UNIVERSITY OF CALIFORNIA SAN DIEGO

A Comparative Transcriptomic Analysis of the Mosquito Blood Feeding Response

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Bryant Cao

Committee in charge:

Professor Matthew Daugherty, Chair
Professor Omar Akbari
Professor Amy Pasquinelli

2021

The thesis of Bryant Cao is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

EPIGRAPH

Stay Hungry. Stay Foolish. And I have always wished that for myself. And now, as you graduate to begin anew, I wish that for you.

Steve Jobs

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge my committee chair, Professor Matthew Daugherty, who has encouraged, supported, and inspired me throughout my academic career. His ability to recognize one's strengths and enthusiasm for mentoring has helped me grow into not only a researcher but also as a better person. Under his guidance, I have been able to pursue a path that I never would have thought was possible, and I will be forever grateful that he offered me the opportunity that started my journey.

I would next like to acknowledge Brian Tsu, my graduate student mentor throughout my time in the Daugherty lab. Brian has always provided everything that I needed to fill the shoes of a scientist, from experimental analyses to the answer to every question I ever had. As I was willing to learn, Brian was willing to teach, and that dynamic has made these past three years more valuable than anything I have ever learned in college.

I would also like to acknowledge Stephanie Gamez, Michelle Bui, Jacob Garrigues, and Professor J.J. Emerson for their assistance in this project. Without their experiments and insights, my current work would not have been possible. It has been an experience for me to collaborate on something entirely new, but more than that, an absolute pleasure to learn from these brilliant individuals along the way.

I would like to thank all current and former members of the Daugherty lab, for their support, friendship, and making every day of coming to lab something to look forward to.

Lastly, I would like to acknowledge my thesis committee members, Professor Omar Akbari and Professor Amy Pasquinelli, for their insights, patience, and consideration of my candidacy.

This thesis is co-authored by Bryant Cao, Stephanie Gamez, Michelle Bui, Brian Tsu, Omar Akbari, and Matthew Daugherty. The thesis author was the primary author of this paper.

ABSTRACT OF THE THESIS

A Comparative Transcriptomic Analysis of the Mosquito Blood Feeding Response

by

Bryant Cao

Master of Science in Biology

University of California San Diego, 2021

Professor Matthew Daugherty, Chair

Female mosquitoes are only able to produce eggs upon feeding on blood from their vertebrate hosts, but this feeding behavior activates a network of genes to ensure their survival following a blood meal. Blood meals are critical to mosquito proliferation despite activating stress and defense responses, however the whole transcriptome at the early time point post-blood meal has not been fully studied across mosquito genera. To explore the transcriptomic responses under environmental thermal stress and early-stage blood feeding, we compared bulk RNA-seq profiles

across *Ae. aegypti*, *Ae. albopictus*, *An. gambiae*, *An. stephensi*, and *Cx. quinquefasciatus* under these experimental conditions. In addition, we conducted a functional and orthology inference analysis to compare differentially expressed genes across species and condition. We found that despite the conservation of the heat shock response, there are deviations in the response across species based on heat shock inducibility. We were also able to identify that a blood meal elicits a heat shock response in *Ae. albopictus* and that intersections with the blood feeding response in *Ae. albopictus* highlights the temporal dynamics to survival thermal stress of a blood meal before the activation of metabolic processes. Our comparative approach and generated RNA-seq datasets illustrates the convergence and divergence in these critical stress responses of mosquitoes that would be beneficial for understanding mosquito survival.

INTRODUCTION

1.1 Mosquito-Borne Vectors of Infectious Diseases

Female mosquitoes require blood meals from vertebrate host organisms for the maturation of their eggs (Klowden & Briegel, 1994). However, this blood feeding behavior is the cause for transmission of mosquito-borne human diseases, some of which include malaria, dengue virus, and Zika virus (Omodior et al., 2018). Combined with the impacts from climate change, urbanization, and international mobility which can further the spread of mosquitoes and their disease burdens, mosquitoes are increasing threats to global health through their blood feeding behaviors (Kraemer et al., 2019; Wilder-Smith et al., 2017). As such, approaches that can identify molecular targets at this critical point of the mosquito's gonotrophic cycle during a blood meal could assist in vector control.

1.2 Transcriptomic Responses During Mosquito Hematophagy

Following a blood meal, a mosquito undergoes genome-wide transcriptional changes, most notably in functions for metabolism to provide nourishment for itself and its eggs (Attardo et al., 2005; Santiago et al., 2017). However, other biological processes important for the mosquito's survival are also involved in the blood feeding response and prior work has highlighted that levels of transcriptional products for defense mechanisms, catalytic activity, and molecule binding also undergo significant changes upon a blood meal (Bryant & Michel, 2014; Dana et al., 2005; Hou et al., 2015). These transcriptomic profiles following hematophagy have been recorded over the course of a mosquito's development or of a limited selection of genes from multiple mosquito species, but little work has been done to explore these early-stage, post-blood meal (PBM) transcriptomic profiles in a whole-genome, comparative approach. Conducting a comparative

analysis at a time point earlier than what has been previously explored may elucidate mechanisms and molecular targets of blood feeding applicable to multiple species that can be further studied.

A particular response that has been studied to be active immediately following a blood meal is the heat shock response, a cellular stress response intended to mitigate proteotoxic effects that can compromise cellular functions due to increasing temperatures (Shibata & Morimoto, 2014). While the heat shock response is active in situations of high environmental temperatures, prior work has explored that blood meals alone in *Ae. aegypti*, *An. gambiae*, and *Cx. pipiens* elicit these heat shock responses and that suppression of heat shock protein 70 (Hsp70) from the canonical response can impair egg production (Benoit et al., 2011). Although Hsp70 is the most well-studied gene in the heat shock response and has been focused on in prior work, there are likely other genes that play a role in this regulatory network that can be detected under blood feeding conditions with a whole-genome RNA-seq analysis.

1.3 A Potential Expansion to Mosquito Stress Responses via Transposition

The need for a blood meal is a necessary yet potentially deadly component for mosquito proliferation – a remarkable case in evolution to need blood while also mitigating all its consequences upon feeding. Mosquitoes are not the only organisms that partake in hematophagy, but their small bodies make thermal stress more prominent (Freitas & Nery, 2020). Because mosquitoes can transmit diseases through feeding, every blood meal is a potential exposure to foreign pathogens (Pakpour et al., 2014). Observing the transcriptomics of the innate immune response and other defense responses has been done before in mosquitoes, but little has been done to specifically explore the evolutionary impacts from these combined stressors of thermal stress and foreign pathogens.

Transposable elements (TEs) are mobile genetic elements that can move from one genomic locus to another (Melo & Wallau, 2020). Because of the repetitive pasting mechanisms of TEs, DNA transposons are a component of almost all eukaryotic genomes and their pasting of repetitive elements has been known to alter gene regulatory networks over the course of evolutionary history (Cosby et al., 2021; Feschotte, 2008). Prior work has observed that TEs are prevalent in most mosquito genomes, specifically with an abnormally high proportion of TE calls in *Ae. albopictus*, *Ae. aegypti*, and *Cx. quinquefasciatus* (Melo & Wallau, 2020). The activity of rampant TEs in mosquito genomes to alter gene regulatory networks is a potential cause to the expansion of critical transcriptomic responses through co-opting unrelated genes into a regulatory network. This evolutionary phenomenon was illustrated with the heat shock response in *Caenorhabditis* through a particular subclass of TE, Helitrons, that could copy and repetitively paste a transcription factor binding site upstream of a gene in order to allow a transcription factor to bind and upregulate the gene's function under thermal stress (Garrigues et al., 2019). As it has been studied that TEs have a propensity to mobilize under cellular stress, the evolutionary pressures exerted by these stressors may have also caused TEs to mobilize in mosquitoes and expand regulatory networks like the heat shock response (Horváth et al., 2017). It has not been observed if similar to *Caenorhabditis*, there are cases that the canonical transcription factor binding site of the heat shock response, heat shock elements (HSEs), have been carried by Helitrons to their current genomic locations where they can affect heat shock inducibility from either environmental thermal stress or a blood meal (Vihervaara et al., 2018).

1.4 Comparative Transcriptomic Profiles from Blood Feeding and Heat Shock Responses

Many of the genes and their functions from the early-stage blood feeding response and the heat shock response have not been comprehensively highlighted before in a single mosquito genome, so generating these profiles would assist in revealing previously undetected transcriptional products and interactions. In addition, little work has been done to compare these profiles across mosquito species, as blood meals and thermal stress are common exposures in all mosquitoes, but it has not been explored if these critical responses have conserved or divergently evolved functions across species to adapt to their stressors. Comparing mosquito transcriptomic profiles of these two critical responses from an intra- and inter-species perspective can greatly augment our understanding of two situations in a mosquito's lifespan.

While these transcriptomic profiles quantify transcript accumulation, they also need to be mapped to biological information in order to generate informative insights about relevant genes, dynamics, and potential molecular targets. However, retrieving this information is limited to the quality of existing genome annotations which is expected to be poorly annotated as it would need significant amounts of experimental evidence to identify unknown genes. As recorded experimental evidence of gene functions is limited and large portions of mosquito genomes are annotated with unknown functions, we conducted homology-based functional analyses using Gene Ontology (GO) terms (Ashburner et al., 2000; Mi et al., 2019; The Gene Ontology Consortium et al., 2021) and the Clusters of Orthologous Groups of proteins (COG) categories (Galperin et al., 2021; Tatusov et al., 1997) to infer gene functions based on the homologous sequences from other genomes.

In this study, we compared transcriptomic profiles across a sampling of five mosquito species: *Ae. aegypti*, *Ae. albopictus*, *An. gambiae*, *An. stephensi*, and *Cx. quinquefasciatus*. The original locations of these five mosquito species span across various habitat ranges (Figure 1A)

and genera (Figure 1B) which we expect can capture a diverse early-stage blood feeding and heat shock response across orthologs (Lippi et al., 2019; Pock Tsy et al., 2003; Samy et al., 2016; Tikar et al., 2011; Vega-Rúa et al., 2020). Intersecting differentially expressed genes under the early-stage blood feeding and heat shock responses can allow us to capture the early-stage blood feeding response that all female mosquitoes encounter but has not been thoroughly studied before, confirm previous findings about the overlap between blood meals and the heat shock response, and capture convergently evolved gene functions under these stressors.

