 EST contigs likely to have been of mosquito origin. None of these had clear identity with any of the putative differentially expressed *A. gambiae* genes in Table 6. Xu *et al.* studied *A. stephensi* genes differentially regulated between 6 h and 20 d PI with *P. berghei* [37]. Though the strategy used by Xu *et al.* in making the subtraction cDNA libraries assayed in the microarray was expected to have biased their identified mosquito genes towards ones upregulated PI and did bias mosquito genes towards ones upregulated between 20 h and 20 d PI, a chymotrypsin 2 precursor was identified as being



**Table 6: ASSEMBLED SEQUENCES WITH STATISTICALLY SIGNIFICANT DIGITAL NORTHERN VALUES**

| AS  | Sequence similarity; Description   | S      | RB | IRB | S (norm) | RB (norm) | IRB (norm) | AC S-RB      | AC S-IRB     | AC RB-IRB    | R            |
|-----|--|--------|----|-----|----------|-----------|------------|--------------|--------------|--------------|--------------|
| 996 | agCP14019; similar to <i>Aedes aegypti</i> vitellogenic cathepsin-B like protease (VCB).   | 0      | 0  | 5   | 0        | 0         | 29         |              | <b>0.015</b> | <b>0.03</b>  | <b>0.007</b> |
| 24  | no known predicted protein; shares 96% nucleotide identity with transcript ENSANG00000021567, a gene product belonging to the vitelline membrane protein family                                    | 0      | 1  | 12  | 0        | 9         | 72         | 0.240        | <b>0</b>     | <b>0.004</b> | <b>0</b>     |
| 139 | ebiP415; Ribosomal protein L44   | 6      | 5  | 17  | 35       | 44        | 102        | 0.132        | <b>0.005</b> | <b>0.017</b> | <b>0.034</b> |
| 475 | agCP12050; AgMucI  | 2      | 3  | 12  | 12       | 26        | 72         | 0.138        | <b>0.002</b> | <b>0.024</b> | <b>0.012</b> |
| 145 | agCP6338; protein of unknown function containing a HMW kininogen protein domain  | 1      | 2  | 8   | 6        | 18        | 48         | 0.172        | <b>0.008</b> | <b>0.046</b> | <b>0.034</b> |
| 313 | CGI4235; Cytochrome c oxidase  | 1      | 2  | 8   | 6        | 18        | 48         | 0.172        | <b>0.008</b> | <b>0.046</b> | <b>0.034</b> |
| 270 | no known predicted protein; shares sequence identity with transcript I7000687051196 A.Gam.ad.cDNA.blood I of unknown function  | 2      | 3  | 10  | 12       | 26        | 60         | 0.138        | <b>0.007</b> | <b>0.042</b> | <b>0.042</b> |
| 230 | agCP2518; Vitellogenin I   | 1      | 7  | 19  | 6        | 61        | 114        | <b>0.005</b> | <b>0</b>     | <b>0.024</b> | <b>0</b>     |
| 98  | agCP4445; hydrogen-transporting two-sector ATPase  | 0      | 3  | 4   | 0        | 26        | 23         | <b>0.038</b> | <b>0.030</b> | 0.119        | <b>0.035</b> |
| 521 | agCP8317; Ribosomal protein L13  | 7      | 14 | 15  | 41       | 122       | 90         | <b>0.005</b> | <b>0.018</b> | <b>0.043</b> | <b>0.039</b> |
| 447 | agCP2518; Vitellogenin I   | 0      | 10 | 8   | 0        | 87        | 48         | <b>0</b>     | <b>0.002</b> | 0.034        | <b>0</b>     |
| 35  | Glutathione S-transferase D1-6   | 0      | 4  | 3   | 0        | 35        | 17         | <b>0.015</b> | 0.060        | 0.081        | <b>0.024</b> |
| 537 | agCP5224; Glycine hydroxymethyltransferase   | 0      | 5  | 2   | 0        | 44        | 11         | <b>0.006</b> | 0.123        | <b>0.034</b> | <b>0.009</b> |
| 273 | agCP12503; ATP synthase B chain mitochondrial precursor (FO-ATP synthase subunit B)  | 1      | 6  | 3   | 6        | 52        | 17         | <b>0.010</b> | 0.123        | <b>0.032</b> | <b>0.038</b> |
| 516 | agCP3409; AgPer I  | 8      | 10 | 3   | 46       | 87        | 17         | <b>0.046</b> | <b>0.045</b> | <b>0.003</b> | <b>0.029</b> |
| 8   | agCP8133; Ribosomal protein L38e   | 1<br>4 | 6  | 3   | 81       | 52        | 17         | 0.076        | <b>0.003</b> | <b>0.032</b> | <b>0.025</b> |
| 208 | agCP7801; ATP dependent RNA helicase   | 5      | 1  | 0   | 28       | 9         | 0          | 0.113        | <b>0.017</b> | <b>0.166</b> | <b>0.030</b> |
| 28  | agCP11956; Serine protease G13   | 1<br>7 | 3  | 0   | 98       | 26        | 0          | <b>0.008</b> | <b>0</b>     | <b>0.027</b> | <b>0</b>     |
| 591 | agCP1095; peptide with no known function, is part of the protein Family mucin 4 tracheobronchial mucin fragment, shares weak identity to <i>Cryptosporidium parvum</i> mucin-like glycoprotein 900 | 6      | 1  | 0   | 35       | 9         | 0          | 0.079        | <b>0.009</b> | 0.17         | <b>0.013</b> |
| 113 | agCP10139; 65 aa peptide of unknown function containing no known protein domains except signal peptide and transmembrane regions.  | 8      | 0  | 0   | 46       | 0         | 0          | <b>0.010</b> | <b>0.002</b> |              | <b>0</b>     |
| 18  | agCP11264; Trypsin I   | 2<br>5 | 2  | 4   | 145      | 18        | 23         | <b>0</b>     | <b>0</b>     | 0.125        | <b>0</b>     |
| 408 | I7000687367332 A.Gam.ad.cDNA1; EST of unknown function   | 5      | 0  | 0   | 28       | 0         | 0          | <b>0.047</b> | <b>0.017</b> |              | <b>0.008</b> |
| 13  | gene ENSANGG0000002077 6; gene located 4 kb 3' to the Peritrophin I that encodes a protein with a chitin binding domain  | 1<br>6 | 2  | 5   | 93       | 18        | 29         | <b>0.004</b> | <b>0.006</b> | 0.104        | <b>0.007</b> |
| 99  | agCP3123; Chymotrypsin 2   | 1<br>6 | 3  | 8   | 93       | 26        | 48         | <b>0.01</b>  | <b>0.025</b> | 0.069        | <b>0.054</b> |
| 170 | agCP12309; putative homolog of <i>Drosophila melanogaster</i> Heat shock 70 kDa protein cognate 4 (Heat shock 70 kDa protein 88E)  | 7      | 0  | 4   | 41       | 0         | 23         | <b>0.017</b> | 0.086        | <b>0.050</b> | <b>0.028</b> |

