UNIVERSITY OF CALIFORNIA,
IRVINE

Insert Brilliant Insightful Title Here

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

William Augustine Dunn

Dissertation Committee:
Anthony A. James, Chair
Xiaohui Xie
Donald F. Senear

2013

© 2013 William Augustine Dunn

DEDICATION

(Optional dedication page)
To ...

TABLE OF CONTENTS

Page

LIST OF FIGURES **vi**

LIST OF TABLES **vii**

ACKNOWLEDGMENTS **viii**

CURRICULUM VITAE **ix**

ABSTRACT OF THE DISSERTATION **xi**

1 Introduction **1**
1.1 Dengue fever and malaria are mosquito borne diseases 1
1.2 Hematophagy (bloodfeeding) is an essential part of mosquito-borne disease 2
1.2.1 Transmission cycle 2
1.3 Public health: *multiple targets* 4
1.3.1 Addressing the human side of the cycle 5
1.3.2 Addressing the vector side of the cycle (conventionally) 5
1.3.3 Transgenesis provides new vector control strategy 6
1.4 Elements of a successful transgenic vector control product 7
1.4.1 Introducing heritable transgenes into mosquito genomes 7
1.4.2 Effector genes 9
1.4.3 Controlling when and where the effector genes are expressed . . 11
1.4.4 Achieving swift, dramatic, and clinically meaningful effects on the native vector population 12
1.4.5 Conservation of bloodmeal-regulated transcription 14
1.5 Evolutionary relationship between mosquitoes and hematophagy 14
1.5.1 Implications for design of transgenic vector control systems? . . . 15

2 Review of Pertinent Publications **17**
2.1 Comparative genomics allows the discovery of *cis*-regulatory elements in mosquitoes. [115] 19
2.2 RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti* [20] 22
2.2.1 Transcripts found exclusively in bloodfed mosquitoes 23

2.2.2 *Cis*-regulatory element discovery 24
2.2.3 RNA-seq identifies annotation corrections 25
2.3 Complex modulation of the *Aedes aegypti* transcriptome in response to dengue virus infection. [21] 27
2.3.1 Results 29
2.3.2 Discussion 34
2.3.3 Conclusions 36
2.4 Strain variation in the transcriptome of the dengue fever vector, *Aedes aegypti*. [22] 39
2.4.1 Results 40
2.4.2 Discussion 46

3 Approach and Custom Software solutions **55**
3.1 The phylogenetic transcriptional correspondence index (PTCI) 59
3.2 Blacktie RNA-seq pipeline 62

4 Analysis of Conservation of Midgut Transcriptional Response to Bloodfeeding in Four Vector Mosquitoes **64**
4.1 Background 64

5 Contributions to Literature Not Otherwise Mentioned **69**
5.1 Purpose 69

6 Conclusions **70**
6.1 Purpose 70

Bibliography **71**

A Methods **87**
A.1 Comparative genomics allows the discovery of *cis*-regulatory elements in mosquitoes 87
A.1.1 Sequence Datasets 87
A.1.2 Discovery of Evolutionarily Conserved Putative CREs Among Mosquitoes 88
A.1.3 Clustering of Temporal- and Spatially Regulated *An. gambiae* . . 89
A.1.4 Determination of Association of Mosquito Motifs Within Expression Clusters 90
A.2 RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti* 91
A.2.1 Mosquito strains and rearing 91
A.2.2 RNA extraction and Illumina library preparation 91
A.2.3 Processing of Illumina sequencing data 92
A.2.4 Real-time quantitative RT-PCR validation of RNA-seq data . . . 92
A.2.5 Splice-site predictions 93
A.2.6 Motif discovery 94

A.3 Strain Variation in the Transcriptome of the Dengue Fever Vector, *Aedes aegypti* . . . 95
A.3.1 *Aedes aegypti* strains . . . 95
A.3.2 Weight comparisons of sugar and blood-fed mosquitoes . . . 95
A.3.3 RNA-seq library preparation and sequencing . . . 96
A.3.4 Reverse transcription gene amplification . . . 96
A.3.5 Quantitative real-time gene amplification . . . 97
A.3.6 RNA-seq data analyses . . . 97
A.4 Complex modulation of the *Aedes aegypti*transcriptome in response to dengue virus infection . . . 98
A.4.1 Mosquitoes . . . 98
A.4.2 Virus and Mosquito Oral Infection . . . 99
A.4.3 RNA-seq Library Preparation and Sequencing . . . 100
A.4.4 Data Analyses . . . 100