METHODS AND MATERIALS

2.1 Mosquito Rearing and Sample Collections

Ae. albopictus (LA county strain), *Ae. aegypti* (Liverpool strain), *An. gambiae* (G3 strain), *An. stephensi* (Indian strain), and *Cx. quinquefasciatus* (Auburn strain) were maintained under standard insectary conditions at a constant 28°C at 70-80% humidity and with a 12:12 light/dark cycle. Larvae from each species were reared on dry fish food (Tetramin fish flakes) and adults were reared on a 10% sucrose solution. Mosquitoes were aged 3-4 days before collection.

Five adult female mosquitoes per replicate were collected into 1.5 mL tubes and three replicates were prepared for each species. Control samples were stored in a -80°C freezer following collection. Blood fed samples were fed on anesthetized mouse blood for 5-10 minutes and then stored in a -80°C freezer following feeding. Environmentally heat shocked samples were placed in a cage in an incubator set to 37°C for one hour and then stored in a -80°C freezer following treatment.

2.2 RNA Extractions and Sequencing

Each sample tube of mosquitoes was lysed with lysis buffer and crushed. Total RNA was extracted from each of the lysates with the RNeasy Mini kit (Qiagen). Following RNA extraction, samples were subjected to a DNase treatment using the Turbo DNA-free kit (Ambion/Thermo Fisher). RNA samples were stored in a -80°C freezer following treatment until prepared for bulk RNA sequencing.

RNA sequencing libraries were prepared with an Illumina TruSeq Stranded mRNA kit. Single-read RNA-seq was performed on *Ae. albopictus* samples with an Illumina HiSeq 4000

platform and paired-end RNA-seq was performed on *Ae. aegypti*, *An. gambiae*, *An. stephensi*, and *Cx. quinquefasciatus* samples with an Illumina NovaSeq 6000 platform.

2.3 Transcript Quantification and Differential Expression Analysis

Raw sequencing read content, read quality, GC content, and sequences with ambiguous base calls were assessed with FastQC. Raw read counts, the percentage of unique reads, and the percentage of reads with a Phred quality score above 30 (Q30) were calculated (Table 1). Sequencing reads of each species were quantified with Salmon in a quasi-mapping-based mode to their corresponding reference genomes of *Ae. albopictus (*RefSeq accession: GCF_006496715.1), *Ae. aegypti* (RefSeq accession: GCF_002204515.2), *An. gambiae* (VectorBase structural annotation version: AgamP4.13), *An. stephensi* (NCBI BioProject accession: PRJNA629843) (Chakraborty et al., 2021), and *Cx. quinquefasciatus* (VectorBase structural annotation version: CpipJ2.5). Salmon was run with parameters to enable the mapping validation algorithm and GC bias correction for each sequencing sample; the single-read flag was passed for single-read samples and the paired-end flag was passed for paired-end samples (Patro et al., 2017). Read quantifications were imported and differentially expressed genes across experimental conditions were identified using the R package DESeq2 (Love et al., 2014).

2.4 Orthology Inference and Functional Annotation

Ortholog groups were detected from reference genome protein sequences using OrthoFinder with default parameters (Emms & Kelly, 2015, 2019). The gene with the lowest Benjamini-Hochberg adjusted p-value of an ortholog group was selected as the species' representative ortholog, otherwise the gene with the lowest non-adjusted p-value was selected as

the species' representative ortholog. To generate paralog-weighted ortholog groups, each ortholog group was multiplied by the number of paralogs from the group. Individual sets and intersections of ortholog groups were visualized using the Python package UpSetPlot to draw UpSet plots (Lex et al., 2014).

GO term annotations were determined from the reference genome protein sequences using OmicsBox/Blast2GO (Götz et al., 2008). The OmicsBox/Blast2GO annotation process consisted of a blastp-fast search against the NCBI RefSeq protein database with an E-value cutoff of 10^{-3} and a maximum number of allowed hits of 50, a mapping step of BLAST hits to GO terms, and an InterPro scan with default parameters against the EMBL-EBI InterPro protein database for GO terms (Altschul et al., 1990; Blum et al., 2021). Due to limitations in existing experimental evidence to confirm GO terms in sampled mosquito species, the "GO annotation" step of the pipeline to validate GO terms was not performed in order to preserve annotations. COG category annotations were determined from the reference genome protein sequences using eggNOG-mapper against the EMBL eggNOG database with default parameters (Huerta-Cepas et al., 2019).

A GO and COG enrichment analysis was conducted using a binomial test between the observed term datasets to the whole-genome annotations. P-values were transformed on a $-\log_{10}$ scale and standardized with the scale function in R, which subtracts each log-transformed value by the mean of all log-transformed values and divides by the standard deviation of all log-transformed values.

2.5 HSE Motif and Helitron Element Identification

The 15-bp HSE motif [ATCG]GAA[ATCG][ATCG]TTC[ATCG][ATCG]GAA[ATCG] was searched for in genomes via regular expression. RepeatMasker with a custom-built library of

species-specific Helitron sequences from VectorBase was used to detect Helitron transposable elements from genomes. Internal scripts were used to integrate the coordinates of HSEs and Helitron calls to the nearest downstream gene in the genome.

RESULTS

3.1 Assessing RNA-seq Quality and Functional Annotations of Transcriptomes

Under standard rearing (CTL), early-stage blood fed (BF), and environmental heat shock (HS) conditions, we generated 11 transcriptomic profiles across *Ae. aegypti*, *Ae. albopictus*, *An. gambiae*, *An. stephensi*, and *Cx. quinquefasciatus* from 33 samples with each sample averaging 36617825 reads (Table 1). The average Phred quality score of each sample was above 35, indicating at least a 99.96% accuracy in sequencing base calls (Ewing et al., 1998; Ewing & Green, 1998).

Using the reference genome sequences, we were able to functionally infer and annotate genes as many of these sequences in genomes are poorly annotated and have unknown functions. From our analysis, we were able to annotate most of the whole genome with either a GO term or a COG category annotation (Table 2). These annotations were then used to infer the functions of our observed transcripts and applied in downstream enrichment analyses.

3.2 Assessing Differential Gene Expression Profiles Under Experimental Conditions

We were able to confirm that under HS and BF conditions, the transcriptomic profiles across conditions were significantly different from each other based on hierarchical clustering and a principal component analysis of samples (Figure 2). To identify differentially expressed genes (DEGs) in the BF and HS samples, we compared the gene expression profiles between the control and each of the experimental conditions. We used a threshold cutoff of a $\log_2$ fold change > 1 and a Benjamini-Hochberg adjusted p-value < 0.05 to identify significant differences in gene expression over the control. Based on these parameters, we were able to identify DEGs between

the HS-CTL and BF-CTL conditions from *Ae. albopictus*, *Ae. aegypti*, *An. gambiae*, *An. stephensi*, and *Cx. quinquefasciatus* (Table 3).

From the proportions of upregulated DEGs under HS conditions in the transcriptome, *Ae. albopictus* and *Cx. quinquefasciatus* exhibit the highest levels of heat shock inducibility of the sampled species, with *Ae. albopictus* having over five times the number of heat shock upregulated DEGs over *Ae. aegypti*. Within the BF samples of *Ae. albopictus*, there are almost four times the number of downregulated DEGs to the number of upregulated DEGs.

3.3 A Transcriptomic Response to Thermal Stress is Present Following *Ae. albopictus* Blood Meals

From the 9857 unique GO term annotations retrieved across all five genomes, we were able to conduct an enrichment analysis to test for genes that might be highly enriched within a GO term functional annotation under each experimental condition. The 25 unique GO terms with the highest standardized and log-scaled p-values from any genome were visualized and compared to each other (Figure 3). Across HS upregulated DEGs, GO terms for "response to heat", "protein folding chaperone", and "protein refolding" are among the most enriched GO terms and are characteristic of the heat shock response (Gidalevitz et al., 2011; Gomez-Pastor et al., 2018; Shibata & Morimoto, 2014; Vihervaara & Sistonen, 2014). With our model for comparing enriched GO terms, we also conducted an analysis from the HS downregulated DEGs and while there are fewer detailed insights into associated biological systems, there are observable molecular functions that are visible with HS downregulation, including "oxidoreductase activity", "carboxylic ester hydrolase activity", and "serine-type peptidase activity".

For a broader understanding of the relevant biological functions, we repeated the enrichment analysis using 24 of the general COG categories (Figure 4). Across the HS upregulated

DEGs, “posttranslational modification, protein turnover, and chaperones” is a consistently enriched COG category, likely expected of the functions of activated genes from the heat shock response including chaperoning functions to mitigate protein misfolding and aggregation (Gidalevitz et al., 2011; Gomez-Pastor et al., 2018; Shibata & Morimoto, 2014; Vihervaara & Sistonen, 2014). Across the HS downregulated DEGs, there are a multitude of enriched categories in addition to certain categories being highly enriched in only one species, including “cell wall/membrane/envelope biogenesis” in *Ae. aegypti*, “defense mechanisms” and “secondary metabolites biosynthesis, transport and catabolism” in *An. gambiae*, or “carbohydrate transport and metabolism” in *An. stephensi*.

Expectedly, the *Ae. albopictus* BF samples display a different set of GO term and COG category enrichments from the HS samples. However, at the early stages PBM, we do observe a moderate enrichment of the associated thermoregulatory GO terms (“response to heat”, “protein folding chaperone”, and “protein refolding”) and COG category (“posttranslational modification, protein turnover, and chaperones”) highlighted in HS upregulated DEGs. These annotations are also consistent with the gene expression profile at the transcript level in which one of the numerous Hsp70 transcripts in the *Ae. albopictus* transcriptome, LOC109423613, exhibits a 2048-fold change in expression under HS conditions over the control, but exhibits a 29-fold change in expression following a typical blood meal.

3.4 Intersections in the Heat Shock and Early-Stage Blood Feeding Response in *Ae. albopictus*

In addition to comparing the experimental conditions to the control, we also conducted an analysis to compare the HS and BF conditions from *Ae. albopictus* to each other. This would help to highlight genes that were differentially expressed in both conditions, suggesting that a blood

feeding response can resemble a heat shock response or are exclusively characteristic of one condition. We used a threshold of an absolute $\log_2$ fold change > 0 and a Benjamini-Hochberg adjusted p-value < 0.05 under both experimental conditions to determine significantly expressed genes, which resulted in a total of 316 significantly expressed genes from the background of 33823 total genes (Figure 5A). We used a cutoff of $\log_2$ fold change > 0 or $\log_2$ fold change < 0 for a high or low expression of a condition, respectively, to highlight significantly expressed genes into quadrants (Figure 5B). Within each quadrant, we performed both a GO term and COG category enrichment analysis to determine highly enriched gene functions under each pair of experimental conditions (Figure 5C).