Actual EST counts (S, RB and IRB) and the normalized number of ESTs (S (norm), RB (norm), and IRB (norm)) are given for each cDNA library. Assembled sequences were identified as differentially expressed among the three libraries if the R statistic was less than 0.05 (Stekel *et al.* 2000). The Audic and Claverie statistic (AC < 0.05) was used to identify libraries in which the assembled sequence was statistically significantly differentially expressed (Audic and Claverie, 1997). Pairwise comparisons between libraries are indicated by AC S-RB, AC S-IRB, and AC RB-IRB, respectively. All statistics were calculated with the IDEG6 Tool (Romualdi *et al.*, 2003). Cells with characters in bold (underlined), bold (italics) and bold indicate statistical values equal to zero, less than 0.01, and less than 0.05, respectively. Cells with characters in *italics* (only), represent different levels of gene expression ranging from no detectable transcripts (*smallest*) to 145 transcripts (*largest*) detected in a library.



**Figure 3**  
 qRT-PCR of selected genes identified as differentially expressed in *A. gambiae* abdomens by digital Northern analysis. A. AS 996, a cathepsin B-like protease; B. AS 447, Vitellogenin; C. AS 475, Mucin; D. AS 145, a putative high molecular weight kininogen; and E. AS 99, Chymotrypsin 2. Transcript abundances were normalized against the abundance of RP S7 under the same condition and shown as genomic equivalents \* 10,000 for S, RB, and *P. berghei*-infected (IRB) adult female mosquitoes 30 h post treatment. The arithmetic means of three biological replicates SEM are shown. ANOVA was used to compare the means, and a *post hoc* Tukey test was used to make pairwise comparisons. Statistically significant differences,  $P < 0.05$  are discussed in the text.

expressed more abundantly at earlier times than at later ones. Since their study did not include a 0 h PI time point, we cannot be certain that there is an actual discrepancy between our data and theirs. They found a peak in NOS activity at 40 h PI though we have no NOS ESTs with which to compare their data. Since Xu *et al.* did not test RNA from uninfected mosquitoes and tested only one time point, 20 h, at all close to our 30 h we can not thoroughly evaluate the single identified gene overlap between the two data sets. In addition to the facts that experimental conditions varied among all three data sets and that at least one of these data sets was small, it is possible that *A. gambiae* and *A. stephensi* have only partially overlapping repertoires of transcriptionally regulated responses to infection with *P. berghei*, simply because they are not extremely closely related anophelines. Though they are both classified within the subgenus *Cellia*, they are grouped differently within the subgenus, *A. gambiae* in the *Pyrethophorus* series and *A. stephensi* among the *Neocellia*.

**Conclusion**

We have described 1975 genes expressed in the abdomens of adult female *A. gambiae* mosquitoes, 13% of which are not predicted by the genome sequence or by identity with known genes in other organisms. The latter provide important information for further *A. gambiae* genome annotation. All together these genes form a resource likely to be very useful for annotation of genes in other organisms. While *D. melanogaster* provides a model for many kinds of biological and bioinformatic analyses, its genome has been highly streamlined, more so than that of *A. gambiae*. Consequently, the *A. gambiae* genome and its associated genomic resources including its EST collections may sometimes prove more useful than *Drosophila* for annotation of genes in taxonomically distant organisms.

We have identified several genes as being induced following blood-feeding and/or *P. berghei* infection using the digital Northern technique. Limitations on this approach include the numbers of ESTs obtained from each library, the limited number of time points or physiological states that can be examined, and the requirement that their mRNAs be of intermediate or high abundance in at least one of the conditions studied. Thus some, indeed, many blood-feeding and immune-responsive genes may not have been identified, simply because they are expressed at low levels despite the importance of their roles in these processes. In addition, transcript abundances do not always correlate highly with protein levels. Some mRNAs have high turnover rates while others may be stabilized yet not translated except under specific conditions. Multiple, as well as more targeted approaches may be required before all gene products involved in responses to infection with *Plasmodium* are identified. Nevertheless, the ESTs

obtained from the three different cDNA libraries have provided at least a rudimentary catalog of genes expressed in the abdomens of adult female *A. gambiae* harvested 30 h after they had fed on sugar, on blood or on blood infected with *P. berghei*. We should acknowledge some caution in the interpretation of cases where gene expression patterns in this study differ from those of comparison studies. This study was based on whole mosquito abdomens, which contain a multitude of organs, tissues and cell types and some comparison studies may involve either whole mosquitoes or subsets of the organs and tissues found in abdomens. Despite the limitations of the digital Northern portion of our study, we have identified several gene products as candidates for involvement in *Plasmodium*-immunity processes. These should be studied further. In addition, the sequences in the three cDNA libraries will certainly inform more detailed microarray and qRT-PCR studies of *A. gambiae* gene expression in both physiologically normal and *Plasmodium*-infected females.

## Methods

### Mosquitoes and *P. berghei* infections

*Plasmodium* infection susceptible (4arr) and transforming ookinete-encapsulating (L-35) strains of *A. gambiae* were reared under standard conditions 25°C, 70% humidity, 12 h light/dark. Adults were maintained on 20% sucrose until they were five to seven days post-eclosion. They were then blood fed on naïve and *P. berghei*-infected rats and maintained under similar conditions except that the temperature was lowered to 19–20°C to be permissive for development of *P. berghei* [91].