B Code **103**

LIST OF FIGURES

Page

1.1 Cross section of a female mosquito 3
1.2 *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti* 8
1.3 Elements of transcriptional regulation 12

2.1 Phylogenetic relationships between four vector mosquitoes 49
2.2 Associations of mosquito motifs with gene expression profiles in *An. gambiae* 50
2.3 Motif map of putative CREs from transcripts detected only in bloodfed female *Ae. aegypti* 51
2.4 Examples of amendments to the *Ae. aegypti* annotation supported by HMMSplicer results 52
2.5 MEME analysis of nine genes with $\mathrm{FPKM}_{\mathrm{DENVI}} \geq 100$ in carcasses and salivary glands at 14 dpi 53
2.6 Fatty acid biosynthesis and elongation in mitochondria 54

3.1 Exploring the parameter space of the expression and TFBS components of the PTCI 60
3.2 Graph model used to integrate data types 61
3.3 Diagram of Abbreviated Tophat/Cufflinks Inputs and Outputs 63

4.1 Pairwise PTCI results 65
4.2 Mean PTCI results 66
4.3 *An. gambiae* clustered abundance profiles 67

LIST OF TABLES

Page

1.1 Primary geographic vectors of dengue fever and malaria 2

2.1 Predicted novel junctions within various symmetric distance-windows from annotated transcripts 25

ACKNOWLEDGMENTS

I would like to thank...

(You must acknowledge grants and other funding assistance.

You may also acknowledge the contributions of professors and friends.

You also need to acknowledge any publishers of your previous work who have given you permission to incorporate that work into your dissertation. See Section 3.2 of the UCI Thesis and Dissertation Manual.)

CURRICULUM VITAE

William Augustine Dunn

EDUCATION

Doctor of Philosophy in Computer Science **2012**
University name *City, State*

Bachelor of Science in Computational Sciences **2007**
Another university name *City, State*

RESEARCH EXPERIENCE

Graduate Research Assistant **2007–2012**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2009–2010**
University name *City, State*

REFEREED JOURNAL PUBLICATIONS

Ground-breaking article **2012**
Journal name

REFEREED CONFERENCE PUBLICATIONS

Awesome paper **Jun 2011**
Conference name

Another awesome paper **Aug 2012**
Conference name

SOFTWARE

Magical tool http://your.url.here/
C++ algorithm that solves TSP in polynomial time.

ABSTRACT OF THE DISSERTATION

Insert Brilliant Insightful Title Here

By

William Augustine Dunn

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2013

Anthony A. James, Chair

The abstract of your contribution goes here.

Chapter 1

Introduction

1.1 Dengue fever and malaria are mosquito borne diseases

There are many mosquito borne diseases that afflict the world's population. Dengue fever and malaria are two important such diseases resulting in major health and productivity deficits in many developing nations worldwide CITEME. The causative agents of these two diseases are a suite of four species of RNA viruses (Dengue Virus 1-4) and a group of five species of protist from the genus *Plasmodium* (*Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*). However, in any **single** geographic location each disease is primarily transmitted by a single species of mosquito CITEME(Table 1.1).

The fact that both of these diseases are caused by multiple pathogen species complicates efforts to prevent human disease: particularly when addressed from the perspective of providing therapeutic or prophylactic care directly to human patients.

Disease	Region	Mosquito Vector
Dengue Viruses	Wolrdwide	*Aedes aegypti*
Human Malaria	Africa	*Anopheles gambiae*
	Asia	*Anopheles stephensi*

Table 1.1: Primary geographic vectors of dengue fever and malaria

TONY: MORE HERE ABOUT WHY THIS COMPLICATES THINGS?

1.2 Hematophagy (bloodfeeding) is an essential part of mosquito-borne disease

Another very important aspect of both of these diseases is that people do not become infected through interaction with other people, and mosquitoes do not become infected through interaction with other mosquitoes. The transmission cycles transmission of both diseases require that the pathogen be passed from human to mosquito or from mosquito to human. This pattern has important consequences that can be exploited to reduce the number of people who will become infected with dengue fever or malaria.