In Quadrant 1 of high HS and high BF, we identified significantly enriched gene functions in which the two conditions were correlated to each other. The most enriched functions of this quadrant were related to thermal stress and thermoregulation, emphasizing the combined impacts of blood feeding in a heat shock response in *Ae. albopictus*. In addition, transcripts corresponding to other genes outside of thermoregulation are also significantly expressed in this quadrant, which include LOC109414457 (putative ammonium transporter 2), LOC109412094 (beta-galactosidase-1-like protein), and LOC109403414 (phenoloxidase-activating factor 2-like). While the first two genes correspond to metabolic functions and may suggest their role in early-stage metabolism of a blood meal that can be impaired under sustained thermal stress (Giri et al., 2017; Nguyen et al., 1989), the third gene corresponds to a major component of the innate immune system in *Ae. albopictus* which is potentially linked to a thermal component from geographically variable environments and seasons (Fedorka et al., 2013; González-Santoyo & Córdoba-Aguilar, 2012).

In quadrant 2 of low HS and high BF, we identified genes that were anticorrelated to each experimental condition. With a small sample of genes and an inconclusive functional analysis, it

is likely that these transcripts, such as those corresponding to LOC115264160 and LOC115264698, are relevant to digestion and metabolism of a blood meal but are not linked to thermoregulatory functions.

In quadrant 3 of low HS and low BF, we identified genes and functions correlated to both experimental conditions but are not upregulated in either condition. Highly abundant transcripts correspond to downregulated genes of LOC109409431 (neutral and basic amino acid transport protein rBAT-like, transcript variant X1), LOC109413895 (heterogeneous nuclear ribonucleoprotein F-like), and LOC109426418 (probable cytochrome P450 9f2). While these descriptions suggest genes that may be relevant in metabolism and immune functions, it may be possible that these isoforms are not involved under these stress situations. Of additional note is that the one of the most downregulated transcripts between HS and BF conditions corresponds to LOC109422095 (uncharacterized protein K02A2.6-like) which has been noted to be *C. elegans* transposon-related in origin but is significantly downregulated in the *Ae. albopictus* transcriptome (Metzger et al., 2018).

In Quadrant 4 of high HS and low BF, we identified the largest intersection of genes of our analysis and that were anticorrelated under HS and BF conditions. We hypothesize that gene functions of these quadrant could be strictly for non-BF thermal stress responses, such as transcripts corresponding to LOC109420715 (larval cuticle protein LCP-30-like) and LOC109397727 (spondin-1-like) or could function in later stages for metabolism with transcripts corresponding to LOC109406109 (facilitated trehalose transporter Tret1, transcript variant X1) or LOC109422852 (serine protease easter). While there are genes with descriptions that would suggest being upregulated immediately following a blood meal, these functions may be relevant

in later stages, consistent with the changes in transcriptome dynamics for certain metabolic, stress, and immune response categories in the later stages PBM in *An. gambiae* (Dana et al., 2005).

3.5 Orthologous Gene Groups Following the Heat Shock Response

We compared the transcriptomic profiles across the five mosquito species under HS conditions to determine the intersection of significantly expressed genes following a heat shock. From our orthology inference analysis, we generated a total of 14890 unique ortholog groups, accounted for the associative weight from paralogs in these groups, and determined 31 sets of ortholog groups unique to each species or in combination of species (Figure 6). While we expect the intersection total to be larger within a genus (i.e., between *Ae. aegypti* and *Ae. albopictus* or between *An. gambiae* and *An. stephensi*) compared to other inter-genus intersections of similar group size, there are certain groups not within a genus that are larger even from both upregulated and downregulated genes, namely with *Ae. albopictus* and *Cx. quinquefasciatus*. These groups combined with the intersections from larger group sizes, may help to highlight HS-relevant genes across species and divergently evolved genes in mosquito genomes under thermal stress.

3.6 The Accumulation of Helitron TEs and HSEs in Mosquito Genomes

From our analysis, we were able to call Helitron TEs and HSEs from each mosquito genome (Table 4). Consistent with previous findings, we also observe that Helitron content is much higher in *Ae. albopictus*, *Ae. aegypti*, and *Cx. quinquefasciatus* than in *An. gambiae* or *An. stephensi* (Melo & Wallau, 2020). Surprisingly, we also observe a substantially higher amount of HSEs in the genomes with higher TE content, most notably with *Ae. albopictus* having nearly ten times the proportion of HSEs in the genome and having a peak HSE count in a 1-kb window 22

times higher than in *An. gambiae* or *An. stephensi*. It may be possible that this contrast in TE and HSE content is genus-specific but would require further analyses from other species.

DISCUSSION

4.1 Insights from the Deviations in Mosquito Transcriptional Responses

The heat shock response has been well studied to be a highly conserved stress response and we confirm that we do observe similar enrichments of functional descriptions and the same canonical genes in the response being upregulated across our sampling of mosquito species under these sustained, environmental heat shock conditions (Vihervaara & Sistonen, 2014). However, this analysis also highlights that despite this response being conserved within each species, there are variations in the response across species. The contrast of upregulated DEGs in *Ae. albopictus* or *Cx. quinquefasciatus* and their substantially larger proportion of overall DEGs to other mosquito species highlights the deviation in heat shock inducibility in these species. Furthermore, it is of note that *Ae. albopictus* and *Cx. quinquefasciatus* share a larger set of ortholog groups compared to other paired comparisons, suggestive of their interconnected heat shock inducible gene groups. In addition to transcripts accumulating for genes corresponding to the heat shock proteins like Hsp70, Hsp60, and Hsp83, there are many other undetermined, upregulated DEGs outside of thermoregulatory functions in *Ae. albopictus* or *Cx. quinquefasciatus*. While these findings are likely strain-dependent, the increased levels of heat shock inducibility may be suggestive of a higher tolerance to thermal stress, which could be potentially applied into real-world environmental conditions linked to climate change and these species' impact on habitat range expansions.

The contrast of heat shock inducibility across mosquito species could be explained by the activity from Helitrons and other TEs. As it is possible that certain subclasses of TEs like Helitrons can "copy-and-paste" genomic segments rather than "cut-and-paste", the correlation between TE content, HSE content, and heat shock inducibility to upregulated DEGs in a mosquito genome remains an unexplored question to answer (Cosby et al., 2019). We had conducted preliminary

analyses similar to the methods in Garrigues et al. (2019) with this sampling of mosquito species under HS conditions to explore if clusters of Helitron-associated HSEs (HSEs that had been captured by formerly-autonomous Helitrons and pasted to their current genomic locations) had an effect on DEGs, and while we had identified 23 transcriptional products to fall under our criteria in *Ae. albopictus*, there were too few products in other species to determine any conclusive results. While we were searching for Helitron-associated HSEs that could influence the gene immediately downstream of the HSE cluster, there is the possibility that these clusters could contribute to the upregulation of distant genes or may not be Helitron-associated but rather associated with another subclass of TE.

Based on the enrichments of descriptions for thermoregulatory functions, the early-stage blood-feeding response mirrors the heat shock response to a substantial degree in *Ae. albopictus*. It is remarkable that a moderately strong response to thermal stress is prevalent following a mosquito blood meal in *Ae. albopictus* when the samples are from whole-body, 15-minute PBM mosquitoes and is comparable to mosquitoes under a 1-hour sustained thermal stress. While previously unknown, these findings support the prior work that mosquito blood meals elicit heat shocks in *Ae. albopictus* in addition to other mosquito species (Benoit et al., 2011). While not conclusive at this time but based on the contrast in the levels of heat shock inducibility across mosquito species, we suspect that early-stage blood feeding transcriptomic profiles outside of *Ae. albopictus* are expected to not show as considerable enrichments for thermoregulatory functions and instead may be more enriched for other undetermined functions.

4.2 Capturing *Ae. albopictus* HS- and BF-Relevant Gene Subsets

Our analysis in contrasting the HS and BF transcriptomic profiles in *Ae. albopictus* highlights the intersection as well as the divergence in these two stress responses which can help to further refine molecular targets relevant under these conditions, specifically for being strictly HS-relevant or early-stage BF-relevant. While we found similar enrichments under high HS and high BF conditions for genes relevant to thermal stress, the other pairs of conditions highlight an unexplored space in mosquito transcriptomics particularly with the intersection of high HS and low BF conditions. It is suspected that this region is for HS-relevant gene functions only, however the large quantity and significant enrichment for metabolic functions despite these genes being downregulated highlights that following mosquito blood meals, thermoregulatory, immune, and other stress mitigation responses precede metabolic processes. While it is known that the accumulation of transcriptional products following a blood meal changes over time and has been profiled into stages, it was not previously known that the thermoregulatory functions contributed to a major role in the early-stage blood feeding response (Chen et al., 2017; Dana et al., 2005; Hou et al., 2015). Further data on the later stages of blood feeding (presumably into hours or days) may reveal genes that are relevant in the initial survival of a blood meal followed by metabolic or immune processes.

While not conclusive at this time, it is likely that the other intersections of HS and BF transcriptomic profiles from other species would not mirror the findings in *Ae. albopictus*. It may be plausible that thermoregulatory functions are not as enriched in other species and that other functions like immune or metabolic functions are more relevant following a blood meal, highlighting the divergence in adapted functions to survive a blood meal. However, a strong enrichment for thermoregulatory functions at this early time point in other species outside of *Ae.*

albopictus would signify the convergence of gene functions across mosquitoes to mitigate thermal stress before metabolic processes are involved.

4.3 Limitations and Future Analyses

Studies about mosquito blood feeding have typically focused on the longer-term dynamics on the organism following a blood meal and have not fully studied the transcriptomic profile immediately PBM. Here, we present a comparative transcriptomics approach to observe the mosquito blood feeding response with bulk RNA-seq of a time point that has not typically been focused on. These findings detect and quantify the entire early-stage BF profile that previous works have not before, and transcripts detected from our approach have also been functionally annotated from two different scopes. Comparing these transcripts to orthologs across mosquito species highlights the variation in responses and understanding these dynamics potentially may play a role in vector control. A point of consideration for conducting a comparative transcriptomic analysis of the blood feeding response is how these heat shock dynamics may differ with habitat ranges. For example, certain strains of *Ae. albopictus* tend to proliferate in more temperate climates while certain strains of *Cx. quinquefasciatus* are localized to more tropical climates, but both species require blood meals that may incorporate different genes for its thermoregulation (Zhong et al., 2013).

