Published in final edited form as:

Insect Biochem Mol Biol. 2014 January ; 44: 33–43. doi:10.1016/j.ibmb.2013.11.002.

Changes in transcript abundance for cuticular proteins and other genes three hours after a blood meal in *Anopheles gambiae*

Laura Vannini^a, W. Augustine Dunn^b, Tyler W. Reed^a, and Judith H. Willis^a

Laura Vannini: vannini@uga.edu; W. Augustine Dunn: dunnw@uci.edu; Tyler W. Reed: treed36@uga.edu; Judith H. Willis: jhwillis@uga.edu

^aUniversity of Georgia, Cellular Biology, Athens, GA 30602, USA

^bUniversity of California Irvine, Molecular Biology and Biochemistry, Irvine, CA 92697, USA

Abstract

Numerous studies have examined changes in transcript levels after *Anopheles gambiae* takes a blood meal. Marinotti et al. (2006) used microarrays and reported massive changes in transcript levels 3 h after feeding (BF3h) compared to non-blood fed (NBF). We were intrigued by the number of transcripts for structural cuticular proteins (CPs) that showed such major differences in levels and employed paired-end (50 bp) RNA-seq technology to compare whole body transcriptomes from 5-day-old females NBF and BF3h. We detected transcripts for the majority of CPs (164/243) but levels of only 12 were significantly altered by the blood meal. While relative transcript levels of NBF females were somewhat similar to the microarray data, there were major differences in BF3h animals, resulting in levels of many transcripts, both for CPs and other genes changing in the opposite direction. We compared our data also to other studies done with both microarrays and RNA-seq. Findings were consistent that a small number of CP genes have transcripts that persist even in 5-day-old adults. Some of these transcripts showed diurnal rhythms (Rund et al., 2013; Rinker et al., 2013). *In situ* hybridization revealed that transcripts for several of these CP genes were found exclusively or predominantly in the eye. Transcripts other than for CPs that changed in response to blood-feeding were predominantly expressed in midgut and Malpighian tubules. Even in these tissues, genes responsible for proteins with similar functions, such as immunity or digestion, responded differently, with transcript levels for some rising and others falling. These data demonstrate that genes coding for some CPs are dynamic in expression even in adults and that the response to a blood meal is rapid and precisely orchestrated.

Keywords

Blood feeding; RNA-seq; Illumina sequencing; Microarray; Cuticular proteins; Diurnal cycles

© 2013 Elsevier Ltd. All rights reserved.

Corresponding author Dr. Judith H. Willis, Department of Cellular Biology, 724 Biological Science Building, University of Georgia, Athens GA, 30602, +1 706-542-0802, jhwillis@uga.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Blood feeding by mosquitoes presents a major physiological challenge to the mosquito and is the essential first step in vector transmission. Thus the transcriptional response to blood feeding has been the focus of numerous studies. Our analysis was stimulated by Marinotti et al. (2006), a microarray analysis with *Anopheles gambiae* that described a massive increase in transcript levels 3 h after blood feeding (BF3h) compared to non-blood fed (NBF) females. They reported 1461 genes with transcripts that increased 2-fold or more, 299 of these more than 10-fold. Among those transcripts was a substantial number that corresponded to genes coding for cuticular proteins (CPs), our particular interest. This rapid increase was surprising because by three days post-eclosion one would anticipate that most CP synthesis should have ceased and there was no obvious explanation for an upsurge in transcription for CP genes after blood feeding since mosquitoes are reported to accommodate the blood meal by unfolding their intersegmental membranes (Klowden and Lea, 1979). Layers are added to the thoracic phragma cuticle up to 13 days after eclosion, but this represents only a minute cuticular region and the totality of growth after the first day is about half what is seen for the first day (Schlein and Gratz, 1973). We wanted to confirm and extend the finding of massive CP transcript increase that accompanies blood feeding in *An. gambiae*. In addition to CP transcripts, we report on the entire set of protein coding genes.

An. gambiae devotes almost 2% (243) of its protein coding genes to structural cuticular proteins. These have been classified into 13 families based on sequence characteristics (See Willis, 2010 for review). The largest family, CPR, is defined by the presence of the Rebers and Riddiford (R&R) Consensus (pfam00379), now known to confer chitin-binding properties (Rebers and Willis, 2001).

There is experimental evidence in other species that members of the TWDL family (12 members in *An. gambiae*) and the CPAP1 (one member) and CPAP3 (seven members) also bind chitin (Tang et al., 2010; Jasrapuria et al., 2010). Other families have distinct sequence motifs, but their roles in the cuticle are not known. There is precedence for increased CP deposition after blood feeding. Cuticular plasticization to accommodate the blood meal in nymphal *Rhodnius* and *Triatoma* occurs rapidly, probably involving simple physical changes in cuticle properties. But in the hard, ixodid tick, *Amblyomma*, the endocuticle also increases in weight (Kaufman et al., 2010). In queen termites there is a massive increase in the weight of the adult cuticle, albeit it slowly over a long time (Bordereau, 1982). So, we thought it prudent to examine further the possibility of CP involvement in response to blood feeding and to learn more about transcripts of CP genes in adults.

We used Illumina sequencing to examine the global transcript response to blood feeding, comparing NBF to BF3h animals, and failed to find such a dynamic response with either CPs or other transcripts. We identified only 33 transcripts from the entire transcriptome that significantly increased 10-fold or more (Supplementary Table 1, TAB 3). When we compared these 33 to the microarray analysis of Marinotti et al. (2006), we found major differences discussed in detail in Section 3.4.

We compared our CP data to that generated in other studies, some while not studying the response to blood feeding, do confirm the presence of many CP transcripts in the adult and reveal strong diurnal rhythms for transcript levels of CPs and other genes. We compared CP data to previously published data from our laboratory that used RT-qPCR to measure CP transcripts across development and carried out a very limited number of RT-qPCR comparisons between NBF and BF3h samples. In addition, we looked at the location of several CP transcripts with *in situ* hybridization with surprising results.

For our analysis of the entire transcriptome, we used data from Baker et al. (2011) to learn what organs were responding so quickly to the blood meal and we organized the transcripts that were most responsive into functional classes.

Finally, we discuss the many reasons why our BF3h data, both for CPs and for other transcripts, might differ so fundamentally from the microarray study of Marinotti et al. (2006).

2. Materials and methods

2.1. Insects

An. gambiae used in this experiment were from the G3 strain, maintained at the University of Georgia in a humidified insectary kept at 28°C and a 16:8 L:D photoperiod and fed powdered Tetramin Fish food. Males and females were kept together after eclosion and provided with 8% fructose. Females, five days after eclosion, were moved to a small chamber for feeding either directly or after 20 min exposure to 9°C. Feeding was on a human arm, over a 10 min period. Non-blood fed control females in a similar small chamber had access to water-soaked wicks. Feeding took place 4 h after light on. Three hours later, all animals were chilled for 20 min and then groups of five females were placed in 500µl Trizol and immediately placed on dry ice. Tubes were subsequently moved to -80°C until processed.

2.2. RNA isolation and preparation for Illumina

RNA was isolated following the Trizol protocol and DNA removed with Turbo DNA-free (Invitrogen). RNA integrity was monitored with an Agilent 2100 Bioanalyzer and quality assessed visually, using the image and accompanying scan, for the absence of degradation since RIN numbers do not apply to mosquito total RNA where the 18S and 28S peaks are not comparable to those obtained from mammals.

Satisfactory RNA preparations were sent to HudsonAlpha where eight libraries were prepared (four from NBF, four from BF3h) with poly-A selection and preparation of 200 bp fragments using the Illumina True-Seq RNA Sample Prep kit and analyzed with HiSEQ paired-end reads of 50 nt.

2.3. Data processing

Messenger RNA abundance was analyzed using the suite of programs known as the “Tuxedo Protocol”: bowtie, tophat, cufflinks, cummeRbund (<http://goo.gl/jLmnVG>). The process was automated using the blacktie pipeline script (<http://dx.doi.org/10.6084/m9.figshare.714149>). Detailed settings and output information is automatically recorded by blacktie at the time of the run. These have been made available on figshare.com (<http://dx.doi.org/10.6084/m9.figshare.768500>).