1.2.1 Transmission cycle

In Brief:

1. A female mosquito bites an infected person.
2. The pathogen is taken into the mosquito's midgut along with bloodmeal.
3. The pathogen must escape the midgut and gain access to the mosquito's circulatory system (the animal is now **infected**).

4. The pathogen must gain access to the mosquito's salivary glands (the animal is now **infectious**).

5. The infectious female mosquito bites an uninfected human and the pathogens in her saliva are introduced to the person's body as she feeds.

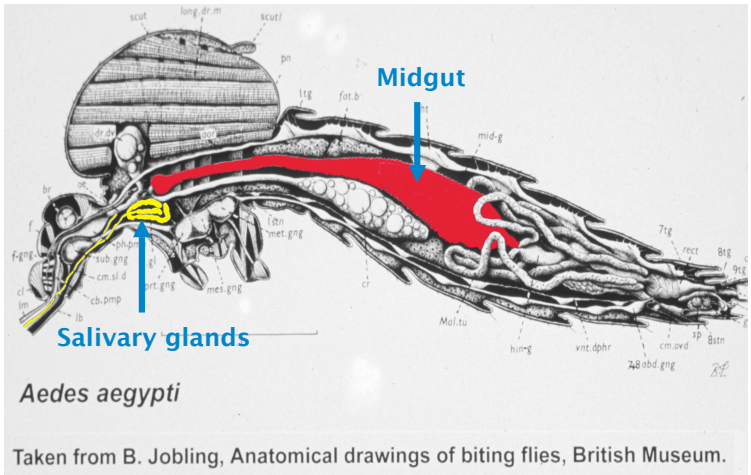


Figure 1.1: Cross section of a female mosquito with tissues important to the transmission cycle highlighted.
CITEME

While the specifics of how viruses and protists live and reproduce in mosquitoes and humans are quite different, the transmission cycles of dengue and malaria are very similar in their critical events. The transmission cycle of either pathogen type can be encapsulated into five fundamental events. First, a female mosquito feeds on an infected human's blood taking up the pathogen as well (Figure 1.1). The bloodmeal is digested in the mosquito's midgut, and escape from the midgut is the first crucial step for the survival of the pathogen inside the mosquito. If it is trapped in the midgut, it will eventually be passed as waste after the bloodmeal is digested. If it manages to escape to the hemolymph, the pathogen must survive long enough to gain access to the mosquito's salivary glands. If it is unable to infiltrate the salivary glands, then the mosquito, while itself **infected**, is not **infectious** to other humans. For this, the

pathogen must be injected into the bloodstream of the next human along with the contents of the mosquito's saliva.

1.3 Public health: *multiple targets*

Because the arrows of transmission essentially[1] points from mosquito to human or from human to mosquito, the cycle will collapse and a local area would eventually be cleared of the pathogen if either of those arrows are broken, provided that the intervention is maintained. This model suggests two primary targets for public health interventions aiming to reduce the incidence (followed by prevalence) of infection in a population:

1. Prevent mosquitoes from infecting humans.
2. Prevent humans from infecting mosquitoes.

However, many interventions can not be clearly divided into addressing purely the human or vector side of the cycle. For example, efforts to reduce the number of bites that humans receive from mosquitoes affects both the probabilities of the vector *and* the human becoming infected. This is one reason that vector control is almost always part of prevention strategies CITEME. For the purposes of this document, I will define human-based interventions to be those that are directly administered to a human body. Essentially: medical interventions.

In this respect, the long term goals of public health are more focused on prevention than the treatment of acute cases. Of course sick people need to be treated. Addition-

[1] While some mosquito to non-human vertebrate transmission may occur, it is not thought to be sufficient to maintain the pathogen if the mosquito:human loop is broken (CITEME).

ally, identifying and curing people who are infected also prevents mosquitoes from acquiring the pathogen as well. However, in the long term prevention is preferred.