Female white rats (*Rattus norvegicus*) were maintained in the Freimann Life Sciences Center according to protocols established by the University of Notre Dame IUCAC. They were infected by intraperitoneal injection with  $4 \times 10^7$  *P. berghei* gametocytes of the ANKA 2.34 clone (gift from Marcelo Jacobs-Lorena) suspended in 10% DMSO. Rat parasitemia levels were determined from Giemsa-stained thin blood smears prepared from eye blood. Unfixed fresh eye blood samples were examined for gametocyte exflagellation, and exflagellation levels calculated as the number of events per 20 random microscopic fields at 400X magnification.

Female mosquitoes were blood fed on *Plasmodium*-infected anaesthetized rats. To determine infection intensity and the time by which the majority of *P. berghei* ookinetes had penetrated to the basal lamina of the mosquito midgut epithelium and were transforming to oocysts, midguts of infected L-35 strain females were dissected in 1% formaldehyde in PBS at 30, 36 and 42 hours PBM, then transected longitudinally to remove the food bolus, washed in PBS and mounted flat on microscope slides [92,93]. Slides were examined microscopically and encap-

sulated parasites counted using bright field illumination and 400 × magnification. A two-way ANOVA was used to compare the mean numbers of parasites per midgut at the three times and a post hoc Tukey test was employed to detect differences in the means [76,94].

### cDNA library construction

5–7 day old female mosquitoes of the 4arr strain were 1) sugar fed (S library), 2) blood fed (RB library), and 3) *P. berghei*-infected (IRB library). Mosquitoes were immediately transferred to 20°C and incubated for 30 hours. Parasitemia of the rat used for these infections was 11.1%. Mean mosquito infection prevalence was 83%. The infection intensity ranged from 2 to 23 with an average of 9 parasites per midgut.

Blood and sugar-fed females were flash frozen in liquid nitrogen at 30 h PBM, or the equivalent age in the case of the S females, then vortexed at -20°C to sever abdomens from heads and thoraces, similarly to [86]. Abdomens were collected at -20°C and total RNA extracted using TRIzol (MRC, Inc.) according to the manufacturer's instructions. Poly-A<sup>+</sup> mRNA was isolated using the PolyA Tract mRNA Isolation System (Promega). S, RB and IRB cDNA libraries were constructed using the SMART<sup>™</sup> cDNA Library Construction Kit (Clontech) from 1.54, 3.82 and 2.1 µg of poly-A<sup>+</sup> mRNA, respectively. Unless otherwise stated, all reagents used were those provided in the kit. Reverse transcription of mRNA was for 1 hr at 42°C using Superscript II Reverse Transcriptase (Invitrogen) and a modified oligo-dT primer, CDSIII/3' PCR Primer: 5' - ATTCTAGAGGCCGAGGCGGCCGACATG-d(T)<sub>30</sub>-3' (Invitrogen), to prime the first strand synthesis reaction. This primer contains the Sfi IB restriction site used for directional cloning. An additional oligonucleotide, either the SMARTIII<sup>™</sup> Oligonucleotide, 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCGGG-3' (Slibrary), or the SMART IV<sup>™</sup> Oligonucleotide, 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' (RB and IRB libraries), contains the Sfi IA restriction site followed by 3 guanines. Second strand synthesis of cDNAs was conducted in a 100 l volume with 11.0 l first strand cDNA, 0.2 M of the indicated 5' and 3' oligonucleotides, 1X Advantage 2 PCR Buffer, 1X dNTP Mix, and 1X Advantage 2 Polymerase Mix. These reactions were performed on a Perkin-Elmer 9600 Thermocycler using the following cycling conditions: 72°C for 10 min, 95°C for 20 s, followed by 3 cycles of 95°C for 5 s and 68°C for 8 min. Primer extended cDNAs were visualized on an ethidium bromide stained 1.25% agarose, 1X TBE gel for quality assessment. They appeared as homogeneous smears ranging from 100 bp to 5 kb. mRNA aliquots not used for cDNA synthesis and cloning were also subjected to PCR amplification using Taq Polymerase and visualized on the same gel. The absence of any visible product on the gel confirmed that

genomic DNA did not contaminate these mRNA populations.

Following proteinase K digestion and phenol:chloroform extraction, the amplified cDNAs were digested with 10  $\mu$ l Sfi I (20 U/ $\mu$ l) at 50°C for 2 h and size fractionated using CHROMA SPIN-400 columns (Clontech). The first three to four fractions containing cDNAs longer than 500 bp were pooled, ethanol precipitated, and concentrated in 4.0 l nuclease free water (Gibco, UltraPure). These cDNAs were directionally cloned into Sfi I digested TripIEx2 (Clontech), and packaged using Gigapack III Gold Packaging Extract (Stratagene) according to the protocol provided. Packaged recombinant phages were incubated with log phase *E. coli* XL1-Blue cells (Stratagene), plated and library titers determined. All three libraries were plated at 100 and 1000 pfu/plate.

White plaques were isolated and recombinant phages eluted overnight in 100 l SM buffer (0.1 M NaCl, 0.01 M MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.05 M Tris-HCl (pH 7.5), 0.01% (w/v) gelatin). The inserts were amplified via PCR using 5' and 3' vector specific primers; 5' LD Amplimer Primer, 5'-CTCGGGAAGCGGCCATTGTGTTGG-3' and 3' LD Amplimer Primer, 5'-ATACGACTCACTATAG-GCGAATTGGC-3' (Invitrogen). Amplification reactions contained 0.4 l eluted phage, 0.03 pmol of each primer, 1X Taq Polymerase Buffer (Invitrogen), 3 mM MgCl<sub>2</sub>, 1 mM of each dNTP, and 0.2 U Taq Polymerase (Invitrogen) in a total volume of 25 ml. Reactions were performed in 96-well plates on a Perkin-Elmer 9700 Thermocycler using the following cycling conditions; initial denaturation at 95°C for 5 min, followed by 25 cycles of denaturation at 94°C for 30 s and annealing/elongation at 70°C for 2 min, and a final elongation step at 68°C for 3 min. Seven samples were chosen randomly from each 96-well plate amplified, and 5 ml of the reaction was electrophoresed on an ethidium bromide stained 1% agarose, 1X TBE gel to confirm that the PCR was not contaminated and that no primer dimers could be visualized (data not shown). 704 PCR-amplified cDNA clone inserts were visualized on ethidium bromide stained 1% agarose, 1X TBE gels, and their insert sizes determined using the KODAK Digital Science 1D software (Scientific Imaging Systems; data not shown).