Briefly, alignments were performed by tophat using the reference GTF file for gene-build AgamP3.7 with the no-coverage-search option and filtering out loci annotated as rRNA. Cufflinks assembled replicate-specific transcriptomes using the reference GTF as a strict guide; frag-bias-correct, multi-read-correct, and upper-quartile-norm options were used. Cuffmerge was used to combine the replicate-specific transcriptomes with the AgamP3.7 reference GTF. Cuffdiff analyzed the transcript abundance differences using the four biological replicates with the following options: frag-bias-correct, multi-read-correct, and upper-quartile-norm.

Finally, exact replication of these analyses may be achieved by downloading the blacktie configuration file from figshare (http://files.figshare.com/1148319/2013.04.20_17_25_30.yaml), substituting the input files with your own, and running blacktie with the following versions of the tuxedo programs: Bowtie v2.0.5.0, TopHat v2.0.6, and Cufflinks v2.1.1.

2.4. In situ hybridization

Six blood-fed and six water fed females collected at the same time as the samples for RNA extraction were fixed in 4% paraformaldehyde in diethylpyrocarbonate (DEPC)-treated phosphate buffered saline (PBS) and processed by the Histology Laboratory at the University of Georgia College of Veterinary Medicine. Embedding was in ParaplastPlus, and sections were cut at 4 μ m. Additional animals at 24 h after pupation (a few hours before adult eclosion) and at specific intervals after eclosion were similarly processed.

Probes were made using the primers in Supplementary Table 2. Primers were originally designed with Primer3, but we now use Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). Although we began with cloning constructs and using different promoters to label sense and anti-sense strands, we now place a T7 promoter on the antisense (downstream) primer and make labeled probe directly from cDNA from animals at the appropriate stage. Probes were sequenced for verification.

In situ hybridization was carried out following a slightly simplified version of an EXIQON protocol (<http://www.exiqon.com/ls/Documents/Scientific/EDC-based-ISH-protocol.pdf>). In particular our protease K treatment was for only 10 min. Slides were washed pre-hybridization in 2X SSC (saline-sodium citrate), and a pre-hybridization buffer of 50% formamide, 10% dextran sulfate, 2X SSC, 10 ng/ml salmon testes ssDNA, 10 ng/ml yeast tRNA for 1h. After 16–20 h hybridization in 250 μ l between two slides with dig (digoxigenin)-labeled probes at 10 or 20 ng/100 μ l, generally at 55 or 65°C depending on the probe, slides were washed in 5X SSC, then 0.2X SSC, rinsed with PBS, incubated on the slide with anti-DIG-AP Fab fragment (Roche) for 2 h and stained in a Lock-Mailer (Ted Pella) with 5 ml detection solution including NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3'-indolylphosphate p-toluidine) for 1–48 h, rinsed in TE (Tris-EDTA) and dehydrated, soaked in xylene and covered with Cytoseal (Thermo Scientific). Control hybridizations using sense strand probes were carried out when we first began to use this technique. They were always negative, and given the different hybridization patterns with different probes, we stopped doing them. Examples of sense-strand controls are given in Supplementary Fig. 1. Hybridization by many probes is frequently found in the acellular lens of the eye, this is non-specific because it occurs even in sections that have been treated with RNase prior to hybridization (Supplementary Fig. 1).

2.5. RT-qPCR

RNA was isolated as described above, concentration was measured using a NanoDrop N-1000 (Thermo Scientific). First strand cDNA was synthesized from 1 μ g of total RNA using the Superscript III First Strand Synthesis Kit (Invitrogen) with oligo(dT) primers in a 20 μ l reaction. RT-qPCR was performed with Bio-Rad's CFX Connect Real Time System. All reactions were carried in triplicate (technical replicates) in a 15 μ l reaction containing 3.75 μ l of 1/100 diluted cDNAs (equivalent to starting with 5.6 ng of total RNA), 250 nM of each primer, and 7.5 μ l SsoAdvanced SYBR[®] Green Supermix (Bio-Rad). PCR conditions were 95°C for 2 min followed by 40 cycles of 95°C for 10 s and 57°C for 30 s. In addition to assuring that efficiency was between 92–104%, all primers pairs were verified to amplify a single gene by comparing reaction kinetics to known single copy genes *RpS7* (AGAP010592) and *chitin synthase* (AGAP001748). After the PCR reactions were

complete, melt curve analyses were done. We used two biological replicates (groups of five animals) for cDNA preparations for each condition. As an added check that there was no residual genomic DNA in the RNA preparations, the transcript level of *RpS7*, using primers that spanned an intron, was measured on every run. Primers used are provided in Supplementary Table 2. Ratios of expression between BF3h and NBF are based on the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). Data were normalized to *RpS7* that our RNA-seq data showed was not significantly different between BF3h and NBF animals.

2.6. Spearman Rank Order Correlations

A Spearman Rank Order Correlation analysis (Wessa, 2012) was run to compare our RNA-seq data with two previously published *An. gambiae* transcriptomes: Marinotti et al. (2006) and Pitts et al. (2011). Our data (Supplementary Table 1, TAB 2) were straightforward to compare to Pitts et al. (2011) using FPKM values per gene for 13,028 genes.

To compare the RNA-seq data to the microarray data, it was necessary to match the FPKM value with its corresponding microarray value. Many microarray probes recognized transcripts from two or more genes, and these were eliminated from the analysis, so a total of 9,602 value pairs were obtained. A correlation analysis was done also matching microarray data with data from Pitts et al. (2011) on NBF females. A total of 9,727 values pairs were obtained.

2.7. Sources of additional *An. gambiae* data gleaned from publications from other labs

We appreciate that different mosquito strains and different rearing and feeding conditions may affect the transcriptome. Hence we provide details on these conditions from the various published studies that we have compared to our results.

Microarray data from Marinotti et al. (2006)—This study was carried out with the Gene Chip Plasmodium/Anopheles genome array (Affymetrix) with RNA isolated from the pink-eye strain, raised at 25°C, 75–85% relative humidity, 18:6 L:D cycle. The adults were fed raisins and water, and fed three days after eclosion on anaesthetized mice. Data analyzed came from an updated annotation of the Affymetrix chip carried out by Dr. O. Marinotti and Dr. J.M. Ribeiro and provided by Marinotti. It differs a bit from the data from the same experiment available on the UC-Irvine *Anopheles gambiae* gene expression profile site: [http://www.angaged.bio.uci.edu/lookup2.php?geneid=Ag.2L.1919.0_CDS_at] and from their data as displayed on VectorBase [<https://www.vectorbase.org/>].

Microarray data from Rund et al. (2013)—This study provided data on diurnal cycles of CP genes obtained from BIOLOCK (<http://www3.nd.edu/~bioclock/>). Females (Pimperena S form), mated but not blood fed were maintained at 85% relative humidity, 27 ± 1°C. The study began with animals 7–9 days post-eclosion. Only data from animals raised in LD 12:12 were used and the q-value from the JTK CYCLE analysis had to be <0.05 to be reported. These data come from a microarray analysis using the same Affymetrix Chip that was used by Marinotti et al. (2006).

RNA-seq data from Rinker et al. (2013)—This Illumina study described transcriptomes from female antennae. It was carried out on both blood-fed (mouse) and non-blood fed mosquitoes of the SUA 2La/2La strain (M-form) originating from Suakoko, Liberia. These animals were raised at 27°C, on a 12:12 L:D photoperiod. Rhythmic behavior of transcripts in the non-blood fed mosquitoes was based on manual inspection of FPKM values (calculated by us based on their read numbers and transcript length) at 12 h intervals over a 48 h period. If a value was high at 1 h it had to be high at 24 and 48 h relative to 12

and 36 h in order to be accepted as indicating a diurnal cycle. At least one of these values had to be two or higher in order for the gene to be scored.