However, there are still many undetermined questions that exist about mosquito blood meals because of our limitations in our methods or in existing knowledge. Mosquito blood meals have a differential role in expression based on their localization, as for example, the midget of a mosquito is known to express different levels of Hsp70 following a blood meal than the wings (Benoit et al., 2011). Understanding the convergence and divergence of early-stage blood feeding

responses could be further augmented by bulk RNA-seq of dissected regions but would add an additional layer of complexity to our comparisons. Furthermore, there are also the unexplored profiles of a mosquito blood meal in its later stages which would help to capture the dynamics in transcriptional products over time, but this analysis has not been conducted yet. Additionally, existing genome annotations and homology-based functional annotations have not been able to comprehensively identify relevant gene and functions following a blood meal as the most highly upregulated transcriptional products have no genome annotations. Our whole-genome RNA-seq and functional analysis captures that these products are relevant in an early-stage blood meal, but it is unknown if these are related to protein-coding genes, ncRNA, or another type of transcriptional product.

Our analysis exhibits functional relationships that could be relevant starting points for further analysis. The divergence in the heat shock response across mosquitoes indicate adapted functions for surviving thermal stress, which we would be able to fully explore with the enrichments of genes across experimental conditions from other species. For example, there are significantly enriched GO terms for "response to hypoxia" in *Ae. albopictus* or "structural constituent of eye lens" in *Ae. aegypti* which are functional descriptions from the heat shock response that currently have not been explored. There are gaps in knowledge not only about the comparisons of upregulated DEGs in the mosquito heat shock response but also in the downregulated DEGs of the heat shock response and even the blood feeding response. We suspect that exploring these deviations in responses would be beneficial in understanding the lifespans and response dynamics in mosquitoes that could affect the evolution and vector control strategies of a mosquito.

FIGURES

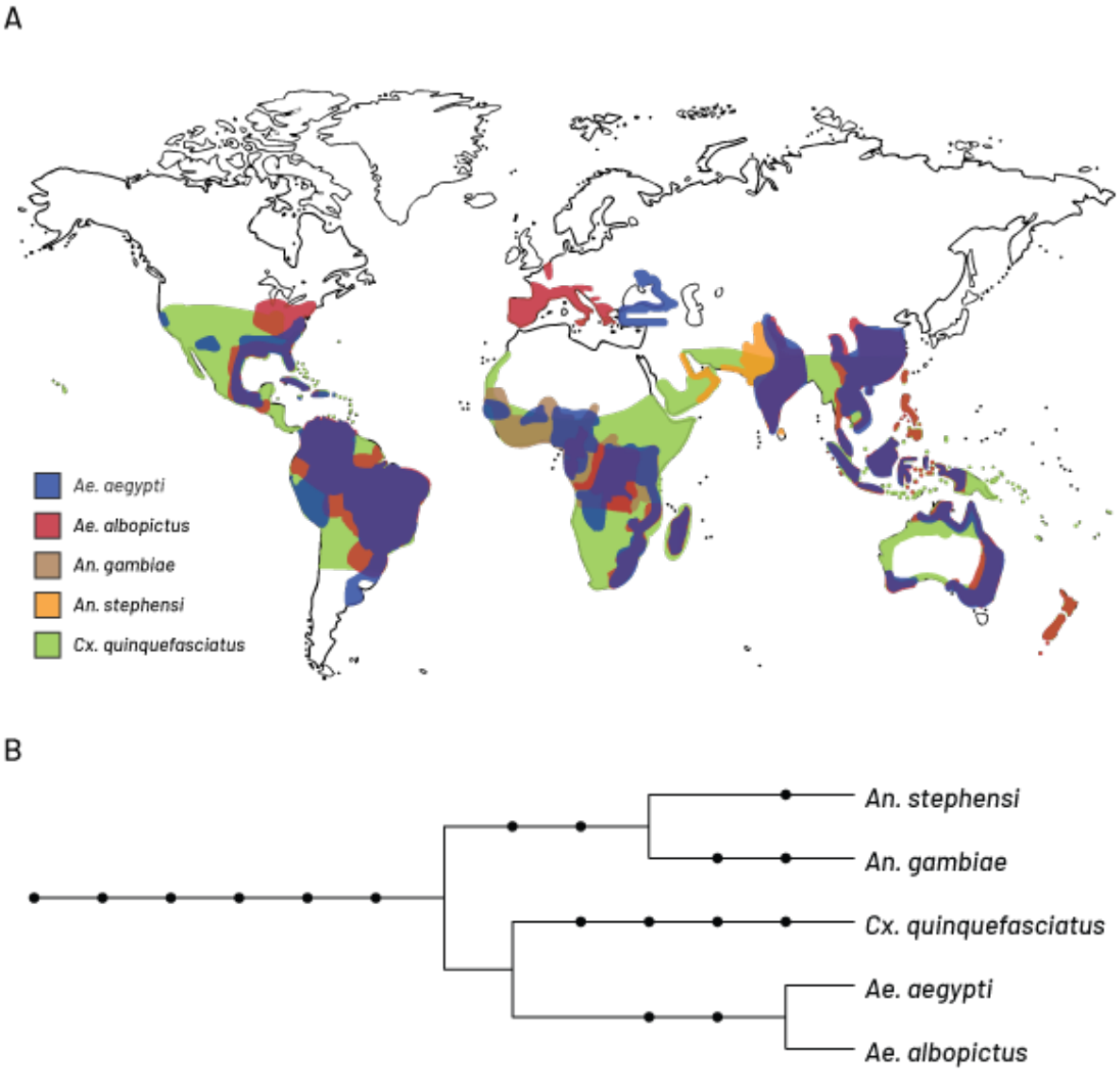
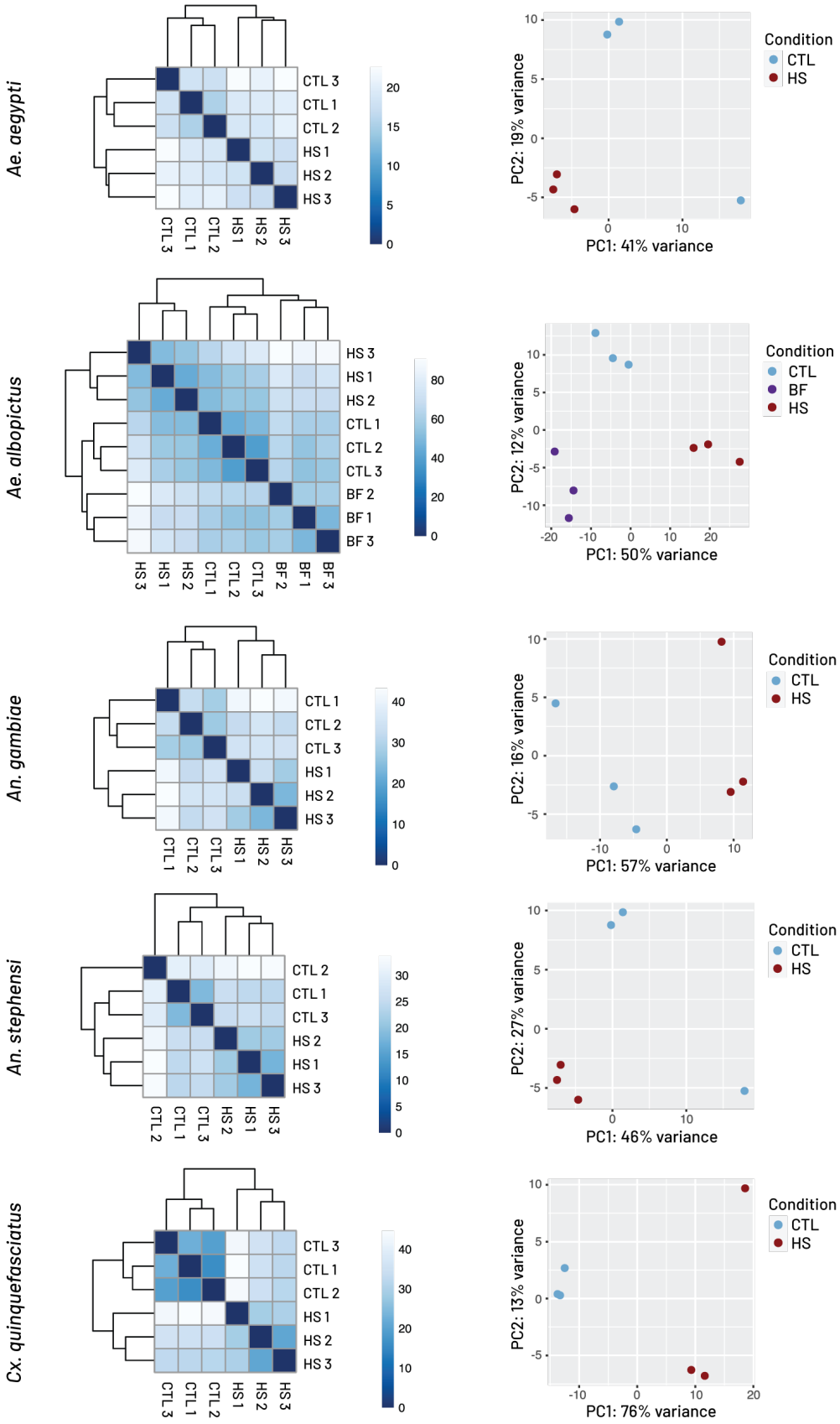


Figure 1. Habitat ranges and phylogenetic relationships of the five sampled mosquito species. (A) A map of the global habitat ranges of the five mosquito species: *Ae. aegypti* and *Ae. albopictus* adapted from Lwande et al. (2019), *An. gambiae* and *An. stephensi* adapted from Kiszewski et al. (2004), and *Cx. quinquefasciatus* adapted from Shocket et al. (2020). (B) A maximum likelihood-based phylogenetic inference of the five mosquito species (Letunic & Bork, 2007).

Figure 2. Sample-level QC of RNA-seq replicates and experimental conditions by hierarchical clustering of sample-to-sample distances and principal component analysis. Replicates are sorted by experimental condition from standard rearing conditions (CTL), environmental heat shock (HS), or early-stage blood feeding (BF).