#### **cDNA clone sequencing and EST assembly**

cDNA clones were picked at random from the S, RB and IRB abdomen libraries and their 5' end sequences obtained through single-pass sequencing of the PCR-amplified inserts using the ABI PRISM Big Dye Terminators 3.0 Cycle Sequencing kit (ABI). All sequencing reactions were performed in 384-well plates on a Perkin-Elmer 9700 Thermocycler using the following cycling conditions; initial denaturation at 94°C for 4 min, fol-

lowed by 25 cycles of denaturation at 94°C for 10 s, annealing at 50°C for 5 s, and elongation at 60°C for 4 min. Each reaction contained 0.7 ml PCR product, 7.4 pmol of the 5' LD Amplimer Primer, 1× Sequencing Buffer (400 mM Tris pH 9.0, 10 mM MgCl<sub>2</sub>), and Big Dye (ABI) in a total volume of 7 ml. Reaction products were ethanol precipitated, resuspended in 20.0 ml HiDi formamide (ABI) and electrophoresed on an ABI 3700 Sequencer.

Sequences were trimmed of low quality and vector sequence, then screened to remove mitochondrial sequences using the SeqMan II software (DNASTAR, Inc.), prior to contig assembly. The options employed for SeqMan II assembly were match size = 12 bases, minimum match % = 80, minimum sequence length = 100, maximum added gaps per kb in contig = 70, maximum added gaps per kb in sequence = 70, maximum register shift difference (maximum base pair separation) between matches = 70, gap penalty = 0, and gap length penalty = 0.7. Consensus sequences derived from the alignment of multiple ESTs were defined as contigs, whereas ESTs that did not assemble into a cluster were defined as singletons. Consensus sequences were called by trace evidence, the majority percentage = 75, using the quality weights option.

#### **Bioinformatic analysis**

Bioinformatic analysis was initiated by subjecting consensus contig and singleton sequences to several blast searches. Initially, sequences were tested against the *A. gambiae* genome using BLASTN 2.2.4 [95,96]. The significance cutoff was chosen as  $E < 1 \times 10^{-4}$ . Sequences were then tested against the non-redundant nucleotide database in GenBank using BLASTX 2.2.4, at the same URL. Sequences that failed to yield significant BLASTX matches were retested against the same database using BLASTN. Finally, sequences lacking any significant BLAST hits were tested against dbEST using BLASTN.

Gene product identities were inferred from BLAST hits and the annotations provided for *A. gambiae* and *D. melanogaster* clones in public databases including The Institute for Genomic Research (TIGR) Gene Indices [97] and GadFly Genome Annotation Database in FlyBase [98-100]. Putative molecular functions of the gene products were determined using KEGG [101] and assigned to categories established by the Gene Ontology Consortium, GOC [102,103]. Gene products were also assigned to hybrid biological process categories by combining the categories used by the GOC and by [1].

#### **Digital northern**

Each contig represents an expressed gene and the number of sequences within a contig represents its transcript abundance. As in [104], only contigs containing more

than five ESTs were used for transcript profiling. Gene expression profiles were created by tabulating the frequencies of cDNAs corresponding to a particular gene in each library and then compared among the three experimental groups. Genes were identified as differentially expressed using the R Statistic. Differences between libraries were determined using the Audic and Claverie pairwise comparison statistic calculated using IDEG6 [104].

#### qRT-PCR

Quantitative real-time PCR (qRT-PCR) analysis was performed using SYBR Green I (Applied Biosystems) technology in order to validate data obtained from the digital Northern. The Primer Express v. 1.5 software (Applied Biosystems) was used to design primer sets for the following 8 transcripts: AS 24 (Forward 5'-GAAGTAGCGAGAGACAGCATCGA-3', Reverse 5'-TACGCTTCGGAGGTCACTTACTG-3'); AS 99 (Forward 5'-TTGCTGTCTCGGTACTCCTAG-3', Reverse 5'-GGTTGACGTAGTTGTCGTCCA-3'); AS 113 (Forward 5'-TGT-TAGTCCGCTGATGCTG-3', Reverse 5'-TCAATGTTATGGGTACACCTTGTGT-3'); AS 145 (Forward 5'-TGGCGATCTTTGTCATCGTG-3', Reverse 5'-GATGACCGTGTGACCACCAT-3'); AS 447 (Forward 5'-TCCACTGCCGTGACGCT-3', Reverse 5'-TCCCTTGCGGATCTGCTG-3'); AS 475 (Forward 5'-TGCCCCACAGGATGTGAAA-3', Reverse 5'-ATCGACATTGCCACGTATGC-3'); AS 996 (Forward 5'-GTGGGGCGATTCCAATGA-3', Reverse 5'-TGTAACCGGGCTGGCAAA-3'); and the ribosomal protein S7 gene, RP S7, (Forward 5'-CATTCTGCCCAAACCGATG-3', Reverse 5'-AACGCGGTCTCTTCTGCTTG-3'). RP S7 [106] is the internal control currently most widely used in studies of *A. gambiae* gene expression [34,35,46,105]. All amplifications and fluorescence quantifications were performed using an ABI 7700 Sequence Detection System and associated Sequence Detector Software v. 1.7 (Applied Biosystems). Standard curves were generated using 10-fold serial dilutions of *A. gambiae* strain 4Arr genomic DNA extracted using the Qiagen DNeasy Tissue Kit protocol for animal tissues (Qiagen, Inc.). The DNA dilutions ranged from 115 to 0.0115 ng per reaction. Total RNA was extracted from rat blood-fed and *P. berghei*-infected whole adult female mosquitoes using TRIzol (MRC, Inc.) RNA samples were incubated with 1.0 µl of DNase I in the supplied DNase I buffer (Invitrogen) for 15 min at room temperature to remove contaminating DNA. The DNase I was then inactivated by addition of 4.0 µl of 25 mM EDTA, pH 8.0 and the total RNA reisolated from TRIzol. RNA sample quality was evaluated by electrophoresis on 1% agarose, 1 × TAE gels, and ethidium bromide staining of the gels. 5 g samples of RNA were reverse transcribed at 42°C for 1.5 h using Superscript II Reverse Transcriptase and oligo(dT) primer (both from Invitrogen) to prime first strand synthesis. qRT-PCR reactions