RNA-seq data from Pitts et al. (2011)—These data came from NBF female mosquitoes from the same colony as Rinker et al. (2013). Data used came from their Illumina transcript analysis of whole bodies of females 4–6 days after eclosion that had been kept with males.

RNA-seq data from Bonizzoni et al. (2012)—This study reported transcriptomes from insecticide resistant, *An. gambiae*. Mosquitoes were NBF adults reared from field collected larvae from Emutete, Kenya. We used the data from their single library of nine resistant animals (apparently both sexes) which had survived 1 h exposure to deltamethrin in the WHO tube bioassay and were then held for 24 h before RNA was extracted. Animals were 4–6 days post-eclosion when sampled. We chose to compare our data only to their resistant animals because their susceptible animals were obviously moribund as they were the first to be knocked down in the assay (Bonizzoni, personal communication).

Microarray data from Baker et al. (2011)—To learn the anatomical location of transcripts that changed in abundance in response to blood feeding, we used the data on VectorBase or the MozAtlas database (<http://www.tissue-atlas.org/>) that had been obtained from non-mated females (G3 strain) that had been blood fed three days after eclosion. Dissected tissues were taken at three-24 h intervals following the blood meal, pooled, and analyzed on the AffymetrixPlasmodium/Anopheles chip.

Comparison to *Aedes aegypti* data—We were unable to compare our data to that from *Aedes* because due to extensive gene amplification only a few orthologs for CPs can be identified (Cornman and Willis, 2008; 2009) and the time course of feeding and subsequent events relative to *Anopheles* has not been established.

3. Results and discussion

3.1. Overview of methodology to examine the *An. gambiae* transcriptome before and 3 h after the blood meal

In order to identify which transcripts for CPs and for other proteins change in abundance 3 h after the blood meal (BF3h), we employed pair-end reads (50 bp) RNA-seq technology to obtain whole body transcriptomes of 5-day-old mosquitoes at BF3h to compare with transcriptomes from non-blood fed (NBF) females. The transcript abundance for each of our eight samples was established by mapping the obtained read sequences against the annotated *An. gambiae* PEST genome 3 (AgamP3) in VectorBase (<https://www.vectorbase.org/navigation/downloads>) and then evaluated as FPKM (Fragments Per Kilobase per Million mapped reads). For each sample, we obtained ~41 million sequence reads (84.8% properly paired). Of 15,559 *An. gambiae* transcripts available, we found 14,008 (93.8%) that had FPKM >0 in NBF 5-day-old females. Supplementary Table 1, Tabs 1 and 2, sorted by isoforms and genes, respectively, contain the complete RNA-seq data. Our analyses are based in the isoforms data.

In order to assess the quality of our identification of differentially expressed transcripts, we first confirmed that the RNA-seq data was free of sequencing bias and displayed a uniform coverage across samples. The range of the FPKM values was evaluated creating a boxplot of log₁₀ transformed FPKM values for each replicate using CummeRbund (Goff et al., 2012) (Supplementary Fig. 2). The overall range and quartile distribution were consistent among replicates, indicating that the data were reproducible and of high quality. The median FPKM values among replicates were similar and slightly more than 1, indicating that very high

levels of sequence coverage allowed the identification of genes with very low levels of expression. These data also show how few differences there were between the two conditions.

CummeRbund was also used to generate a scatter plot that shows limited differences between NBF and BF3h. The FPKM values for all genes were plotted for the NBF and BF3h samples, following averaging across replicates and normalization (Fig. 1).

3.2. Structural cuticular proteins

Our Illumina data establish that mRNAs for many structural CPs are found in adults five days after eclosion. Previous work established that the cuticulome of *An. gambiae* has at least 243 genes that code for structural CPs (Cornman et al., 2008; Cornman and Willis, 2009; reviewed in Willis, 2010). A few came from our analysis of the G3 strain and are not present/annotated in the standard PEST genome. One new gene, *CPR157*, was added when we found two paired end reads that spanned a 13.7 kb intron that allowed us to complete its annotation by providing 62% of the signal peptide. An analysis by tandem mass spectrometry using exuvia (shed pupal cuticles or larval head capsules) (He et al., 2007) or other cuticle preparations (He and Willis, unpublished) revealed peptides, either unique or shared, for 90% of the CPs we had identified. Hence they are authentic, not putative, CPs. Another family, called CPLCP has 27 members, but only four have been confirmed by MS and only they were included in the total count and subsequent analysis. Transcripts were detected for 165 CPs in the complete set of RNA-seq data (Supplementary Table 1, TAB 4, but, importantly, in contrast to the situation in *An. gambiae* actively engaged in CP synthesis immediately before or after a molt (Togawa et al., 2008; Cornman and Willis, 2009), the levels of transcripts are low, with none of the CP genes having transcript levels higher than one-third that of *RpS7*. In our previous analyses carried out with RT-qPCR, CP transcript levels were as high as 35X *RpS7*, with 15 genes higher than 10X *RpS7*. These high levels were most evident 24 h after pupation, a few hours before adult eclosion. Even in larvae, where *RpS7* levels were approximately 5X higher than in pupae and pharate adults, several genes had transcript levels 2-8X *RpS7*.

If we use a conservative value of FPKM of 2, only 41 CP sequences were significantly expressed, six reaching this level only in the BF3h sample. These 41 genes belong to eight of the 13 CP families recognized in *An. gambiae* (Table 1; Supplementary Table 1, TAB 4). The majority (27) are CPRs, the largest family with 65% of the total CP sequences. Two distinct groups of the CPR family have been recognized, RR-1 and RR-2, with the former (n=47) frequently associated with flexible cuticle, the latter (n=102) with rigid cuticles, but it has also been suggested that RR-1 contributes to post-ecdysial endocuticle, while RR-2 will be present in exocuticle (Andersen, 2000). Nine could not be assigned using the classifier on cuticleDB (Magkrioti et al., 2004). Of the 22 significantly expressed CPRs that could be classified, 14 were RR-1 proteins. If the speculation about endo- and exo-cuticle is correct, the presence of eight transcripts of the RR-2 group is surprising, as the animals were taken five days after eclosion when they should only be producing endo-cuticle and that should be RR-1. One family of chitin-binding CPs, the CPAP3 family, is well represented in these data with five of its seven members, and the related CPAP1 family with one member in *An. gambiae* is also present. These families were only recently recognized as distinct from the peritrophins that contribute to the midgut matrix (Jasrapuria et al., 2010). They have cysteine-based chitin-binding domains (ChtBD2), whereas cysteine residues are rarely found in members of the other CP families (Willis, 2010).

We previously carried out an analysis by RT-qPCR of expression at 19 developmental periods for almost all (147/157) of the CPR genes, with the final set collected from adults within the first 12 h after eclosion (Togawa et al., 2008). The resulting data were clustered

into 21 co-expression groups that provide the best way to summarize the developmental periods when transcripts were detected. We have superimposed our RNA-seq data onto that summary (Fig. 2) showing the eight CPR genes whose transcripts changed significantly in response to blood feeding (Table 2). We also indicate the relative expression level of the 18 additional CPR genes that had transcript values greater than 2. Two of the co-expression groups (#3 and #10) contained 11 genes that were expressed in all developmental stages and seven of the genes with NBF transcript levels >2 resided in these groups. Eight were in groups with expression in pharate adults plus another stage; none were members of the four groups with a total of 30 genes where expression was exclusively in pharate adults. Three genes with high transcript levels were in group #11 that has four genes with expression in pharate adults and adults. None of the 26 CPR genes with transcript levels >2 belonged to sequence clusters that we suggested were formed to allow for the rapid development that is characteristic of many mosquitoes (Cornman and Willis, 2008).

We also evaluated the 53 CPR genes for which no expression was detected in either BF3h or NBF (FPKM=0) using the same co-expression summary table. Thirty-two were in sequence clusters, further evidence that these “amplified” genes are not active in the adult (Fig. 2).