1.3.1 Addressing the human side of the cycle

Because of the definition of human-based interventions above, there are relatively few effective options in this category for malaria and dengue fever. Normally, this section would include vaccines and swift, effective patient identification and treatment to clear the infection. However, at this time these types of interventions simply do not exist on the market or in a cost-effective form applicable to the isolated and impoverished areas that are most affected CITEME.

1.3.2 Non-transgenic options for addressing the vector side of the cycle

Medical options are scarce generally in the regions of the world where these diseases are most common. The alternative is to focus attention toward controlling access of the vector populations to human contact. This can include removal of nearby mosquito breeding sites (usually various forms of standing water), spraying of insecticides, introduction of biological predators, and/or bed nets, etc. One fairly recently developed approach that has shown great promise is sterile insect technique (SIT).

1.3.2.1 Sterile insect technique

SIT exploits a peculiar aspect of some insects' reproductive behavior. In many insects, the first mating event provides the only genetic material that the female will

use to fertilize her eggs for the rest of her life CITEME. Subsequent mating events contribute little to no genetic material to the females progeny, **even if the first event involves a sterile male**. This means that inundative numbers of sterilized males introduced into a native population will effectively sterilize significant proportions of wild females. This has dramatic effects on the local population. Negative population growth results from the inability of each generation to fully replace itself, and the effect is compounded for each subsequent generation as long as the program is maintained. Perhaps most famously, SIT was used successfully to eradicate the screwworm from the southern United States in the 1950s [25].

1.3.3 Transgenic techniques provide a new strategy for controlling disease transmission

Most conventional vector control strategies involve what might be termed vector population reduction. In addition to population reduction, transgenic techniques provide a novel vector control strategy that has considerable implications to the sustainability of the vector control aspect of public health interventions for dengue and malaria. In contrast to population reduction, this novel strategy could be conceptualized as vector population conversion. It's goal is not to reduce or eliminate contact between humans and vectors but to drive a transmission-refractory trait through the vector population that renders the vector unable to become infectious at all, converting it into a non-vector.

1.4 Elements of a successful transgenic vector control product

1. the ability to introduce transgenes into the vector genome in a heritable way
2. the discovery or construction of effector genes
3. the ability to control when and where the effector genes are expressed
4. the ability of your transgenic vectors to have swift, dramatic, and clinically meaningful effects on the native vector population (gene drive)

1.4.1 Introducing heritable transgenes into mosquito genomes

Before one can employ transgenic methods for vector control, one must have the capacity to make transgenic vectors. For mosquitoes, this is more labor intensive than for other fly species like *Drosophila melanogaster*. However, it has been accomplished, and protocols to reliably introduce transgenes into many mosquito species now exist CITEME. Unfortunately, the relative difficulty is high. It takes roughly 2-3 months to establish a stable line. Thus many experimental approaches that are straightforward or even routine in *D. melanogaster*or other species remain impractical for mosquitoes.

The method of introducing a transgene into a mosquito genome generally relies on re-purposing the activities of transposons (transposable elements). These nucleotide elements have the ability to cut themselves out of one part of the DNA and re-insert themselves into another. It is possible to engineer them to insert selected transgenes into the mosquito genome instead of the genes that they usually contain.

Until recently, most if not all transposons used for transgenesis in mosquitoes inserted the transgenes into more-or-less random locations. The number of locations and num-

ber of copies at any **single** genomic location was unpredictable [2, 113]. This makes characterization of the effects of a transgene complicated. However, with the successful implementation of the ϕC31 site-specific integration system in *Ae. aegypti* and *An. stephensi* [122, 93, 60], we can now reliably target specific, pre-characterized regions of these genomes for transgene integration of a single copy.

Heritability is achieved through specifically targeting the germline cells of mosquito embryos for transgenesis. In flies like mosquitoes, the cells that will become the germline always position themselves at one end (or pole) of the embryo (Figure 1.2). The transgenesis cocktail containing the transgene construct is injected into the embryos where the pole cells are located. This means that the first generation of mosquitoes that may contain the transgene in every cell of their body is **not** the injected generation but the progeny of the injected mosquitoes.