were performed in duplicate in total volumes of 25 l containing 12.5 l of SYBR Green I PCR Master Mix, 300 nmol of each gene-specific primer, 50 ng of first strand cDNA template, and nuclease free water (Gibco, UltraPURE). All qRT-PCR reactions were performed using the following conditions: 50°C for 2 min, then 95°C for 10 min followed by 45 cycles of denaturation at 95°C for 15 s, annealing and extension at 60°C for 1 min. Amplification plots were generated as fluorescence of SYBR Green I (Rn) vs. PCR cycle number using the Sequence Detector Software v. 1.7 (Applied Biosystems). The abundance of each gene product in an RNA sample was estimated from its standard curve and normalized against the RP S7 transcript abundance in the same RNA sample. Expression levels were represented as genomic equivalents × 10,000. All comparisons were replicated on at least three biological samples, and the means and SEMs reported.

#### Abbreviations

AS, assembled sequence; bp, base pairs; EST, expressed sequence tag; GST, glutathione-S-transferase; h, hour; IRB, [*Plasmodium*-] infected red blood; kb, kilobases; min, minutes; nt, nucleotides; PBM, post blood meal; PI, post-infection; RB, rat blood; S, sugar, 20% sucrose; s, seconds; SEM, standard error of the mean; SP, serine protease; Vg, vitellogenin.

#### Authors' contributions

AND carried out the study with contributions from PAR, MEH, NFL and MKK. PAR and AND drafted the manuscript with contributions from NFL and FHC. All authors read and approved the final manuscript.

#### Additional material

##### Additional File 1

Gene identification and ontology of EST contigs identified in this study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-119-S1.xls>]

#### Acknowledgements

We wish to thank Dr. Marcelo Jacobs-Lorena for his kind gift of *Plasmodium berghei*, James Hogan for mosquito rearing, Drs. Nora Besansky and Michael Ferdig for their constructive comments during the development of this research, and Dr. Rick Goetz and Priscilla Duman for assistance with technical aspects of cDNA library construction. This research would not have been possible without the assistance of the Freimann Life Sciences Center with animal care and Giemsa slide preparation. This project was supported by grants U01-AI48846 and R01-AI44273 from NIH/NIAID to F.H.C. Additional support was provided by a discovery grant to P.R. from the Natural Sciences and Engineering Research Council of Canada.

## References

- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusser DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscos D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL: **The genome sequence of the malaria mosquito *Anopheles gambiae***. *Science* 2002, **298(5591)**:129-149.
- Mongin E, Louis C, Holt RA, Birney E, Collins FH: **The *Anopheles gambiae* genome: an update**. *Trends In Parasitology* 2004, **20(2)**:49-52.
- Clegg N, Eroglu B, Ferguson C, Arnold H, Moorman A, Nelson PS: **Digital expression profiles of the prostate androgen-response program**. *Journal of Steroid Biochemistry and Molecular Biology* 2002, **80(1)**:13-23.
- Hwang DM, Dempsey AA, Lee CY, Liew CC: **Identification of differentially expressed genes in cardiac hypertrophy by analysis of expressed sequence tags**. *Genomics* 2000, **66(1)**:1-14.
- Okubo K, Itoh K, Fukushima A, Yoshii J, Matsubara K: **Monitoring Cell Physiology By Expression Profiles And Discovering Cell-Type-Specific Genes By Compiled Expression Profiles**. *Genomics* 1995, **30(2)**:178-186.
- Carulli JP, Artinger M, Swain PM, Root CD, Chee L, Tulig C, Guerin J, Osborne M, Stein G, Lian J, Lomedico PT: **High throughput analysis of differential gene expression**. *Journal of Cellular Biochemistry* 1998:286-296.
- Audic S, Claverie JM: **The significance of digital gene expression profiles**. *Genome Research* 1997, **7(10)**:986-995.
- Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries**. *Genome Research* 2000, **10(12)**:2055-2061.
- Clements AN: **The Biology of Mosquitoes. Volume 1**. London: Chapman & Hall; 1992.
- Raikhel AS: **Hormonal control of reproductive processes**. In *Comprehensive Molecular Insect Science Volume 3*. Edited by: Gilbert LI, Iatrou K, Gill SS. Oxford: Elsevier Pergamon; 2004:433-491.
- Romoser WS: **The vector alimentary system**. In *The Biology of Disease Vectors* Edited by: Beaty BJ, Marquardt WC, Niwot CO: University Press of Colorado; 1996:298-317.
- Dana AN, Hong YS, Kern MK, Hillenmeyer ME, Harker BW, Lobo NF, Hogan JR, Romans P, Collins FH: **Gene expression patterns associated with blood-feeding in the malaria mosquito *Anopheles gambiae***. *BMC Genomics* 2005, **6**:5.