Ae. aegypti
CTL 3
CTL 1
CTL 2
HS 1
HS 2
HS 3
20
15
10
5
0
PC2: 19% variance
PC1: 41% variance
Condition
CTL
HS
Ae. albopictus
HS 3
HS 1
HS 2
CTL 1
CTL 2
CTL 3
BF 2
BF 1
BF 3
80
60
40
20
0
PC2: 12% variance
PC1: 50% variance
Condition
CTL
BF
HS
An. gambiae
CTL 1
CTL 2
CTL 3
HS 1
HS 2
HS 3
40
30
20
10
0
PC2: 16% variance
PC1: 57% variance
Condition
CTL
HS
An. stephensi
CTL 2
CTL 1
CTL 3
HS 2
HS 1
HS 3
30
25
20
15
10
5
0
PC2: 27% variance
PC1: 46% variance
Condition
CTL
HS
Cx. quinquefasciatus
CTL 3
CTL 1
CTL 2
HS 1
HS 2
HS 3
40
30
20
10
0
PC2: 13% variance
PC1: 76% variance
Condition
CTL
HS

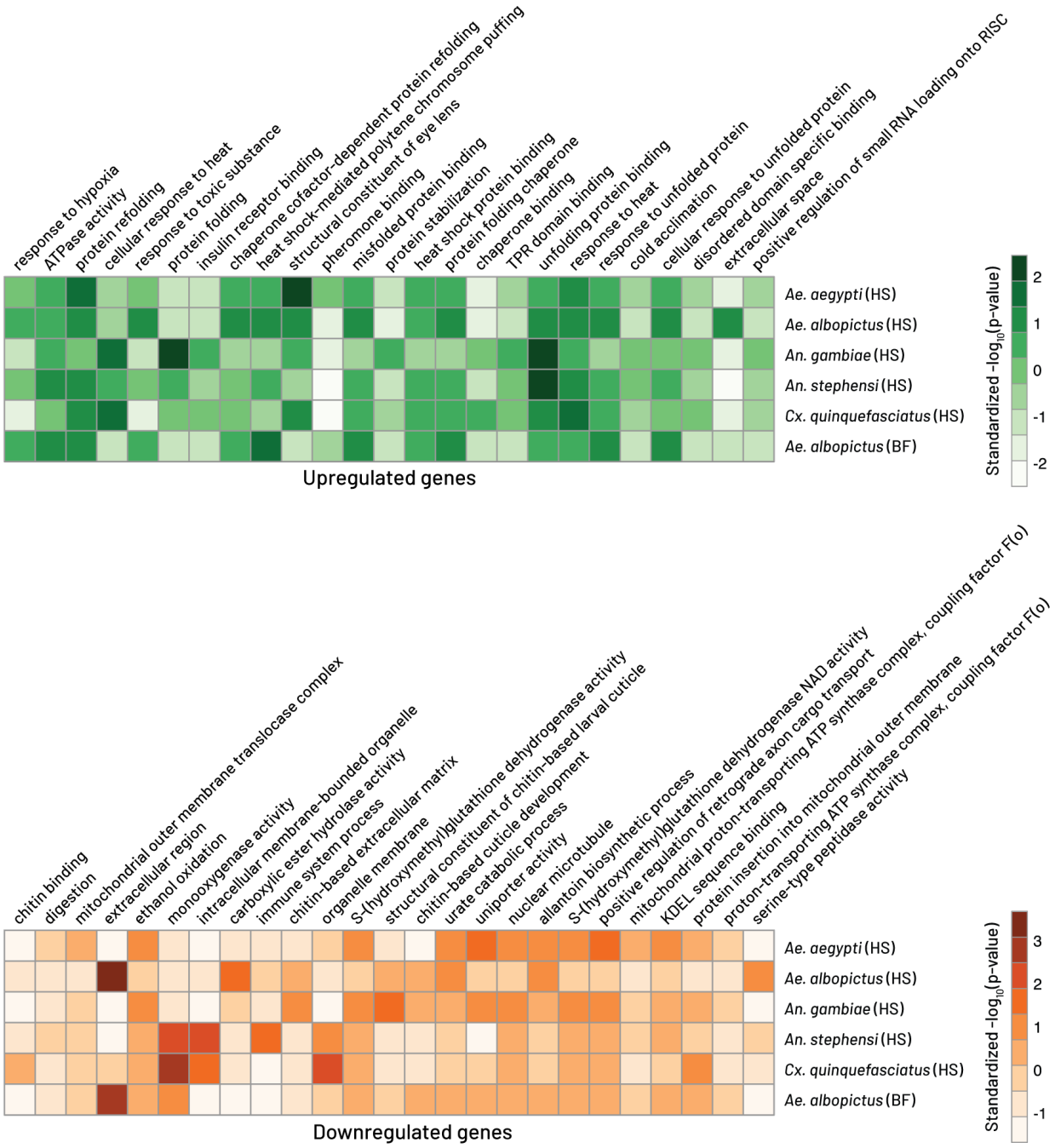


Figure 3. Heatmap of a subset of the standardized and log-transformed p-values from a GO term enrichment analysis. Enrichment p-values were determined from a binomial test of the observed GO terms from upregulated or downregulated DEGs under HS and BF conditions to the whole-genome functional annotations. The subset of GO term enrichments visualized are selected from the top 25 unique GO terms with the highest enrichment values from any sampled species.

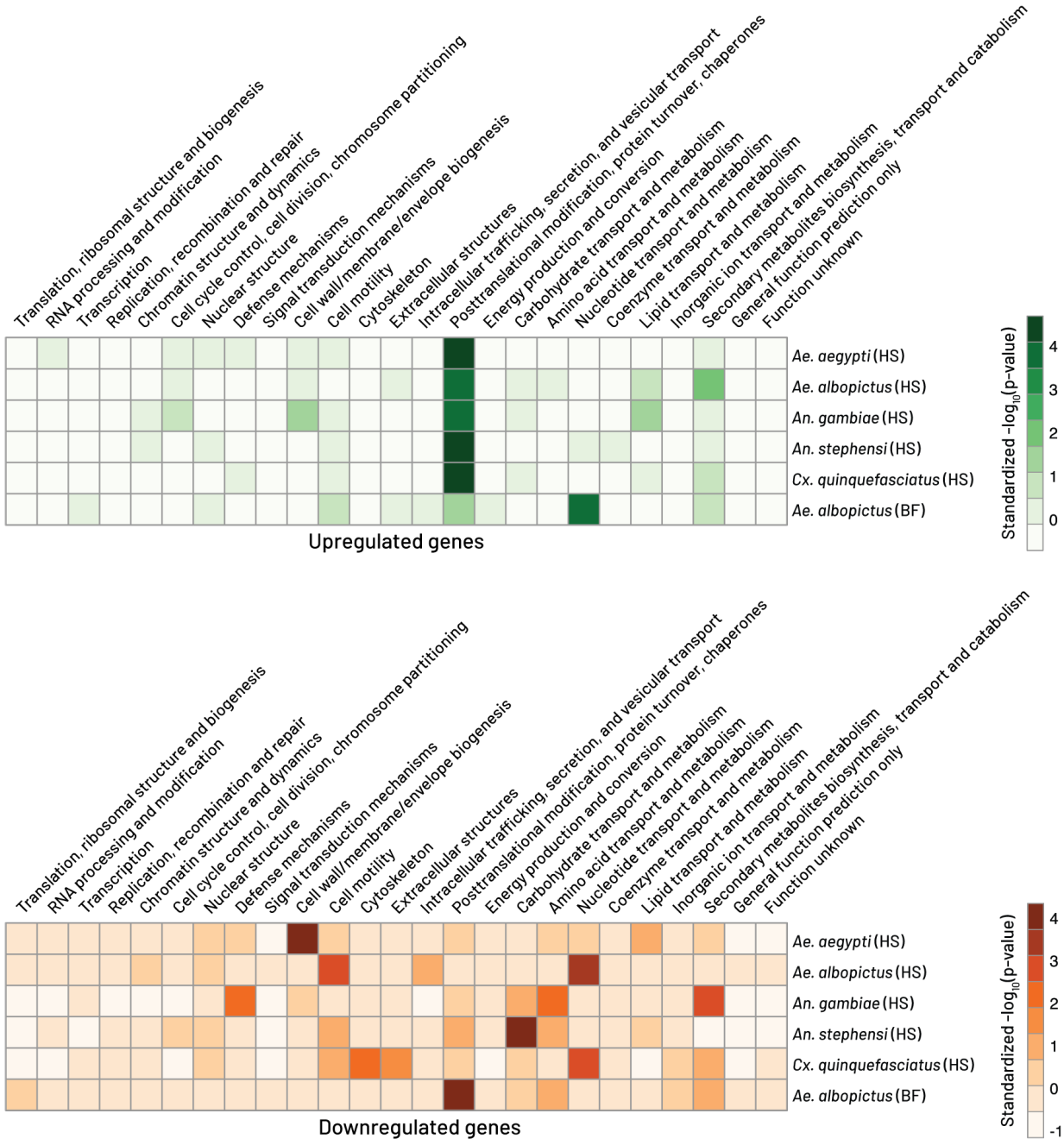
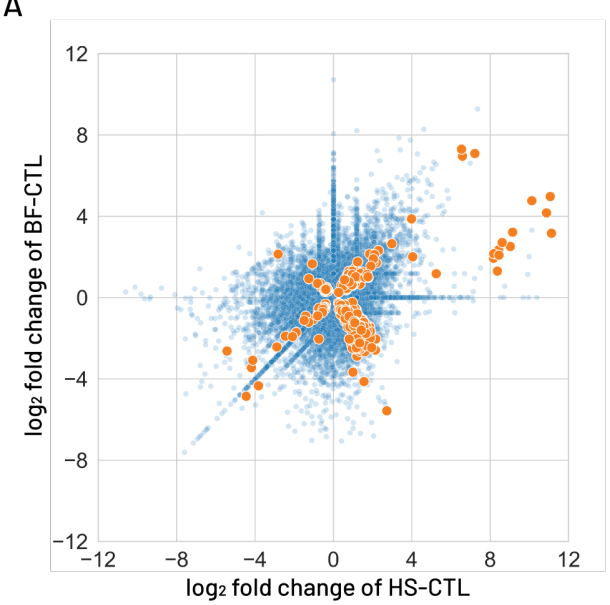


Figure 4. Heatmap of the standardized and log-transformed p-values from a COG category enrichment analysis. Enrichment p-values were determined from a binomial test of the observed upregulated or downregulated DEGs from HS and BF conditions to the whole-genome functional annotations.