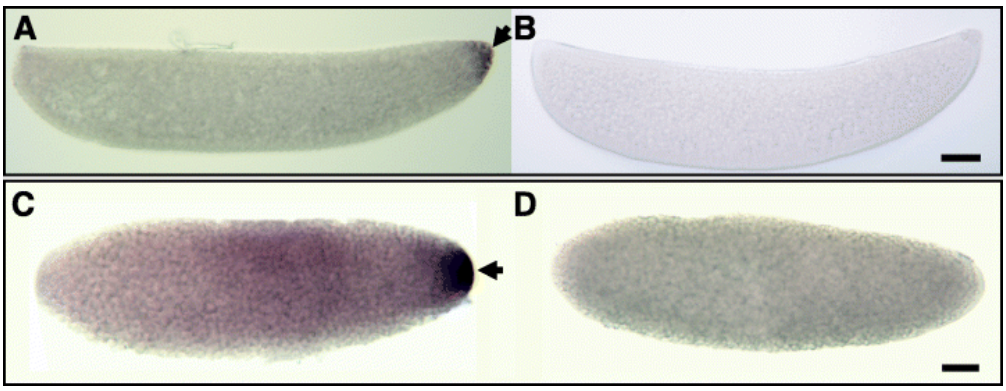


Figure 1.2: oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti:
Hybridizations *in situ* of *Anga osk* and *Aeae osk* RNA probes to whole-mount *An. gambiae* and *Ae. aegypti* embryos. Embryos are orientated with anterior on the left. *An. gambiae* embryos (A,B) shown have reached cellular blastoderm with pole cells indicated by an arrow. *Ae. aegypti* embryos (C,D) shown are in the preblastoderm stage. Strong staining is exclusive to the posterior pole cells (A) and posterior end (C) (arrows), when compared with embryos of the same stage, determined by DAPI staining (not shown), probed with *Anga* and *Aeae osk* sense RNA probes (B,D). Bar = 50 μm.

Excerpted from [63]

1.4.2 Effector genes

In addition to the ability to integrate transgene constructs into the vectors' genomes, it is necessary to design constructs that contain transgenes that will effect a useful response in the vector. Two broad strategies are commonly applied when designing an anti-transmission phenotype: population reduction and population conversion.

1.4.2.1 Vector population reduction:

The goal of vector population reduction is the same as conventional vector control modalities. *Reduce the local population of a disease vector with the goal of reducing the probability that an infectious individual will have an opportunity to participate in a transmission event.*

A **transgenic tactic** for the population reduction strategy is illustrated by Oxitec's *Ae. aegypti*strain OX3604C developed with support from the Bill and Melinda Gates Foundation's Grand Challenges for Global Health Initiative. Strain OX3604C represents a female-specific release of insects carrying a dominant lethal phenotype (RIDL) approach which uses a poison transgene that kills any female adult expressing the transgene unless the female is provided an "antidote" through its water supply [135, 13, 39].

Releasing enough of these mosquitoes will affect the local mosquito species population in a way that is analogous to SIT.

1.4.2.2 Vector population conversion:

Vector population conversion is a novel strategy for vector control that primarily exists within the context of genetically altered vectors (CITEMEwolbacia). It promises

to be the most long-lasting vector-focused intervention. Unlike **all** vector population reduction tactics, there is a practical point in a conversion intervention (the anti-transmission trait works and is present in local populations at near 100%) when the human management of the intervention can be ceased, yet it continues to function. For population reduction to approach this result, the vector species must not simply be eliminated from the local area, but approach elimination on a continental scale, or more realistically achieve global eradication.

The reason is that these mosquitoes (especially *Ae. aegypti*) can and do travel long distances, as eggs or larvae, in the backs of trucks (between villages) or in the pools of water collected in super-tankers (transcontinentally). So a local village is only free of the vectors until more migrate into the area. But in a conversion scenario, those migrants mate with the local vector population, and their offspring are assimilated into transmission-deficient mosquitoes. The protection of the local village can be preserved *even if some surrounding villages fail to maintain control of their mosquitoes*.

An example of a **transgenic tactic** for population conversion of *Ae. aegypti* into a transmission-deficient phenotype involves an effector gene that produces a double stranded RNA molecule that contains a portion of the dengue virus genome [43, 84]. Because most animal cells have a system that detects and degrades double stranded RNA[2] in a sequence specific manner, this primes the mosquito cells' antiviral response to specifically attack the dengue virus if the effector gene is expressed in the cell before it gets infected.