- Marinotti O, Nguyen QK, Calvo E, James AA, Ribeiro JMC: **Microarray analysis of genes showing variable expression following a blood meal in *Anopheles gambiae***. *Insect Molecular Biology* 2005, **14(4)**:365-373.
- Prevot GI, Laurent-Winter C, Rodhain F, Bourgoignie C: **Sex-specific and blood meal-induced proteins of *Anopheles gambiae* midguts: analysis by two-dimensional gel electrophoresis**. *Malaria Journal* 2003, **2**:1.
- Ribeiro JMC: **A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal**. *Insect Biochemistry and Molecular Biology* 2003, **33(9)**:865-882.
- Sanders HR, Evans AM, Ross LS, Gill SS: **Blood meal induces global changes in midgut gene expression in the disease vector, *Aedes aegypti***. *Insect Biochemistry and Molecular Biology* 2003, **33(11)**:1105-1122.
- Abraham EG, Islam S, Srinivasan P, Ghosh AK, Valenzuela JG, Ribeiro JMC, Kafatos FC, Dimopoulos G, Jacobs-Lorena M: **Analysis of the Plasmodium and Anopheles transcriptional repertoire during ookinete development and midgut invasion**. *Journal of Biological Chemistry* 2004, **279(7)**:5573-5580.
- Ahmed AM, Maingon R, Romans P, Hurd H: **Effects of malaria infection on vitellogenesis in *Anopheles gambiae* during two gonotrophic cycles**. *Insect Molecular Biology* 2001, **10(4)**:347-356.
- Collins FH, Sakai RK, Vernick KD, Paskewitz S, Seeley DC, Miller LH, Collins WE, Campbell CC, Gwadz RW: **Genetic Selection Of A Plasmodium-Refractory Strain Of The Malaria Vector *Anopheles gambiae***. *Science* 1986, **234(4776)**:607-610.
- Lowenberger CA, Kamal S, Chiles J, Paskewitz S, Bulet P, Hoffmann JA, Christensen BM: **Mosquito - Plasmodium interactions in response to immune activation of the vector**. *Experimental Parasitology* 1999, **91(1)**:59-69.
- Ensembl Mosquito Genome** [[http://www.ensembl.org/Anopheles\\_gambiae](http://www.ensembl.org/Anopheles_gambiae)]
- Boguski MS, Tolstoshev CM, Bassett DE: **Gene Discovery In dbEST**. *Science* 1994, **265(5181)**:1993-1994.
- Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE: **Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee**. *Genome Research* 2002, **12(4)**:555-566.
- ClustalW WWW Service at the European Bioinformatics Institute** [<http://www.ebi.ac.uk/clustalw/>]
- Thompson JD, Higgins DG, Gibson TJ: **Clustal-W - Improving The Sensitivity Of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties And Weight Matrix Choice**. *Nucleic Acids Research* 1994, **22(22)**:4673-4680.
- Mathe C, Sagot MF, Schiex T, Rouze P: **Current methods of gene prediction, their strengths and weaknesses**. *Nucleic Acids Res* 2002, **30(19)**:4103-4117.
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC: **Complementary-DNA Sequencing - Expressed Sequence Tags And Human Genome Project**. *Science* 1991, **252(5013)**:1651-1656.
- Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE, Smith CD, Tupy JL, Whitfield EJ, Bayraktaroglu L, Bertram BP, Bettencourt BR, Celniker SE, de Grey AD, Drysdale RA, Harris NL, Richter J, Russo S, Schroeder AJ, Shu SQ, Stapleton M, Yamada C, Ashburner M, Gelbart WM, Rubin GM, Lewis SE: **Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review**. *Genome Biol* 2002, **3**:12.
- Sheehan D, Meade G, Foley VM, Dowd CA: **Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily**. *Biochemical Journal* 2001, **360**:1-16.
- De Gregorio E, Spellman PT, Rubin GM, Lemaitre B: **Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98(22)**:12590-12595.
- De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaitre B: **The Toll and Imd pathways are the major regulators of the immune response in *Drosophila***. *EMBO Journal* 2002, **21(11)**:2568-2579.
- Imler JL, Zheng LB: **Biology of Toll receptors: lessons from insects and mammals**. *Journal of Leukocyte Biology* 2004, **75(1)**:18-26.
- Irving P, Troxler L, Heuer TS, Belvin M, Kopczynski C, Reichhart JM, Hoffmann JA, Hetru C: **A genome-wide analysis of immune responses in *Drosophila***. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98(26)**:15119-15124.
- Dimopoulos G, Richman A, della Torre A, Kafatos FC, Louis C: **Identification and characterization of differentially expressed cDNAs of the vector mosquito, *Anopheles gambiae***. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93(23)**:13066-13071.
- Dimopoulos G, Richman A, Muller HM, Kafatos FC: **Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94(21)**:11508-11513.
- Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, Kafatos FC: **Genome expression analysis of**