Only seven CPs showed a significant increase in transcript level following the blood meal and five were lower. We used RT-qPCR with the same batch of animals that had been used for the Illumina analysis. Although there were quantitative differences, the direction of change in every case was confirmed (Table 3). We analyzed two additional batches of mosquitoes blood-fed months later but at the same time of day as the Illumina samples. There were quantitative differences in RT-qPCR data between the two groups, but with one exception (*CPLCX2*) the direction of change was the same (data not shown).

In order to learn what these transcripts might be doing in the adult, we carried out *in situ* hybridization to visualize where they were located (Fig. 3, Supplementary Table 3). The 16 genes analyzed (highlighted in Table 1) included those whose expression levels were in the top ten, three others had significantly altered expression levels, and a final three were those from the top 41 for which we already had data. We found almost no signal in the 5-day-old adults, whether or not they had been blood fed. The exceptions were the presence of label for *CPR10* in ovarian nurse cells (Fig. 3) and some faint epidermal labeling for *CPR75* (not shown). The absence of detectable labeling in these mature adults is not surprising given their very low levels of CP transcripts. Only when a specific transcript is restricted to a tiny anatomical area have we been able to see evidence of hybridization with low transcript levels (unpublished observations). But data from pharate adults taken 24 h after pupation (hence a few hours before adult eclosion) and from those within a few hours after eclosion were informative. At this stage, many CP transcript levels are an order of magnitude higher than RpS7 (Togawa et al., 2008).

Ten were expressed in eye cells and with two exceptions, (*CPR59* and *CPR75*), they were found in the four cone cells that along with primary pigment cells secrete the corneal lens and pseudocone that lies beneath it. For two genes, the eye was the only site where transcripts were found. *In situ* hybridization is far less sensitive than the methods that measure transcripts directly, so the absence of visible stain in other regions should not be over interpreted. *In situ* hybridization is only effective for visualizing abundant transcripts. Data available on MozAtlas, however, confirms that three of the genes, *CPR10*, *CPR15* and *CPR16*, with strong hybridization in the eye of pharate and/or young adults, had at least an eight-fold enrichment of transcript levels in the head over the whole body in blood fed females (Supplementary Table 3). It is surprising that transcripts from so many genes appear to be involved in forming the lens and pseudocone and that they belong to both groups of CPRs as well as other CP gene families. But the lenses of vertebrates also have a diversity of

proteins (Piatigorsky, 2007). Of course, we plan to learn whether proteins from these transcripts are actually present in the eyes.

While it is tempting to speculate, based on the localization in the protected environment of the eye, that some of the transcripts we detected in these mature adults might represent old but stable mRNAs, rather than new transcription occurring long after eclosion, data from Rund et al. (2013) revealed that many of these CP genes had transcripts that showed diurnal cycles, implying new transcription. Data on rhythms were available for 24 of the 26 CPR genes and 12 of the 14 in other families that had expression levels of FPKM >2. Twelve (46%) of the CPR genes had diurnal cycles as did nine (64%) of those in other families (Table 1). All but one, *CPR132*, of the genes we found with mRNA in the eye had diurnal cycles and there were no data available for that gene.

A tiny number of CP transcripts increased or decreased significantly 3 h after the blood meal. These results from our Illumina analysis, especially those in Fig. 2 showing that CP transcripts detected in adults correspond to specific categories of genes, combined with data described in Section 3.4, is compelling evidence that their presence is not due to leaky genes. If new cuticle is being secreted in adults, one would expect to find transcripts for the cuticular chitin synthase (AGAP001748). It is present, albeit at low levels and the increase in BF3h is not significant (Supplementary Table 1, Tab 1).

Illumina analyses by Bonizzoni et al. (2012), the Zwiebel laboratory (Pitts et al., 2011; Rinker et al., 2013) along with data from the microarray analyses of Rund et al. (2013) establish that respectable levels of transcripts from about 15% of CP genes can be found in “mature” adults and that a subset of these are dynamic in that their expression levels vary throughout the day and/or in response to a blood meal.

3.3. Significantly differentially expressed transcripts other than CPs

Including the 12 CP transcripts discussed in Section 3.2, 1809 transcripts were significantly differentially accumulated (q -value < 0.05) between NBF and BF3h mosquitoes. A total of 876 transcripts had different transcript levels with at least a 2-fold change (310 increased and 566 decreased) (Supplementary Table 1, Tab 3). To gain insight into the major processes that take place in the female upon blood-feeding, these differentially expressed transcripts were classified using Bast2GO (Conesa et al., 2005) and then manually grouped in functional classes following Marinotti et al. (2005) and Das et al. (2010) (Supplementary Table 1, Tab 3). All the functional classes considered (except for replication/transcription/translation) had a majority of transcripts that were reduced in abundance after the blood meal. This is most evident for the immunity and stress class that is also the group with most transcripts (136). An overview of the number of transcripts for each functional class is given in Fig. 4.

Supplementary Table 1, Tab 3 also gives the anatomical location of the transcripts in females that had been collected over three days after blood feeding based on data in Baker et al. (2011). It is immediately obvious that two of the six regions from females that they assayed predominate in having significantly changed transcript levels. These are the midgut and the Malpighian tubules. For the genes with transcripts that increased significantly at BF3h, 26% were expressed in midgut, Malpighian tubules or both. For those transcripts that changed in the opposite direction, 39% were predominantly expressed in these two tissues. The predominance of these two tissues is even more evident (Table 4) with only transcripts that increased or decreased by at least 10-fold. Here 74% of the top ranking genes for which data were available had transcripts located primarily in midgut or Malpighian tubules.

We also looked at the predicted function of the genes whose transcripts were most affected by the blood meal. *AGAP013705* topped the list because no transcripts were found in the NBF animals thus giving a fold-increase of at least 1100. It codes for the microRNA, aga-miR-989, that resides near the 5' region of *AGAP010412*. Our data apparently came from the pri-miRNA, because pre-miRNAs (~90 nucleotides) and mature miRNAs (~20–22 nucleotides) were too short to be included in the Illumina libraries. The transcript levels of the neighboring gene did not change significantly. The top five protein coding genes were a *synaptic vesicle protein*, the *protein g12 precursor*, the *vitellogenin 1 (Vg1)*, an *ornithine decarboxylase (ODC)* and a *phenylalanine hydroxylase (PAH)* (*AGAP004350-RA*, *AGAP006187-RA*, *AGAP004203-RA*, *AGAP011808-RA* and *AGAP005712-RB*, respectively) with increased transcript values (176-, 153-, 104-, 80- and 47-fold, respectively).

Synaptic vesicle proteins are involved in secondary transport facilitating the movement of neurotransmitters across synaptic membranes. Two other genes for synaptic vesicle proteins (*AGAP013063-RA* and *AGAP000732-RA*) were among those whose transcript values increased more than 10-fold. Importantly, the transcripts of all three of these genes were found primarily in the Malpighian tubules, making us wonder about their true function in *An. gambiae*. An *Aedes aegypti* gene (*AAEL005533*), the ortholog of *AGAP004350*, also had increased transcript levels 3 h after the blood meal (Dissanayake et al., 2010).

The g12 precursor is a regulator of actin cytoskeletal remodeling (Wang et al., 2006), and has been implicated in cytoskeletal reorganization in *Drosophila* ovarian follicle cells during maturation (Bohrmann and Biber, 1994). Once again the data of Baker et al. (2011) challenge this function for *An. gambiae* where the transcripts of this gene are over 400-fold more abundant in midgut than ovaries.

It is largely accepted that RNAs for vitellogenins are detectable a few hour after the blood meal in *An. gambiae* (Marinotti et al., 2006; Dana et al., 2005; Ribeiro, 2003) and rise by 12 h reaching maximal level by 24 h. *Vg1* is the only vitellogenin transcript that we found to be more abundant at 3 h after the blood meal.

Phenylalanine hydroxylase is the rate-limiting enzyme of the metabolic pathway producing tyrosine essential for synthesis of melanin, a component of immune capsules. Transcripts from this top ranking gene, like those for the vitellogenin, are far higher in carcass than in midgut or Malpighian tubules.