Figure 5. Intersection of significantly expressed genes between HS and BF conditions in *Ae. albopictus*. (A) Scatterplot highlighting the 316 significantly expressed genes (orange) common to both HS and BF conditions from all 33823 genes in the *Ae. albopictus* transcriptome (blue). (B) Table quantifying the number of significantly expressed genes from each quadrant and condition parameters from (A). (C) Bar plots of the $-\log_{10}$-scaled p-values determined from a GO term (purple) and COG category (teal) enrichment analysis of the observed genes in each quadrant from (A). (A-C) Significantly expressed genes used in comparisons have an absolute $\log_2$ fold change > 0 over the control and a Benjamini-Hochberg adjusted p-value < 0.05.

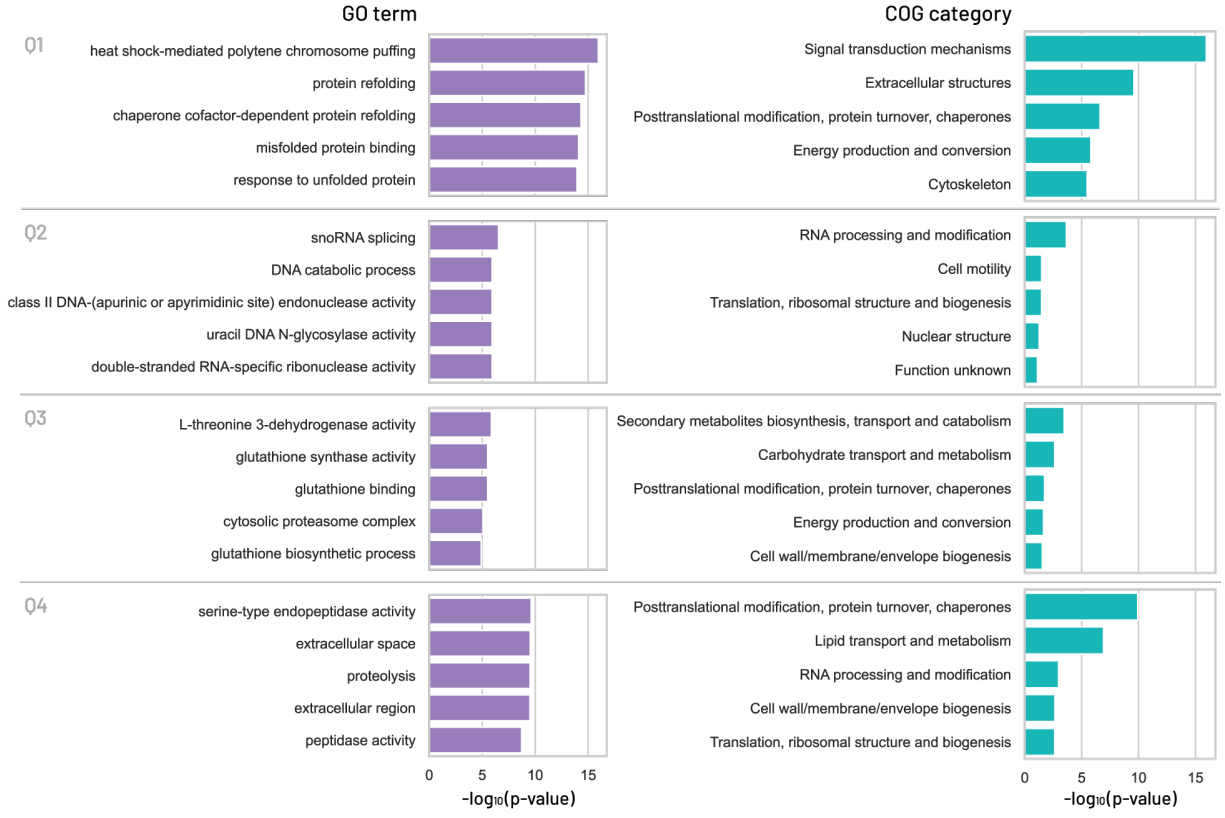
A



B

Quadrant	Condition	Significant genes
1	High HS, High BF (HS LFC > 0, BF LFC > 0)	91
2	Low HS, High BF (HS LFC < 0, BF LFC > 0)	24
3	Low HS, Low BF (HS LFC < 0, BF LFC < 0)	26
4	High HS, Low BF (HS LFC > 0, BF LFC < 0)	175

C



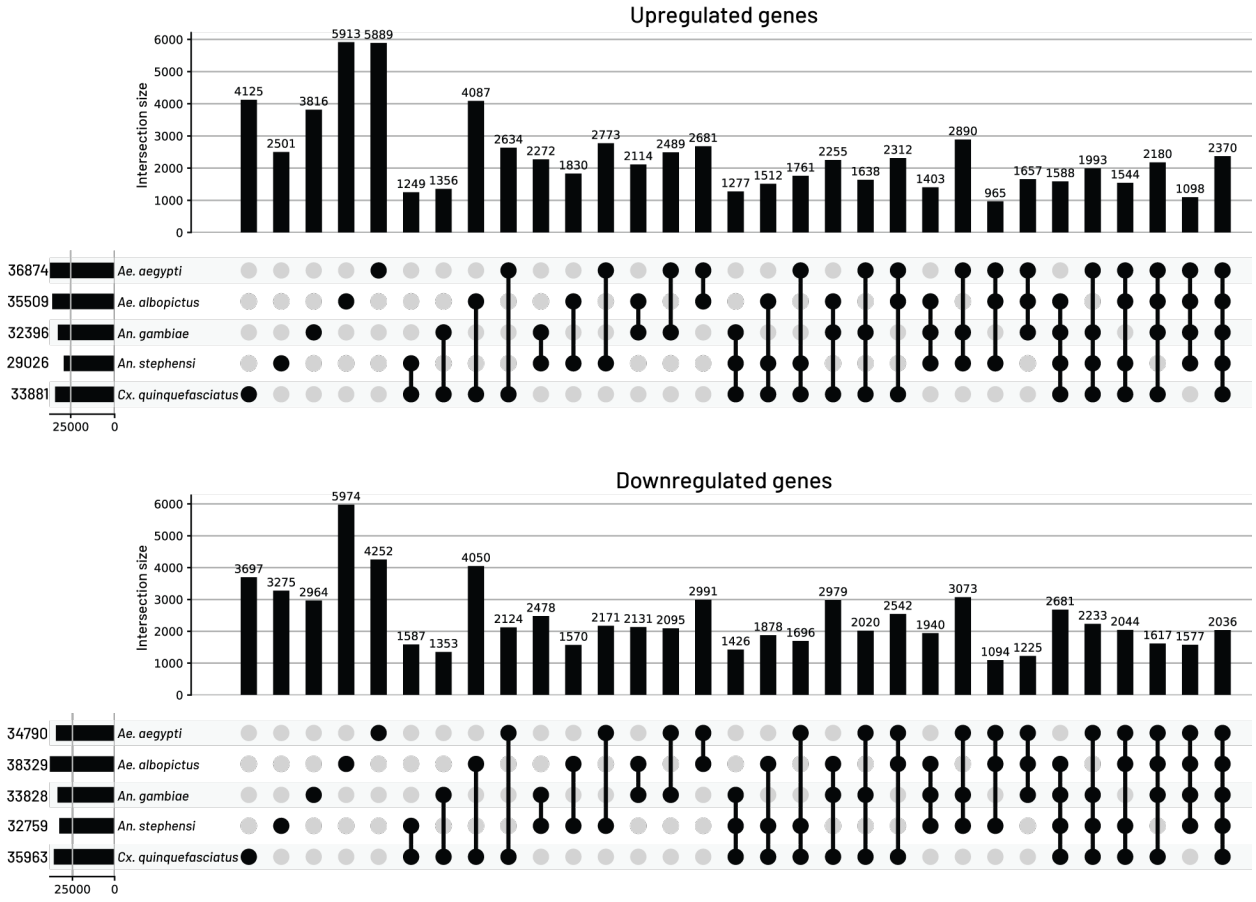


Figure 6. UpSet plots of paralog-weighted orthologous gene groups across mosquito species under HS conditions. Only genes that have orthologs from all other species and an absolute $\log_2$ fold change > 0 over control conditions are mapped. The orthologous gene from each species that was selected was the most statistically significant of its paralog group (either the highest expression level among Bonferroni-adjusted p-values below 0.05 or the expression level with the lowest non-adjusted p-value).

TABLES

Table 1. Overview of RNA-seq sample quality assessments from FastQC. Assessments are sorted by species; experimental condition from standard rearing conditions (CTL), environmental heat shock (HS), or early-stage blood feeding (BF); replicate; and sequencing end from the forward end (+) or the reverse end (-). Q30 is the percentage of reads with a Phred quality score of 30 or higher.

Species	Condition / Replicate / End	Read counts	Unique reads (%)	Q30 (%)
Ae. aegypti	CTL 1 (+)	26307541	48.10	98.83
	CTL 1 (-)	26307541	53.20	95.47
	CTL 2 (+)	40688516	39.29	98.71
	CTL 2 (-)	40688516	41.98	96.56
	CTL 3 (+)	44648520	37.66	98.75
	CTL 3 (-)	44648520	42.92	93.05
	HS 1 (+)	35814327	43.65	98.31
	HS 1 (-)	35814327	47.24	95.39
	HS 2 (+)	32148732	45.69	98.97
	HS 2 (-)	32148732	49.56	95.26
	HS 3 (+)	27130526	49.09	98.62
	HS 3 (-)	27130526	54.18	95.68
Ae. albopictus	CTL 1 (+)	39572779	52.32	97.90
	CTL 2 (+)	46030689	48.78	97.98
	CTL 3 (+)	47688301	49.25	97.68
	HS 1 (+)	44499875	49.05	97.90
	HS 2 (+)	39422924	50.85	97.74
	HS 3 (+)	39310058	45.57	97.80
	BF 1 (+)	37682081	54.15	97.72
	BF 2 (+)	28405546	22.43	97.87
	BF 3 (+)	45084893	51.93	97.83
An. gambiae	CTL 1 (+)	35598460	44.89	98.58
	CTL 1 (-)	35598460	48.89	95.35
	CTL 2 (+)	39377616	42.37	98.13
	CTL 2 (-)	39377616	44.20	95.95
	CTL 3 (+)	39037985	42.60	98.58
	CTL 3 (-)	39037985	44.96	93.97
	HS 1 (+)	34978870	46.91	98.23

Table 1. Overview of RNA-seq sample quality assessments from FastQC, Continued. Assessments are sorted by species; experimental condition from standard rearing conditions (CTL), environmental heat shock (HS), or early-stage blood feeding (BF); replicate; and sequencing end from the forward end (+) or the reverse end (-). Q30 is the percentage of reads with a Phred quality score of 30 or higher.