- Anopheles gambiae** : Responses to injury, bacterial challenge, and malaria infection. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(13):8814-8819.
37. Xu X, Dong Y, Abraham EG, Kocan A, Srinivasan P, Ghosh AK, Sinden RE, Ribeiro JM, Jacobs-Lorena M, Kafatos FC, Dimopoulos G: **Transcriptome analysis of *Anopheles stephensi*-*Plasmodium berghei* interactions.** *Molecular and Biochemical Parasitology* 2005, **142**(1):76-87.
  38. Nappi AJ, Ottaviani E: **Cytotoxicity and cytotoxic molecules in invertebrates.** *Bioessays* 2000, **22**(5):469-480.
  39. Nappi AJ, Sugumaran M: **Some biochemical aspects of eumelanin formation in insect immunity.** In *Insect Immunity Volume 48*. Edited by: Pathak JPN. London: Kluwer; 1993:131-148.
  40. Luckhart S, Vodovotz Y, Cui LW, Rosenberg R: **The mosquito *Anopheles stephensi* limits malaria parasite development with inducible synthesis of nitric oxide.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(10):5700-5705.
  41. Han YS, Thompson J, Kafatos FC, Barillas-Mury C: **Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes.** *Embo Journal* 2000, **19**(22):6030-6040.
  42. Dimopoulos G, Seeley D, Wolf A, Kafatos FC: **Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle.** *Embo Journal* 1998, **17**(21):6115-6123.
  43. Tahar R, Boudin C, Thiery I, Bourgooin C: **Immune response of *Anopheles gambiae* to the early sporogonic stages of the human malaria parasite *Plasmodium falciparum*.** *Embo Journal* 2002, **21**(24):6673-6680.
  44. Cotran RS, Kumar V, Collins T: **Robbins Pathological Basis of Disease.** 6th edition. Philadelphia: WB Saunders; 1999.
  45. Fernandez-Checa JC: **Redox regulation and signaling lipids in mitochondrial apoptosis.** *Biochemical and Biophysical Research Communications* 2003, **304**(3):471-479.
  46. Richman AM, Dimopoulos G, Seeley D, Kafatos FC: ***Plasmodium* activates the innate immune response of *Anopheles gambiae* mosquitoes.** *EMBO Journal* 1997, **16**(20):6114-6119.
  47. Paskewitz SM, Christensen BM: **Immune responses of vectors.** In *The Biology of Disease Vectors* Edited by: Beaty BJ, Marquardt WC. Niwot CO: University Press of Colorado; 1996:371-392.
  48. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa NT, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Moita LF, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C, Xu J, Zheng L, Bork P, Kafatos FC: **Immunity-related genes and gene families in *Anopheles gambiae*.** *Science* 2002, **298**(5591):159-165.
  49. Zheng XL, Zheng L: **Genomic organization and regulation of three Cecropin genes in *Anopheles gambiae*.** *Insect Molecular Biology* 2002, **11**(6):517-525.
  50. Vizioli J, Bulet P, Charlet M, Lowenberger C, Blass C, Muller HM, Dimopoulos G, Hoffmann J, Kafatos FC, Richman A: **Cloning and analysis of a cecropin gene from the malaria vector mosquito, *Anopheles gambiae*.** *Insect Molecular Biology* 2000, **9**(1):75-84.
  51. Lu JH, Teh C, Kishore U, Reid KBM: **Collectins and ficolins: sugar pattern recognition molecules of the mammalian innate immune system.** *Biochimica et Biophysica Acta-General Subjects* 2002, **1572**(2-3):387-400.
  52. Reid KB, Colomb M, Petry F, Loos M: **Complement component C1 and the collectins-first-line defense molecules in innate and acquired immunity.** *Trends Immunol* 2002, **23**:115-117.
  53. Vandivier RW, Ogden CA, Fadok VA, Hoffmann PR, Brown KK, Botto M, Walport MJ, Fisher JH, Henson PM, Greene KE: **Role of surfactant proteins D, D, and C1q in the clearance of apoptotic cells in vivo and in vitro: Calreticulin and CD91 as a common collectin receptor complex.** *Journal of Immunology* 2002, **169**(7):3978-3986.
  54. Matsushita M, Fujita T: **The role of ficolins in innate immunity.** *Immunobiology* 2002, **205**(4-5):490-497.
  55. Zakeri Z, Bursch W, Tenniswood M, Lockshin RA: **Cell-Death – Programmed, Apoptosis, Necrosis, or Other.** *Cell Death and Differentiation* 1995, **2**(2):87-96.
  56. Kang DW, Romans P, Lee JY: **Analysis of a lysozyme gene from the malaria vector mosquito, *Anopheles gambiae*.** *Gene* 1996, **174**(2):239-244.
  57. Hultmark D: **Insect lysozymes.** In *Lysozymes: Model Enzymes in Biochemistry and Biology* Basel JP: Birkhauser; 1996:87-102.
  58. Morishima I, Horiba T, Iketani M, Nishioka E, Yamano Y: **Parallel induction of cecropin and lysozyme in larvae of the silkworm, *Bombyx mori*.** *Developmental And Comparative Immunology* 1995, **19**(5):357-363.
  59. Vizioli J, Bulet P, Hoffmann JA, Kafatos FC, Muller HM, Dimopoulos G: **Gambicin: A novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(22):12630-12635.
  60. Luna C, Wang XL, Huang YM, Zhang JA, Zheng LB: **Characterization of four Toll related genes during development and immune responses in *Anopheles gambiae*.** *Insect Biochemistry and Molecular Biology* 2002, **32**(9):1171-1179.
  61. Gorman MJ, Andreeva OV, Paskewitz SM: **Molecular characterization of five serine protease genes cloned from *Anopheles gambiae* hemolymph.** *Insect Biochemistry and Molecular Biology* 2000, **30**(1):35-46.
  62. Paskewitz SM, Reese-Stardy S, Gorman MJ: **An easter-like serine protease from *Anopheles gambiae* exhibits changes in transcript abundance following immune challenge.** *Insect Molecular Biology* 1999, **8**(3):329-337.
  63. Danielli A, Kafatos FC, Loukeris TG: **Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion.** *Journal of Biological Chemistry* 2003, **278**(6):4184-4193.
  64. Hamilton JV, Munks RJJ, Lehane SM, Lehane MJ: **Association of midgut defensin with a novel serine protease in the blood-sucking fly *Stomoxys calcitrans*.** *Insect Molecular Biology* 2002, **11**(3):197-205.
  65. Grunclova L, Fouquier H, Hypsa V, Kopacek P: **Lysozyme from the soft tick *Ornithodoros moubata*: the sequence, phylogeny and post-feeding regulation.** *Developmental and Comparative Immunology* 2003, **27**(8):651-660.
  66. Nakajima Y, Natori S: **Identification and characterization of an anterior fat body protein in an insect.** *Journal of Biochemistry* 2000, **127**(5):901-908.
  67. Nakajima Y, van Naters-Yasui AV, Taylor D, Yamakawa M: **Two isoforms of a member of the arthropod defensin family from the soft tick, *Ornithodoros moubata* (Acari: Argasidae).** *Insect Biochemistry and Molecular Biology* 2001, **31**(8):747-751.
  68. Nakajima Y, van Naters-Yasui AV, Taylor D, Yamakawa M: **Antibacterial peptide defensin is involved in midgut immunity of the soft tick, *Ornithodoros moubata*.** *Insect Molecular Biology* 2002, **11**(6):611-618.
  69. Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Hieter P, Vogelstein B, Kinzler KW: **Characterization of the yeast transcriptome.** *Cell* 1997, **88**(2):243-251.
  70. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD: **Molecular Biology of the Cell.** New York: Garland; 1994.
  71. Bishop JO, Morton JG, Rosbash M, Richardson M: **Three abundance classes in HeLa cell messenger RNA.** *Nature* 1974, **250**:199-204.
  72. Hastie ND, Bishop JO: **The expression of three abundance classes of messenger RNA in mouse tissues.** *Cell* 1976, **9**:761-774.
  73. Lee NH, Weinstock KG, Kirkness EF, Earleughes JA, Fuldner RA, Marmaros S, Glodek A, Gocayne JD, Adams MD, Kerlavage AR: **Comparative Expressed-Sequence-Tag Analysis Of Differential Gene-Expression Profiles In Pc-12 Cells Before And After Nerve Growth-Factor Treatment.** *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**(18):8303-8307.
  74. Claverie JM: **Computational methods for the identification of differential and coordinated gene expression.** *Human Molecular Genetics* 1999, **8**(10):1821-1832.
  75. Cho WL, Tsao SM, Hays AR, Walter R, Chen JS, Snigirevskaya ES, Raikhel AS: **Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor.** *Journal of Biological Chemistry* 1999, **274**(19):13311-13321.