ODCs catalyzed a committed step in polyamine synthesis that is essential for normal vitellogenesis in mosquitoes (Kogan and Hagedorn, 2000). All of the three ODCs known in *An. gambiae* (*AGAP011808-RA*, *AGAP011806-RA* and *AGAP011807-RA*) had increased transcript levels (80-, 8- and 5-fold, respectively). The two, for which data were available (*AGAP011808* and *AGAP011806*), have transcript levels in the midgut at least 50-fold higher than either carcass or ovaries (Baker et al., 2011).

Among the top ranking genes whose transcript levels fall, it is apparent that there have been significant changes for genes involved in immunity and digestion, and here too midgut and Malpighian tubules are a major site where these transcripts were detected.

Our data, with but one time point after blood feeding, reveal that the response to blood feeding is precisely orchestrated, with transcripts expressed in the same tissue or acting in a similar manner, i.e. immunity, digestion, rising or falling significantly.

3.4. Comparisons of RNA-seq and microarray data

We initiated this study based on the microarray data of Marinotti et al. (2006) where of the 125 CP transcripts that were uniquely recognized with their microarray, 28 increased >10 fold and an additional 31 increased > 2-fold when NBF were compared to BF3h while 11 decreased >10-fold and 10 additional ones > 2-fold. Two sequences had transcripts that changed by more than 1000-fold, one increased, the other decreased, and four others increased and four decreased by more than 100-fold (Supplementary Table 4). These microarray data presented a dynamic and rapid response of CP transcripts to blood feeding with 47% of the CP transcripts increasing and 17% decreasing by at least 2-fold.

Our RNA-seq results with CPs are in marked contrast to these microarray results. Four of the seven genes where transcript levels increased in the Illumina analysis, fell in the microarray analysis, while those that decreased significantly in the Illumina analysis also decreased in the microarray data (Table 2). The greatest difference was for *CPR26* that increased 3-fold with Illumina and fell 2526-fold in the microarray data.

Fortunately, there are other data available to try to sort out these differences and establish what is actually going on with CP transcripts in adult *An. gambiae*. Most comparable was an Illumina analysis by Pitts et al. (2011) designed to identify transcripts specific to appendages that might be candidates for coding for sensory receptors. Although done on a different strain and fed on mice (See Section 2.7), other conditions were similar; animals were also five days post-eclosion. We compared the rank order of CP transcripts in the whole bodies of their non-blood fed females and found impressive agreement. Data were available for 31 of the 35 NBF transcripts with values >2 in our data and all but four were also in their top 35 most abundant CP transcripts (Table 1). Another recent RNA-seq analysis (Bonizzoni et al., 2012) was carried out on *An. gambiae* adults that came from larvae collected in Emutete, Kenya. Analysis of the relative expression levels also confirmed the majority of the CP genes with highest expression. Twenty-five (71%) of our top 35, were in the top 35 of these data. This is in marked contrast to the microarray data from Marinotti et al. (2006), where only 19 (54%) met this criterion (Table 1).

Rinker et al. (2013) used the same strain as Pitts et al. (2011) and restricted their analysis to transcripts in the antennae, examining with Illumina NBF and BF at times corresponding to 1, 12, 24, 36 and 48 h after a mouse blood meal. Two findings with implications for all analyses of *An. gambiae* transcripts are reported. First, many antennal transcripts show strong diurnal rhythms, only a small number of which they discussed, and there is a decoupling of diel rhythmicity in the blood fed. We identified strong rhythms for many CP transcripts in their data for NBF samples (Table 2), confirming the microarray results of Rund et al. (2013) discussed below. The other unexpected finding that may explain why so many CP transcripts are found in the antennae is that Rinker et al. (2013) detected transcripts for proteins that are known to be restricted to non-antennal tissues such as transcripts for vitellogenin precursors, proteases and other proteins associated with ovaries and embryos. Their sensible explanation is that diffusion of hormones or other strong transcriptional activation signals may elicit a transcriptional response from non-target tissues regardless of whether these transcripts will be effectively translated or mature peptides correctly processed and localized. Of the 12 CP transcripts in our data that changed significantly at BF3h, the data of Rinker et al. (2013) revealed 10 that changed in the same direction at 1 and/or 12 h after a blood meal (Table 2).

Just like what we found with the CP transcripts, there was little agreement between our data for the influence of blood-feeding on the complete set of transcripts and the microarray data from Marinotti et al. (2006). Table 4 compares the fold-increase or decrease that we found to that from Marinotti et al. (2006) for those 60 genes where transcripts levels changed by at

least 10-fold in our data. Half of these changed in the opposite direction in the microarray data. Conversely, 55% of the 110 genes that changed 10-fold or more in the data from Marinotti et al. changed significantly in the opposite direction in our data (Supplementary Table 1, Tab 5.) We were interested in learning if diurnal rhythms might play a role in this, so we examined the data of Rund et al. (2013) available on BIOLOCK. While many of those where there was conflict, had significant diurnal rhythms, they had peak expression values at different times of the day (Table 4). So, the possibility that diurnal rhythms could help to explain the lack of similarity between our results and those of the earlier analysis is unlikely.

The results of a Spearman Rank Correlation analysis (Wessa, 2012) are another indication of just how much our data differed from the microarray data of Marinotti et al. (2006). In addition to comparing our RNA-seq data with that microarray data, we also looked at the RNA-seq data from Pitts et al. (2011) from NBF 4–6-day-old females. The comparison with Pitts et al. (2011) gave an impressive degree of correlation for NBF mosquitoes ($\rho=0.90$). A far lower degree of correlation was observed between our RNA-seq data and the microarray data from Marinotti et al. (2006) for both NBF and BF3h mosquitoes ($\rho=0.70$ and 0.47 , respectively). A correlation analysis was done also matching the Marinotti et al. (2006) microarray and Pitts et al. (2011) RNA-seq data from NBF females; a $\rho=0.70$ was obtained.

Even though the correlation between our NBF and the microarray data in Marinotti et al. (2006) was lower than obtained with another RNA-seq study, the data are not totally incompatible. This suggests that the two different methods are not the basis for the enormous differences in the BF3h data from the two studies, especially the numerous cases where changes are in the opposite directions (Tables 2 and 4 and Supplementary Table 1, Tab 5). Indeed, other studies have found good agreement between the two methods (Malone and Oliver, 2011; Nookaew et al., 2012).

We have presented all the differences between the two studies in terms of animal husbandry in Section 2.7. While the different sources of blood meals (human vs. mouse) might seem to be critical, it must be remembered that our CP data and the RNA-seq data from mouse fed *An. gambiae* from Rinker et al. (2013) with just antennal transcripts and from whole animals from Pitts et al. (2011) were quite congruent. An additional factor is that our animals had been adults for two additional days (5 vs. 3) when fed. Since our goal had been to learn about CPs in adults rather than to confirm Marinotti et al. (2006), we chose to use slightly older animals. We have found, using RT-qPCR, that CP transcripts levels decline with time after adult eclosion (data not shown). Nonetheless, the difference between the two studies was far greater in the BF3h than in the NBF.

Unknown are subtle differences in handling, whether CO₂ or cold was used for immobilization. Possibly a significant difference would be the size of the blood meal and resulting expansion, and this could reflect the state of the animals before blood feeding. Most perplexing are the massively different numbers of significant differences and that so many transcripts that increased significantly in the Illumina analysis had decreased in the microarray analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Osvaldo Marinotti (University of California-Irvine) for providing up-to-date data for his microarray data and Dr. Mariangela Bonizzoni for information on sample collection. We are also grateful to Dr. Daniela Lino Lourenco (University of Georgia) and Dr. Eleonora Biagetti (Polytechnic University of Marche, Italy) for technical support in data analysis and to Dr. Scott Cornman and Dr. Hitoshi Tsujimoto for designing and making some of the *in situ* probes and Dr. Sarah Jardeleza and Chase Naples for helping with *in situ* hybridization. Drs. Cornman and Mark R. Brown provided useful comments on an early version of this MS and two anonymous referees made helpful suggestions. We also thank Dr. Brown and Anne Robertson for maintaining the mosquito facility from which these animals were obtained and Dr. Michael Strand for access to his Leica photomicroscope. Work was supported by a grant from the National Institutes of Health (AI55624).