Species	Condition / Replicate / End	Read counts	Unique reads (%)	Q30 (%)
An. gambiae	HS 1 (-)	34978870	49.93	94.73
	HS 2 (+)	32064250	47.40	98.21
	HS 2 (-)	32064250	51.96	91.34
	HS 3 (+)	30269332	50.73	98.68
	HS 3 (-)	30269332	54.60	94.86
An. stephensi	CTL 1 (+)	45027987	36.10	98.65
	CTL 1 (-)	45027987	37.76	95.96
	CTL 2 (+)	29590523	42.08	98.81
	CTL 2 (-)	29590523	44.13	96.64
	CTL 3 (+)	45795101	35.34	98.09
	CTL 3 (-)	45795101	38.34	94.61
	HS 1 (+)	27436551	45.67	98.96
	HS 1 (-)	27436551	46.36	97.07
	HS 2 (+)	28346313	44.36	99.02
	HS 2 (-)	28346313	45.61	97.29
	HS 3 (+)	33953670	42.04	98.21
	HS 3 (-)	33953670	45.07	94.64
Cx. quinquefasciatus	CTL 1 (+)	39341788	42.94	98.25
	CTL 1 (-)	39341788	45.89	94.29
	CTL 2 (+)	41566758	42.39	99.01
	CTL 2 (-)	41566758	44.34	97.26
	CTL 3 (+)	35657073	45.81	98.81
	CTL 3 (-)	35657073	49.32	96.51
	HS 1 (+)	34345435	46.69	98.49
	HS 1 (-)	34345435	50.27	96.35
	HS 2 (+)	41395965	41.96	98.57
	HS 2 (-)	41395965	46.18	95.41
	HS 3 (+)	39237600	43.96	98.58
	HS 3 (-)	39237600	46.74	96.33

Table 2. Overview of transcriptome functional annotation quality. The percentage of GO annotation is calculated from if any hits of the BLAST or InterProScan searches returned GO terms. The percentage of COG annotation is calculated from if any protein match was made to any COG category. The percentage of annotation to known COG categories excludes undetermined matches and matches to the COG category "Function unknown".

Species	Transcriptome annotated with any GO term (%)	Transcriptome annotated with any COG category (%)	Transcriptome annotated with known COG categories (%)
Ae. aegypti	88.17	92.06	66.16
Ae. albopictus	83.78	91.32	64.17
An. gambiae	86.24	91.78	65.23
An. stephensi	83.85	92.72	68.43
Cx. quinquefasciatus	87.84	91.44	64.50

Table 3. Differentially expressed genes from experimental conditions over the control. The analysis was performed between the environmentally heat shocked samples to the control (HS-CTL) or the short-term blood fed samples to the control (BF-CTL). “Total DEGs” is the number of differentially expressed genes (DEGs) that meet the threshold cutoff of an absolute $\log_2$ fold change > 1 and a Benjamini-Hochberg adjusted p-value < 0.05. “Upregulated DEGs” and “Downregulated DEGs” are the number of genes from the “Total DEGs” that have a $\log_2$ fold change > 1 or a $\log_2$ fold change < -1, respectively.

Species	Condition	Total genes in transcriptome	Total DEGs	Upregulated DEGs	Downregulated DEGs	Upregulated DEGs of transcriptome (%)	Downregulated DEGs of transcriptome (%)
Ae. aegypti	HS-CTL	17866	77	42	35	0.24	0.20
Ae. albopictus	HS-CTL	33823	637	475	162	1.40	0.48
	BF-CTL	33823	823	177	646	0.52	1.91
An. gambiae	HS-CTL	13052	209	96	113	0.74	0.87
An. stephensi	HS-CTL	16435	122	83	39	0.51	0.24
Cx. quinquefasciatus	HS-CTL	18884	235	198	37	1.05	0.20

Table 4. Summary of Helitron TE and HSE calls in mosquito genomes. “Helitron calls” are the number of transposable elements detected from a search using a species-specific Helitrons library in RepeatMasker. “HSE calls” are the number of HSEs detected using a regular expression search for the canonical HSE motif. “HSEs in genome” is the proportion of the genome in bases that is comprised of HSEs. “Peak number of HSEs in a 1-kb window” refers to the highest sum of HSEs detected and clustered in a 1-kb section of the genome assembly.

Species	Helitron calls	HSE calls	HSEs in genome (%)	Peak number of HSEs in a 1-kb window
Ae. aegypti	205962	62569	0.0734	40
Ae. albopictus	1460743	179009	0.1059	91
An. gambiae	3266	3299	0.0181	4
An. stephensi	12866	2978	0.0183	4
Cx. quinquefasciatus	29686	21529	0.0558	56

This thesis is co-authored by Bryant Cao, Stephanie Gamez, Michelle Bui, Brian Tsu, Omar Akbari, and Matthew Daugherty. The thesis author was primary author of this paper.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Attardo, G. M., Hansen, I. A., & Raikhel, A. S. (2005). Nutritional regulation of vitellogenesis in mosquitoes: Implications for anautogeny. *Insect Biochemistry and Molecular Biology*, *35*(7), 661–675. https://doi.org/10.1016/j.ibmb.2005.02.013

Benoit, J. B., Lopez-Martinez, G., Patrick, K. R., Phillips, Z. P., Krause, T. B., & Denlinger, D. L. (2011). Drinking a hot blood meal elicits a protective heat shock response in mosquitoes. *Proceedings of the National Academy of Sciences*, *108*(19), 8026–8029. https://doi.org/10.1073/pnas.1105195108

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Bateman, A., & Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Bryant, W. B., & Michel, K. (2014). Blood feeding induces hemocyte proliferation and activation in the African malaria mosquito, Anopheles gambiae Giles. *The Journal of Experimental Biology*, *217*(8), 1238–1245. https://doi.org/10.1242/jeb.094573

Chakraborty, M., Ramaiah, A., Adolfi, A., Halas, P., Kaduskar, B., Ngo, L. T., Jayaprasad, S., Paul, K., Whadgar, S., Srinivasan, S., Subramani, S., Bier, E., James, A. A., & Emerson, J. J. (2021). Hidden genomic features of an invasive malaria vector, Anopheles stephensi, revealed by a chromosome-level genome assembly. *BMC Biology*, *19*(1), 28. https://doi.org/10.1186/s12915-021-00963-z

Chen, Q., Pei, D., Li, J., Jing, C., Wu, W., & Man, Y. (2017). The antenna transcriptome changes in mosquito Anopheles sinensis, pre- and post- blood meal. *PLoS ONE*, *12*(7). https://doi.org/10.1371/journal.pone.0181399

Cosby, R. L., Chang, N.-C., & Feschotte, C. (2019). Host–transposon interactions: Conflict, cooperation, and cooption. *Genes & Development*, *33*(17–18), 1098–1116. https://doi.org/10.1101/gad.327312.119

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., & Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*, *371*(6531). https://doi.org/10.1126/science.abc6405

Dana, A. N., Hong, Y. S., Kern, M. K., Hillenmeyer, M. E., Harker, B. W., Lobo, N. F., Hogan, J. R., Romans, P., & Collins, F. H. (2005). Gene expression patterns associated with blood-feeding in the malaria mosquito Anopheles gambiae. *BMC Genomics*, *6*, 5. https://doi.org/10.1186/1471-2164-6-5

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Ewing, B., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, *8*(3), 186–194. https://doi.org/10.1101/gr.8.3.186

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*, *8*(3), 175–185. https://doi.org/10.1101/gr.8.3.175

Fedorka, K. M., Copeland, E. K., & Winterhalter, W. E. (2013). Seasonality influences cuticle melanization and immune defense in a cricket: Support for a temperature-dependent immune investment hypothesis in insects. *The Journal of Experimental Biology*, *216*(Pt 21), 4005–4010. https://doi.org/10.1242/jeb.091538

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, *9*(5), 397–405. https://doi.org/10.1038/nrg2337

Freitas, L., & Nery, M. F. (2020). Expansions and contractions in gene families of independently-evolved blood-feeding insects. *BMC Evolutionary Biology*, *20*. https://doi.org/10.1186/s12862-020-01650-3

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., & Koonin, E. V. (2021). COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, *49*(D1), D274–D281. https://doi.org/10.1093/nar/gkaa1018

Garrigues, J. M., Tsu, B. V., Daugherty, M. D., & Pasquinelli, A. E. (2019). Diversification of the Caenorhabditis heat shock response by Helitron transposable elements. *ELife*, *8*, e51139. https://doi.org/10.7554/eLife.51139

Gidalevitz, T., Prahlad, V., & Morimoto, R. I. (2011). The stress of protein misfolding: From single cells to multicellular organisms. *Cold Spring Harbor Perspectives in Biology*, *3*(6). https://doi.org/10.1101/cshperspect.a009704

Giri, A., Heckathorn, S., Mishra, S., & Krause, C. (2017). Heat Stress Decreases Levels of Nutrient-Uptake and -Assimilation Proteins in Tomato Roots. *Plants*, *6*(1). https://doi.org/10.3390/plants6010006

Gomez-Pastor, R., Burchfiel, E. T., & Thiele, D. J. (2018). Regulation of heat shock transcription factors and their roles in physiology and disease. *Nature Reviews Molecular Cell Biology*, *19*(1), 4–19. https://doi.org/10.1038/nrm.2017.73

González-Santoyo, I., & Córdoba-Aguilar, A. (2012). Phenoloxidase: A key component of the insect immune system. *Entomologia Experimentalis et Applicata*, *142*(1), 1–16. https://doi.org/10.1111/j.1570-7458.2011.01187.x

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Horváth, V., Merenciano, M., & González, J. (2017). Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. *Trends in Genetics*, *33*(11), 832–841. https://doi.org/10.1016/j.tig.2017.08.007

Hou, Y., Wang, X.-L., Saha, T. T., Roy, S., Zhao, B., Raikhel, A. S., & Zou, Z. (2015). Temporal Coordination of Carbohydrate Metabolism during Mosquito Reproduction. *PLoS Genetics*, *11*(7). https://doi.org/10.1371/journal.pgen.1005309