[2]Double stranded RNA generally signals that a virus is active in the cell.

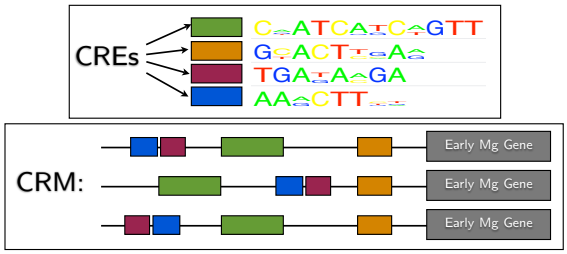
1.4.3 Controlling when and where the effector genes are expressed

Even with an effector gene that clears 100% of the pathogen 100% of the time, the intervention will not be successful in limiting transmission if the transgene is not turned on at the right time and place. If the transgene is most effective when the pathogen is in the midgut, it is not useful if the gene is only expressed in the antennae.

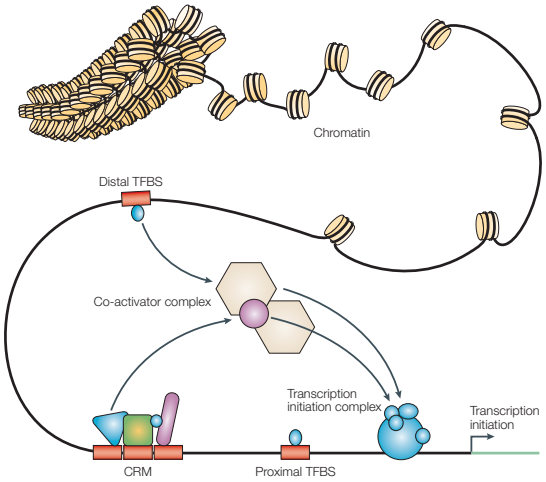
The region of DNA directly before the sequence of the gene is a major determinant of the pattern of expression and is referred to as the promoter. It influences when the gene will be turned on (e.g., directly following the ingestion of a bloodmeal), and in which tissue type (e.g., the midgut). Promoters influence the transcription of genes through the binding of special proteins called transcription factors that recognize specific DNA sequences called *cis*-regulatory elements (CREs). CREs are sometimes organized into modules that represent groups of binding sites for a specific combination of transcription factors. These modules are *cis*-regulatory modules (CRMs) CITEME(Figure 1.3A). Once a set of transcription factors are bound to the promoter they recruit the special machinery needed for the gene to be turned on or silenced depending on the context of the cellular environment (Figure 1.3B).

A typical strategy to obtain regulatory sequences suited to drive a transgene following the expression characteristics used as an example above (expressed directly following ,a bloodmeal, in the midgut) would be to identify genes in the mosquito that already have a similar expression pattern to the ideal. The promoter from that gene would then be copied and inserted in front of our transgene in the final integration construct.

This is a very common process used to engineer the expression patterns of real transgenes in mosquitoes. In [91], the promoter sequences of a gene called carboxypeptidase



(A) CREs make up CRMs



(B) One model of transcriptional regulation CITEME

Figure 1.3: Elements of transcriptional regulation

(normally expressed in the midgut after a meal) from *Ae. aegypti* and *An. gambiae* were inserted in front of a reporter gene.

From the abstract of [91]:

> Six independent transgenic lines were obtained with the AeCP construct and one with the AgCP construct. Luciferase mRNA and protein were abundantly expressed in the guts of transgenic mosquitoes in four of the six AeCP lines and in the AgCP line. Expression of the reporter gene was gut-specific and reached peak levels at about 24 h post-blood ingestion.

1.4.4 Achieving swift, dramatic, and clinically meaningful effects on the native vector population

In many ways, this is the most difficult part of the system. In order for the transgene to have its self-sustaining properties as well as achieve effective anti-transmission

results, it must spread through the native mosquito population to the point that the percent of individuals possessing the gene approaches 100%.