Abbreviations

NBF	non-blood fed
BF3h	blood fed 3 h
CP	cuticular protein
RT-qPCR	real-time quantitative PCR
FPKM	fragments per kilobase per million mapped reads
Vg1	vitellogenin 1
ODC	ornithine decarboxylase
PAH	phenylalanine hydroxylase

References

- Andersen SO. Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. *Insect Biochem Mol Biol*. 2000; 30:569–577. [PubMed: 10844249]
- Baker DA, Nolan T, Fischer B, Pinder A, Crisanti A, Russell S. A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. *BMC Genomics*. 2011; 12:296. [PubMed: 21649883]
- Bohrmann J, Biber K. Cytoskeleton-dependent transport of cytoplasmic particles in previtellogenic to mid-vitellogenic ovarian follicles of *Drosophila*: time-lapse analysis using video-enhanced contrast microscopy. *J Cell Sci*. 1994; 107:849–858. [PubMed: 8056841]
- Bonizzoni M, Afrane Y, Dunn WA, Atieli FK, Zhou G, Zhong D, Li J, Githenjo A, Yan G. Comparative transcriptome analyses of deltamethrin-resistant and -susceptible *Anopheles gambiae* mosquitoes from Kenya by RNA-Seq. *PLoS ONE*. 2012; 7:e44607. [PubMed: 22970263]
- Bordereau C. Ultrastructure and formation of the physogastric termite queen cuticle. *Tissue Cell*. 1982; 14:371–396. [PubMed: 6214044]
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21:3674–3676. [PubMed: 16081474]
- Cornman RS, Togawa T, Dunn WA, He N, Emmons AC, Willis JH. Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae*. *BMC Genomics*. 2008; 9:22. [PubMed: 18205929]
- Cornman RS, Willis JH. Extensive gene amplification and concerted evolution within the CPR family of cuticular proteins in mosquitoes. *Insect Biochem Molec Biol*. 2008; 38:661–676. [PubMed: 18510978]
- Cornman RS, Willis JH. Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. *Insect Mol Biol*. 2009; 18:607–622. [PubMed: 19754739]
- 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. [PubMed: 15651988]

- Das S, Radtke A, Choi YJ, Mendes AM, Valenzuela JG, Dimopoulos G. Transcriptomic and functional analysis of the *Anopheles gambiae* salivary gland in relation to blood feeding. *BMC Genomics*. 2010; 11:566. [PubMed: 20946652]
- Dissanayake SN, Ribeiro JM, Wang MH, Dunn WA, Yan G, James AA, Marinotti O. aeGEPUCI: a database of gene expression in the dengue vector mosquito, *Aedes aegypti*. *BMC Res Notes*. 2010; 3:248. [PubMed: 20920356]
- Goff L, Trapnell C, Kelley D. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2.2.0. 2012
- He N, Botelho JM, McNall RJ, Belozero V, Dunn WA, Mize T, Orlando R, Willis JH. Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. *Insect Biochem Mol Biol*. 2007; 37:135–146. [PubMed: 17244542]
- Jasrapuria S, Arakane Y, Osman G, Kramer KJ, Beeman RW, Muthukrishnan S. Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function. *Insect Biochem Mol Biol*. 2010; 40:214–227. [PubMed: 20144715]
- Kaufman WR, Flynn PC, Reynolds SE. Cuticular plasticization in the ixodid tick *Amblyomma hebraeum* (Acari: Ixodidae): possible roles of monoamines and cuticular pH. *J Exp Biol*. 2010; 213:2820–2831. [PubMed: 20675552]
- Klowden MJ, Lea AO. Abdominal distention terminates subsequent host-seeking behavior of *Aedes aegypti* following a blood meal. *J Insect Physiol*. 1979; 25:583–585. [PubMed: 544697]
- Kogan PH, Hagedorn HH. Polyamines, and effects from reducing their synthesis during egg development in the yellow fever mosquito, *Aedes aegypti*. *J Insect Physiol*. 2000; 46:1079–1095. [PubMed: 10817834]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) method. *Methods*. 2001; 25:402–408. [PubMed: 11846609]
- Magkrioti CK, Spyropoulos IC, Ikonomidou VA, Willis JH, Hamodrakas SJ. cuticleDB: a relational database of Arthropod cuticular proteins. *BMC Bioinformatics*. 2004; 5:138. [PubMed: 15453918]
- Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol*. 2011; 9:34. [PubMed: 21627854]
- 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 Mol Biol*. 2005; 14:365–373. [PubMed: 16033430]
- Marinotti O, Calvo E, Nguyen QK, Dissanayake S, Ribeiro JM, James AA. Genome-wide analysis of gene expression in adult *Anopheles gambiae*. *Insect Mol Biol*. 2006; 15:1–12. [PubMed: 16469063]
- Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlén M, Nielsen J. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2012; 40:10084–10097. [PubMed: 22965124]
- Piatigorsky, J. Gene Sharing and Evolution: the diversity of protein functions. 4. Harvard University Press; Cambridge, Massachusetts: 2007.
- Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ. Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics*. 2011; 12:271. [PubMed: 21619637]
- Rebers JE, Willis JH. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem Mol Biol*. 2001; 31:1083–1093. [PubMed: 11520687]
- Ribeiro JMC. A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal. *Insect Biochem Mol Biol*. 2003; 33:865–882. [PubMed: 12915178]
- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006; 38:500–501. [PubMed: 16642009]
- Rinker DC, Pitts RJ, Zhou X, Suh E, Rokas A, Zwiebel LJ. Blood meal induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in *Anopheles gambiae*. *Proc Natl Acad Sci USA*. 2013; 110:8260–8265. [PubMed: 23630291]

- Rund SSC, Gentile JE, Duffield GE. Extensive circadian and light regulation of the transcriptome in the malaria mosquito *Anopheles gambiae*. BMC Genomics. 2013; 14:218. [PubMed: 23552056]
- Schlein Y, Gratz NG. Determination of the age of some anopheline mosquitos by daily growth layers of skeletal apodemes. Bull World Health Organ. 1973; 49:371–5. [PubMed: 4368577]
- Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA. 1999; 96:2907–2912. [PubMed: 10077610]
- Tang L, Liang J, Zhan Z, Xiang Z, He N. Identification of the chitin binding proteins from the larval proteins of silkworm, *Bombyx mori*. Insect Biochem Mol Biol. 2010; 40:228–234. [PubMed: 20149871]
- Togawa T, Dunn WA, Emmons AC, Nagao J, Willis JH. Developmental expression patterns of cuticular protein genes with the R&R Consensus from *Anopheles gambiae*. Insect Biochem Mol Biol. 2008; 38:508–519. [PubMed: 18405829]
- Wang D, Tan YC, Kreitzer GE, Nakai Y, Shan D, Zheng Y, Huang XY. G proteins G12 and G13 control the dynamic turnover of growth factor-induced dorsal ruffles. J Biol Chem. 2006; 281:32660–32667. [PubMed: 16943201]
- Wessa, P. Spearman Rank Correlation (v1.0.1) in Free Statistics Software (v1.1.23 -r7). Office for Research Development and Education; 2012. http://www.wessa.net/rwasp_spearman.wasp/
- Willis JH. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. Insect Biochem Mol Biol. 2010; 40:189–204. [PubMed: 20171281]

HIGHLIGHTS

We investigate the transcriptional response to blood feeding in 5-day-old *Anopheles gambiae* with particular attention to cuticular proteins (CPs).

Most transcripts that change at least 10-fold three hours after the blood meal are located in midgut or Malpighian tubules.

Transcripts from genes coding for proteins involved in digestion and immunity/stress are most affected.

Transcripts for many CPs (41) are present in adult mosquitoes and some (12) are differentially expressed after the blood meal.