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., & Sachs, J. (2004). A GLOBAL INDEX REPRESENTING THE STABILITY OF MALARIA TRANSMISSION. *The American Journal of Tropical Medicine and Hygiene*, *70*(5), 486–498. https://doi.org/10.4269/ajtmh.2004.70.486

Klowden, M. J., & Briegel, H. (1994). Mosquito gonotrophic cycle and multiple feeding potential: Contrasts between Anopheles and Aedes (Diptera: Culicidae). *Journal of Medical Entomology*, *31*(4), 618–622. https://doi.org/10.1093/jmedent/31.4.618

Kraemer, M. U. G., Reiner, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Bisanzio, D.,

Perkins, T. A., Lai, S., Lu, X., Jones, P., Coelho, G. E., Carvalho, R. G., Van Bortel, W., Marsboom, C., Hendrickx, G., Schaffner, F., Moore, C. G., Nax, H. H., Bengtsson, L., Wetter, E., Tatem, A. J., Brownstein, J. S., Smith, D. L., Lambrechts, L., Cauchemez, S., Linard, C., Faria, N. R., Pybus, O. G., Scott, T. W., Liu, Q., Yu, H., Wint, G. R. W., Hay, S. I., & Golding, N. (2019). Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nature Microbiology*, *4*(5), 854–863. https://doi.org/10.1038/s41564-019-0376-y

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), 127–128. https://doi.org/10.1093/bioinformatics/btl529

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

Lippi, C. A., Stewart-Ibarra, A. M., Loor, M. E. F. B., Zambrano, J. E. D., Lopez, N. A. E., Blackburn, J. K., & Ryan, S. J. (2019). Geographic shifts in Aedes aegypti habitat suitability in Ecuador using larval surveillance data and ecological niche modeling: Implications of climate change for public health vector control. *PLOS Neglected Tropical Diseases*, *13*(4), e0007322. https://doi.org/10.1371/journal.pntd.0007322

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lwande, O. W., Obanda, V., Lindström, A., Ahlm, C., Evander, M., Näslund, J., & Bucht, G. (2019). Globe-Trotting Aedes aegypti and Aedes albopictus: Risk Factors for Arbovirus Pandemics. *Vector-Borne and Zoonotic Diseases*, *20*(2), 71–81. https://doi.org/10.1089/vbz.2019.2486

Melo, E. S. de, & Wallau, G. L. (2020). Mosquito genomes are frequently invaded by transposable elements through horizontal transfer. *PLOS Genetics*, *16*(11), e1008946. https://doi.org/10.1371/journal.pgen.1008946

Metzger, M. J., Paynter, A. N., Siddall, M. E., & Goff, S. P. (2018). Horizontal transfer of retrotransposons between bivalves and other aquatic species of multiple phyla. *Proceedings of the National Academy of Sciences*, *115*(18), E4227–E4235. https://doi.org/10.1073/pnas.1717227115

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1), D419–D426. https://doi.org/10.1093/nar/gky1038

Nguyen, V. T., Morange, M., & Bensaude, O. (1989). Protein denaturation during heat shock and related stress. Escherichia coli beta-galactosidase and Photinus pyralis luciferase inactivation in mouse cells. *The Journal of Biological Chemistry*, *264*(18), 10487–10492.

Omodior, O., Luetke, M. C., & Nelson, E. J. (2018). Mosquito-borne infectious disease, risk-perceptions, and personal protective behavior among U.S. international travelers. *Preventive Medicine Reports*, *12*, 336–342. https://doi.org/10.1016/j.pmedr.2018.10.018

Pakpour, N., Riehle, M. A., & Luckhart, S. (2014). Effects of ingested vertebrate-derived factors on insect immune responses. *Current Opinion in Insect Science*, *3*, 1–5. https://doi.org/10.1016/j.cois.2014.07.001

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon: Fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Pock Tsy, J.-M. L., Duchemin, J.-B., Marrama, L., Rabarison, P., Le Goff, G., Rajaonarivelo, V., & Robert, V. (2003). Distribution of the species of the Anopheles gambiae complex and first evidence of Anopheles merus as a malaria vector in Madagascar. *Malaria Journal*, *2*(1), 33. https://doi.org/10.1186/1475-2875-2-33

Samy, A. M., Elaagip, A. H., Kenawy, M. A., Ayres, C. F. J., Peterson, A. T., & Soliman, D. E. (2016). Climate Change Influences on the Global Potential Distribution of the Mosquito Culex quinquefasciatus, Vector of West Nile Virus and Lymphatic Filariasis. *PLOS ONE*, *11*(10), e0163863. https://doi.org/10.1371/journal.pone.0163863

Santiago, P. B., de Araújo, C. N., Motta, F. N., Praça, Y. R., Charneau, S., Bastos, I. M. D., & Santana, J. M. (2017). Proteases of haematophagous arthropod vectors are involved in blood-feeding, yolk formation and immunity—A review. *Parasites & Vectors*, *10*(1), 79. https://doi.org/10.1186/s13071-017-2005-z

Shibata, Y., & Morimoto, R. I. (2014). How the Nucleus Copes with Proteotoxic Stress. *Current Biology*, *24*(10), R463–R474. https://doi.org/10.1016/j.cub.2014.03.033

Shocket, M. S., Verwillow, A. B., Numazu, M. G., Slamani, H., Cohen, J. M., El Moustaid, F., Rohr, J., Johnson, L. R., & Mordecai, E. A. (2020). Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23°C and 26°C. *ELife*, *9*, e58511. https://doi.org/10.7554/eLife.58511

Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. *Science (New York, N.Y.)*, *278*(5338), 631–637. https://doi.org/10.1126/science.278.5338.631

The Gene Ontology Consortium, Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L.-P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang,

X., Mushayahama, T., LaBonte, S. A., Siegele, D. A., Antonazzo, G., Attrill, H., Brown, N. H., Garapati, P., Marygold, S. J., Trovisco, V., dos Santos, G., Falls, K., Tabone, C., Zhou, P., Goodman, J. L., Strelets, V. B., Thurmond, J., Garmiri, P., Ishtiaq, R., Rodríguez-López, M., Acencio, M. L., Kuiper, M., Lægreid, A., Logie, C., Lovering, R. C., Kramarz, B., Saverimuttu, S. C. C., Pinheiro, S. M., Gunn, H., Su, R., Thurlow, K. E., Chibucos, M., Giglio, M., Nadendla, S., Munro, J., Jackson, R., Duesbury, M. J., Del-Toro, N., Meldal, B. H. M., Paneerselvam, K., Perfetto, L., Porras, P., Orchard, S., Shrivastava, A., Chang, H.-Y., Finn, R. D., Mitchell, A. L., Rawlings, N. D., Richardson, L., Sangrador-Vegas, A., Blake, J. A., Christie, K. R., Dolan, M. E., Drabkin, H. J., Hill, D. P., Ni, L., Sitnikov, D. M., Harris, M. A., Oliver, S. G., Rutherford, K., Wood, V., Hayles, J., Bähler, J., Bolton, E. R., De Pons, J. L., Dwinell, M. R., Hayman, G. T., Kaldunski, M. L., Kwitek, A. E., Laulederkind, S. J. F., Plasterer, C., Tutaj, M. A., Vedi, M., Wang, S.-J., D'Eustachio, P., Matthews, L., Balhoff, J. P., Aleksander, S. A., Alexander, M. J., Cherry, J. M., Engel, S. R., Gondwe, F., Karra, K., Miyasato, S. R., Nash, R. S., Simison, M., Skrzypek, M. S., Weng, S., Wong, E. D., Feuermann, M., Gaudet, P., Morgat, A., Bakker, E., Berardini, T. Z., Reiser, L., Subramaniam, S., Huala, E., Arighi, C. N., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Bateman, A., Blatter, M.-C., Boutet, E., Bowler, E., Breuza, L., Bridge, A., Britto, R., Bye-A-Jee, H., Casas, C. C., Coudert, E., Denny, P., Estreicher, A., Famiglietti, M. L., Georghiou, G., Gos, A., Gruaz-Gumowski, N., Hatton-Ellis, E., Hulo, C., Ignatchenko, A., Jungo, F., Laiho, K., Le Mercier, P., Lieberherr, D., Lock, A., Lussi, Y., MacDougall, A., Magrane, M., Martin, M. J., Masson, P., Natale, D. A., Hyka-Nouspikel, N., Orchard, S., Pedruzzi, I., Pourcel, L., Poux, S., Pundir, S., Rivoire, C., Speretta, E., Sundaram, S., Tyagi, N., Warner, K., Zaru, R., Wu, C. H., Diehl, A. D., Chan, J. N., Grove, C., Lee, R. Y. N., Muller, H.-M., Raciti, D., Van Auken, K., Sternberg, P. W., Berriman, M., Paulini, M., Howe, K., Gao, S., Wright, A., Stein, L., Howe, D. G., Toro, S., Westerfield, M., Jaiswal, P., Cooper, L., & Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Tikar, S. N., Mendki, M. J., Sharma, A. K., Sukumaran, D., Veer, V., Prakash, S., & Parashar, B. D. (2011). Resistance status of the malaria vector mosquitoes, Anopheles stephensi and Anopheles subpictus towards adulticides and larvicides in arid and semi-arid areas of India. *Journal of Insect Science*, *11*(1). https://doi.org/10.1673/031.011.8501

Vega-Rúa, A., Marconcini, M., Madec, Y., Manni, M., Carraretto, D., Gomulski, L. M., Gasperi, G., Failloux, A.-B., & Malacrida, A. R. (2020). Vector competence of Aedes albopictus populations for chikungunya virus is shaped by their demographic history. *Communications Biology*, *3*(1), 1–13. https://doi.org/10.1038/s42003-020-1046-6

Vihervaara, A., Duarte, F. M., & Lis, J. T. (2018). Molecular mechanisms driving transcriptional stress responses. *Nature Reviews Genetics*, *19*(6), 385–397. https://doi.org/10.1038/s41576-018-0001-6

Vihervaara, A., & Sistonen, L. (2014). HSF1 at a glance. *Journal of Cell Science*, *127*(2), 261–266. https://doi.org/10.1242/jcs.132605

Wilder-Smith, A., Gubler, D. J., Weaver, S. C., Monath, T. P., Heymann, D. L., & Scott, T. W. (2017). Epidemic arboviral diseases: Priorities for research and public health. *The Lancet. Infectious Diseases*, *17*(3), e101–e106. https://doi.org/10.1016/S1473-3099(16)30518-7