76. Sokal RR, Rohlf FJ: **Biometry: The Principles and Practice of Statistics In Biological Research**. 3rd edition. New York: Freeman; 1995.
77. Edwards MJ, Severson DW, Hagedorn HH: **Vitellic envelope genes of the yellow fever mosquito, *Aedes aegypti***. *Insect Biochemistry and Molecular Biology* 1998, **28(12)**:915-925.
78. Han YS, Barillas-Mury C: **Implications of Time Bomb model of ookinete invasion of midgut cells**. *Insect Biochemistry and Molecular Biology* 2002, **32(10)**:1311-1316.
79. Shen ZC, Dimopoulos G, Kafatos FC, Jacobs-Lorena M: **A cell surface mucin specifically expressed in the midgut of the malaria mosquito *Anopheles gambiae***. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96(10)**:5610-5615.
80. Carraway KL, Ramsauer VP, Haq B, Carrothers Carraway CA: **Cell signaling through membrane mucins**. *Bioessays* 2003, **25(1)**:66-71.
81. Morishima N: **Control of cell fate by Hsp70: More than an evanescent meeting**. *Journal of Biochemistry* 2005, **137(4)**:449-453.
82. Shen ZC, Jacobs-Lorena M: **A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin - Cloning, expression, and characterization**. *Journal of Biological Chemistry* 1998, **273(28)**:17665-17670.
83. Langer RC, Li FW, Popov V, Kurosky A, Vinetz JM: **Monoclonal antibody against the *Plasmodium falciparum* chitinase, PfCHT1, recognizes a malaria transmission-blocking epitope in *Plasmodium gallinaceum* ookinetes unrelated to the chitinase PgCHT1**. *Infection and Immunity* 2002, **70(3)**:1581-1590.
84. Langer RC, Vinetz JM: ***Plasmodium* ookinete-secreted chitinase and parasite penetration of the mosquito peritrophic matrix**. *Trends In Parasitology* 2001, **17(6)**:269-272.
85. Tsuboi T, Kaneko O, Eitoku C, Suwanabun N, Sattabongkot J, Vinetz JM, Torii M: **Gene structure and ookinete expression of the chitinase genes of *Plasmodium vivax* and *Plasmodium yoelii***. *Molecular and Biochemical Parasitology* 2003, **130(1)**:51-54.
86. Li FW, Templeton TJ, Popov V, Comer JE, Tsuboi T, Torii M, Vinetz JM: ***Plasmodium* ookinete -secreted proteins secreted through a common micronemal pathway are targets of blocking malaria transmission**. *Journal of Biological Chemistry* 2004, **279(25)**:26635-26644.
87. Muller HM, Catteruccia F, Vizioli J, DellaTorre A, Crisanti A: **Constitutive and blood meal-induced trypsin genes in *Anopheles gambiae***. *Experimental Parasitology* 1995, **81(3)**:371-385.
88. Muller HM, Crampton JM, Dellatorre A, Sinden R, Crisanti A: **Members Of A Trypsin Gene Family in *Anopheles gambiae* are Induced in the Gut by Blood Meal**. *Embo Journal* 1993, **12(7)**:2891-2900.
89. Vizioli J, Catteruccia F, della Torre A, Reckmann I, Muller HM: **Blood digestion in the malaria mosquito *Anopheles gambiae* - Molecular cloning and biochemical characterization of two inducible chymotrypsins**. *European Journal of Biochemistry* 2001, **268(14)**:4027-4035.
90. Lemos FJA, Cornel AJ, JacobsLorena M: **Trypsin and aminopeptidase gene expression is affected by age and food composition in *Anopheles gambiae***. *Insect Biochemistry and Molecular Biology* 1996, **26(7)**:651-658.
91. Gwadz R, Collins FH: **Anopheline mosquitoes and the agents they transmit**. In *The Biology of Disease Vectors* Edited by: Beaty BJ, Marquardt WC. Niwot CO: University Press of Colorado; 1996:73-83.
92. Sinden RE: **Infection of mosquitoes with rodent malaria**. In *Molecular Biology of Insect Disease Vectors: A methods manual* Edited by: Crampton JM, Beard CB, Louis C. New York: Chapman and Hall; 1997:67-91.
93. Vernick KD, Fujioka H, Seeley DC, Tandler B, Aikawa M, Miller LH: ***Plasmodium gallinaceum* - A Refractory Mechanism of Ookinete Killing in the Mosquito, *Anopheles gambiae***. *Experimental Parasitology* 1995, **80(4)**:583-595.
94. Rohlf FJ, Sokal RR: **Statistical Tables**. New York: Freeman; 1995.
95. **BLAST** [<http://www.ncbi.nlm.nih.gov/BLAST/>]
96. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Research* 1997, **25(17)**:3389-3402.
97. **The Institute for Genomic Research (TIGR)** [<http://www.tigr.org/tdb/tgi/>]
98. **BDGP: GadFly: Genome Annotation Database of Drosophila** [<http://www.fruitfly.org/annot/>]
99. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Perteza G, Sultana R, White J: **The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species**. *Nucleic Acids Research* 2001, **29(1)**:159-164.
100. Quackenbush J, Liang F, Holt I, Perteza G, Upton J: **The TIGR Gene Indices: reconstruction and representation of expressed gene sequences**. *Nucleic Acids Research* 2000, **28(1)**:141-145.
101. **KEGG: Kyoto Encyclopedia of Genes and Genomes** [<http://www.genome.ad.jp/kegg/>]
102. **The Gene Ontology Consortium** [<http://www.geneontology.org/>]
103. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene Ontology: tool for the unification of biology**. *Nature Genetics* 2000, **25(1)**:25-29.
104. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA: **IDE6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments**. *Physiological Genomics* 2003, **12(2)**:159-162.
105. Richman AM, Bulet P, Hetru C, BarillasMury C, Hoffmann JA, Kafatos FC: **Inducible immune factors of the vector mosquito *Anopheles gambiae* : Biochemical purification of a defensin antibacterial peptide and molecular cloning of preprodefensin cDNA**. *Insect Molecular Biology* 1996, **5(3)**:203-210.
106. Salazar CE, Mills-Hamm D, Kumar V, Collins FH: **Sequence of a cDNA from the Mosquito *Anopheles gambiae* Encoding a Homolog of Human Ribosomal Protein-S7**. *Nucleic Acids Research* 1993, **21(17)**:4147-4147.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