In situ hybridization reveals that many of the “adult” cuticular protein transcripts are present in the eyes of pharate adults.

Fig. 1. Scatterplot generated using CummeRbund. The FPKM values for all transcripts were plotted for the NBF and BF3h samples, following averaging across replicates and normalization. Each dot represents data from one transcript. The solid line is the best fit, dotted line shows what would be equal expression in both conditions.

Co-expression Cluster	Stage	Gene	Group	Sequence Cluster
0	PP	CPR88	RR.2	3RB
		CPR82	RR.2	3RC
		CPR84+97+109	RR.2	3RC
		CPR85	RR.2	3RC
		CPR86	RR.2	3RC
		CPR100	RR.2	3RC
		CPR109	RR.2	3RC
		CPR83	RR.2	3RC
		CPR83+99 or 142	RR.2	3RC
		CPR87	RR.2	3RC
1	PP	CPR155	RR.2	3RC
		CPR1	RR.2	2RA
		CPR2	RR.2	2RA
		CPR3	RR.2	2RA
		CPR4	RR.2	2RA
		CPR5	RR.2	2RA
		CPR6	RR.2	2RA
		CPR24	RR.1	
		CPR28	RR.1	
		CPR78	RR.1	
2	PP-P	CPR111	RR.2	
		CPR125	RR.1	
		CPR79	RR.1	
		CPR140	RR.2	
		CPR26 >>	RR.1	
		CPR75	RR.1	
		CPR17	RR.2	2LA
		CPR18	RR.2	2LA
		CPR57	RR.2	2LC
		CPR72	RR.2	2LC
3	ALL	CPR119	RR.2	
		CPR120	RR.2	
		CPR145	RR.2	2LC
		CPR19	RR.2	2LA
		CPR20	RR.2	2LA
		CPR55	RR.2	
		CPR56	RR.2	
		CPR58	RR.2	2RB
		CPR115	RR.2	2RB
		CPR117	RR.2	2RB
4	PA	CPR118	RR.2	2RB
		CPR122	RR.2	2RB
		CPR123	RR.2	2RB
		CPR154	RR.2	2RB
		CPR28	RR.2	3RB
		CPR87	RR.2	3RB
		CPR88	RR.2	3RB
		CPR90	RR.2	3RB
		CPR148	RR.2	3RB
		CPR149	RR.2	3RB
5	PL+PP	CPR98	RR.2	
		CPR131	RR.2	
		CPR130	RR.2	
		CPR22	RR.1	
		CPR90	RR.1	
		CPR134	--	
		CPR139	RR.1	
		CPR84 <<	RR.2	
		CPR59 >>	RR.2	
		CPR88	RR.2	
6	IM	CPR112	RR.1	
		CPR113	RR.1	
		CPR129	RR.1	
		CPR127 <<	--	
		CPR138	RR.2	
		CPR19	RR.1	
		CPR81	RR.2	
		CPR132	RR.2	
		CPR152	RR.2	
		CPR152	RR.2	
7	PA-A			

Fig. 2.

CP transcripts mapped onto co-expression clusters from prior RT-qPCR analysis across developmental stages. Figure modified from Togawa et al. (2008) to provide information about CPR genes with interesting transcript levels. Genes were organized into 21 co-expression clusters from self-organizing maps (Tamayo et al., 1999; Reich et al., 2006). Boxes with gene names in orange had transcript levels (FPKM) >100, those in dark gray 10–100, those in light gray 2–9.9 and those in pink only >2 in BF3h. Gene names highlighted in yellow had uniform values of zero. Stage indicates major stages, of the 20 examined, where transcripts were found by RT-qPCR. A, adult; PA, pharate adult; P, pupa; PP, pharate pupa; L, post-ecdysial larva; PL, pharate larval instar; IM, inter-molt; L4, fourth instar larva. Sequence clusters are genes that code for proteins with almost identical sequences; names are based on chromosomal location (Cornman et al., 2008).

Fig. 3.

In situ hybridization of paraffin embedded sections with probes for different CYPs.

Hybridization to the outer surface of the eye lens for *CPLCX2*, *CPR26*, *CPR59* and *CPR75* is an artifact. No cells are present and hybridization remains even on sections pre-treated with RNase. Hybridization to pupal cuticle in *CPLCG1* is also an artifact. Note that *CPR75* and *CPR59* probes recognize different cells than the primary pigment cells seen with *CPLCX2*, *CPR26*, *CPR32* and *CPR10*. All the mosquitoes were collected 24 h after pupation, except for *CPLCG5* (young adult) and *CPR10* ovary (5-day-old adult).

Fig. 4. Transcript abundance for each functional class. Transcripts related to several classes were identified as diverse, whereas transcripts with no similarity with known genes were identified as unknown if a BLAST search did not clarify their function.

TABLE 1
Cuticular proteins with transcript levels (FPKM) at least 2 in either NBF or BF3h samples compared to data from other studies.

Gene name	CP FAMILY (group)	NBF	BF3h	Rank Order Illumina NBF	Rank Order Pitts NBF	Rank Order Bonizzoni Resistant NBF	Rank Order Marinotti Microarray NBF	DIURNAL from BIOLOCK NBF
<i>RpS7</i>		2803	2176					no
<i>CPLCG1</i>	CPLCG	826	694	1	3	1	1	head, body
<i>CPR132</i>	CPR (RR-2)	118	134	3	6	3	56	
<i>CPR10</i>	CPR (RR-2)	130	129	2	1	8	13	head
<i>CPR75</i>	CPR (RR-1)	106	94	4	7	2	2	body
<i>CPAP3-A1b</i>	CPAP3	64	18	5	16	13	24	head, body
<i>CPR16</i>	CPR (RR-1)	61	54	6	2	7	11	head, body
<i>CPR26</i>	CPR (RR-1)	18	60	17	5	5	4	head
<i>CPLCX2</i>	CPLCX	58	16	7	14	11	10	head
<i>CPR76</i>	CPR (RR-1)	47	37	8	4	4	7	no
<i>CPR113</i>	CPR (RR-1)	42	39	9	11	14	34	head
<i>CPR+C1</i>	CPR+C	39	26	14	8	30	40	head, body
<i>CPR81</i>	CPR (RR-1)	28	38	10	12	10	5	head
<i>CPR131</i>	CPR (RR-2)	29	33	12	68	38	65	
<i>CPR15</i>	CPR (RR-1)	31	32	11	10	6	12	body
<i>CPR9</i>	CPR (RR-1)	29	14	13		25	36	no
<i>CPR21</i>	CPR (RR-1)	21	27	16	13	20	42	body
<i>CPAP3-E</i>	CPAP3	23	14	15	15	21	32	head, body
<i>CPAP1-C</i>	CPAP1	10	8	18	18	23	46	no
<i>CPAP3-D</i>	CPAP3	10	8	19	22	24	28	head
<i>CPR70</i>	CPR (RR-2)	8	9	20	47	42	26	no
<i>CPR30</i>	CPR (RR-1)	4	7	26	9	22	47	head, body
<i>CPR127</i>	CPR (RR-not classified)	6	2	21		19	19	no
<i>CPAP3-B</i>	CPAP3	5	6	22	32	29	29	head
<i>CPLCG21</i>	CPLCG	4	6	27	33	52	85	
<i>CPR79</i>	CPR (RR-1)	5	4	24	25	34	30	head

Gene name	CP FAMILY (group)	NBF	BF3h	Rank Order Illumina NBF	Rank Order Pitts NBF	Rank Order Bonizzoni Resistant NBF	Rank Order Marinotti Microarray NBF	DIURNAL from BIOLOCK NBF
<i>CPR59</i>	CPR (RR-2)	2	5	34	59	17	6	head
<i>CPLCG5</i>	CPLCG	2	5	35	33	9	101	
<i>TWDL12</i>	TWDL	5	5	23	20	69	83	no
<i>CPF3</i>	CPF	4	2	25	28	80	27	no
<i>CPR146</i>	CPR (RR-2)	4	3	28	24	32	53	no
<i>TWDL2</i>	TWDL	4	3	29		66	55	
<i>CPR73</i>	CPR (RR-1)	3	4	30	70	57	109	no
<i>CPR25</i>	CPR RR-1)	1	4					body
<i>~CPR157</i>	CPR (RR-not classified)	3	3	31				no
<i>CPR130</i>	CPR (RR-not classified)	1	3					no
<i>CPR74</i>	CPR (RR-1)	0.5	3					no
<i>CPR124</i>	CPR (RR-2)	1	2					no
<i>CPAP3-Alt</i>	CPAP3	1	2					head
<i>CPR138</i>	CPR (RR-not classified)	2	2	33	31	53	75	no
<i>CPR134</i>	CPR (RR-not classified)	2	2	32	21	50	96	no
<i>CPR116</i>	CPR (RR-2)	1	2					no

Data expressed as FPKM is arranged by maximum level in highest condition. Rps7 given for comparison. The group is given for CPR family members. Bolded gene names and numbers were significantly different between conditions. Rank orders for NBF from our data are compared to Illumina data from Pitts et al. (2011) and Bonizzoni et al. (2012) and to the microarray data of Marinotti et al. (2006). Those that did not fall within the top 35 are shown in italics (and red in Web version). CPs with diurnal rhythms based on Rund et al. (2013) are indicated. Transcripts with no or low data have no entry. Gene names shaded in gray indicate those for which *in situ* hybridization is reported.

TABLE 2

CPs with significantly different transcript levels in BF3h vs. NBF compared to Marinotti et al. (2006) and Rinker et al. (2013).

CP name	Our RNA-Seq analysis			Marinotti et al.	Rinker et al. antennae		
	NBF FPKM	BF3h FPKM	Fold-change BF3h/NBF		Fold- change 1 h	Fold- change 12 h	Diurnal rhythm?
CPR74	0.46	2.67	5.80	10.57	No data	No data	
CPR25	1.02	4.16	4.08	36.18	1.6	0.95	yes
CPR26	17.98	59.85	3.33	.0004	0.95	0.95	yes
CPR130	0.96	3.05	3.18	0.08	1.13	4.57	yes
CPLCG5	1.63	4.74	2.91	0.24	0.1	7.61	no
CPR59	1.8	4.93	2.74	0.52	1.99	1.22	yes
CPR124	0.89	2.37	2.66	7.88	1.18	2.27	no
CPAP3-E	23.13	14.29	0.62	0.63	0.99	0.59	no
CPR9	28.54	13.6	0.48	0.58	0.47	0.56	no
CPR127	6.48	2.12	0.33	0.57	0.75	1.01	no
CPAP3-A1b	63.67	18.42	0.29	.82	0.46	0.35	yes
CPLCX2	58.31	16.38	.28	.38	0.49	0.74	no

Data from Marinotti et al. that differ in direction of change are in italics (and red in Web version). Data from Rinker et al. that agree with direction of change are in bold (and green in Web version). Criteria used for diurnal rhythms are given in Sect. 2.7.

TABLE 3

Comparison of significant BF3h/NBF ratios of CP transcript levels from samples collected at the same time and analyzed by RNA-seq and RT-qPCR.

Transcript from Gene	RNA-seq	RT-qPCR
CPR74	5.85	5.12
CPR26	3.33	5.46
CPR130	3.18	2.95
CPLCG5	2.90	3.45
CPR59	2.74	2.32
CPR124	2.64	3.66
CPR9	0.48	0.65
CPR127	0.33	0.56
CPAP3-A1b	0.29	0.35
CPLCX2	0.28	0.25

TABLE 4

Spatial, temporal, and comparative data on transcripts that were most affected by blood feeding.

gene	location	>10-FOLD INCREASE BF3h/NBF			gene	location	>10-FOLD DECREASE BF3h/NBF		
		our data	Marinotti	BIOCLOCK			our data	Marinotti	BIOCLOCK
AGAP013705	no data	>1100		no data					
AGAP004350	MT	176.52	14.78	low	AGAP010781	MG	0.02	0.07	yes 32
AGAP006187	MG	152.85	1.34	low	AGAP011939	MG	0.02	0.01	yes 24
AGAP004203	CAR	103.81	0.35	low	AGAP004133	no data	0.02	0.76	no
AGAP011808	MG	79.85	0.26	low	AGAP012394	no data	0.03	0.22	yes 12
AGAP005712	CAR	46.87	2.78	no	AGAP009591	CAR	0.04	7.68	yes 24
AGAP011047	MT	40.38	9.52	low	AGAP012395	no data	0.04	0.22	yes 12
AGAP008628	MT	37.38	1.08	low	AGAP009122	BROAD	0.05	2.36	no
AGAP012247	MG	33.95	1.57	yes 24	AGAP003623	MG MT O	0.05	6.76	yes 32
AGAP010243	MG	29.71	0.59	low	AGAP008688	CAR	0.05	0.26	yes 32
AGAP007165	MG	27.32	0.4	yes 24	AGAP004809	MG	0.05	1.97	yes 32
AGAP003205	MT	22.62	5.11	low	AGAP006709	MG	0.05	0.00	yes 24
AGAP008900	MT	20.38	1.48	low	AGAP012296	MT	0.06	5.94	yes 24
AGAP013063	MT	18.86	0.03	yes odd	AGAP006047	CAR	0.06	1.65	low
AGAP003206	MG	15.58	1.44	no	AGAP006736	MT	0.06	0.47	yes 24
AGAP003714	CAR	14.53	0.15	yes 32–36	AGAP012757	no data	0.07	0.99	yes 28
AGAP008363	CAR	14.43	0.58	yes 20	AGAP003692	MG	0.07	0.99	yes 28
AGAP006707	MG	13.7	0.01	yes 16	AGAP008547	no data	0.07	no data	no data
AGAP005065	MG	12.73	0.16	yes 16–20	AGAP006731	no data	0.07	6.38	low
AGAP008849	BROAD	12.39	0.35	no	AGAP008218	MT	0.08	0.87	yes 32
AGAP002850	MG	12.35	155.05	low	AGAP003020	MG	0.08	1.54	no
AGAP010854	MG, MT	11.24	1.7	low	AGAP004383	BROAD	0.08	22.28	no
AGAP008254	SG	11.1	0.57	no data	AGAP004900	MG	0.09	2.09	yes 24
AGAP010132	no data	11.09	39.7	no	AGAP013155	MG	0.09	72.15	yes 24–28
AGAP010118	no data	10.95	12.97	no	AGAP001881	MG	0.09	1.50	yes 24–32
AGAP000603	MG, MT	10.94	0.004	yes 16	AGAP008370	MG	0.10	1.11	yes 28–32

gene	location	>10-FOLD INCREASE BF3h/NBF			gene	location	>10-FOLD DECREASE BF3h/NBF		
		our data	Marinotti	BIOCLOCK			our data	Marinotti	BIOCLOCK
AGAP008632	MT	10.9	2.67	yes	AGAP010571	MG	0.10	0.54	low
AGAP000558	SG	10.82	1.54	no	AGAP007676	MG	0.10	11.64	low
AGAP009385	CAR, MG	10.72	0.51	low					
AGAP012849	no data	10.4	39.7	no					
AGAP005945	CAR	10.18	0.14	yes 24					
AGAP002587	MT	10.15	1.92	yes 12					
AGAP000732	MT	10.1	0.57	yes 12					

Comparison of our data on all transcripts with 10-fold changes to data from Marinotti et al. (2006). BF3h/NBF ratios that are bolded had lowest value of 2 or greater. Numbers italicized (and red in Web version) indicated that change in transcript levels was in the opposite direction in data from Marinotti et al. (2006). Transcript locations are from Baker et al. (2001). Abbreviations: MT (Malpighian tubules); MG (midgut); CAR (carcass); O (ovary); BROAD (no obvious localization). BIOCLOCK data are from bioclock.crc.nd.edu; based on LD body; maximum values below 20 recorded as low; #s indicate zeitgeber time of peak values.