Making the optimistic assumption that the transgene carries a negligible fitness cost [3], or even a slight fitness advantage, achieving near 100% conversion could take decades. Funding terms for efforts like this in poor nations can be closer to 5 years or less. To enable the population conversion strategies to work, genetic "tricks" must be exploited that cause the gene to spread through a mosquito population **much** faster than could happen naturally. Efforts to discover or design these gene drive systems are an on going and active area in this field.

An example of such a method is the Medea system successfully engineered in *D. melanogaster* (CITEME).

In Brief:

1. The construct consists of a **death gene** under control of a maternal effect promoter and a **rescue gene** under control of an early embryonic promoter.
2. The **death gene** is produced by the mother and deposited into 100% of her developing embryos: *even those that did not inherit the transgene due to mating with a wild type male.*
3. However, since the **rescue gene** is produced by only those embryos that inherited the transgene construct, only those embryos will survive.
4. The female has on average half as many progeny survive as usual if she mated with a wild type male, but 100% of those that survive inherit the transgene.

[3] By fitness cost/advantage here, I mean that the transgene causes the mosquitoes that inherit it to be either less or more successful at producing offspring, respectively.

5. Over time this results in the fixation or near fixation of the transgene in the population.

1.4.5 Conservation of bloodmeal-regulated transcription

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

1.5 Evolutionary relationship between mosquitoes and hematophagy

Closely Related Lineages, Share Similar Sets of Protein Families in Their Sialomes, Even If No Apparent Orthologous or Conserved Functional Relationships Exist For example, while the D7-proteins in mosquitoes possess an extra alpha-helix involved in the binding of biogenic amines, this helix is absent from the D7-proteins from sand flies and biting midges, implying that they will not possess biogenic amine-binding function (Mans et al., 2007).

The **Same Protein Families are Prone to Expansion in Closely Related Lineages** The odorant-binding/D7-family expanded in mosquitoes, sand flies and bitingmidges (Anderson et al., 2006; Campbell et al., 2005; Francischetti et al., 2002a).

Excerpted from [82]

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

1.5.1 Implications for design of transgenic vector control systems?

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus.

Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Chapter 2

Review of Pertinent Publications

Our previous attempts to infer CREs bioinformatically have been relatively successful. Comparative genomics by way of the Motif Discovery using Orthologous Sequences (MDOS) program allowed us to infer a set of k-mers that should represent CREs conserved among three species of mosquitoes (*Ae. aegypti*, *An. gambiae*, *Cx. quinquefasciatus*). This method used no expression information in the process of predicting the CREs. However, we were able to show that a sub-set of the inferred CREs were enriched significantly in abundance profiles derived from a previous affymetrix-based bloodfeeding experiment in *An. gambiae* [81],[80],[115]. Comparing transcriptional abundance profiles between sugarfed and bloodfed (5h post bloodmeal (PBM)) *Ae. aegypti* mosquitoes allowed us to predict multiple putative CREs when we focused on transcripts that were detected at high levels in in bloodfed mosquitoes but undetectable in sugarfed females [20]. Possible CRMs were observed by visual inspection as well, but efforts to confirm biological meaningfulness have not yet been completed. Putative CRMs were again observed by visual inspection in an experiment comparing mRNA abundance profiles between *Ae. aegypti* fed with blood infected with dengue virus and those fed with non-infected blood. In this study, specific tissues were dis-

sected and analyzed separately providing a level of tissue specificity as well. These CRM results also remain in the process of being confirmed experimentally.

These studies have provided useful lessons that I have incorporated into the approach described in Chapter 3. A common theme among these lessons is the importance of the relative ratio of signal-to-noise contained in the data-set being examined. For example, in Bonizzoni et al. [20], the most successful CRE prediction results arose from a gene-set characterized by a fairly restrictive abundance profile definition (the requirement that transcripts be detected highly at 5h PBM but completely undetectable in sugarfed mosquitoes). This yielded a gene-set of just 40 members. Bonizzoni et al. [21] further demonstrates that results are sensitive to the definition of gene-sets used for analysis. In addition to abundance profile definition, Bonizzoni et al. [20] illustrated that, in *Ae. aegypti*, the official annotation is incomplete and contains at least some level of annotation errors that may be considered particularly troublesome to analyses that rely on accurate definition of putative promoter sequences; however, it also provides support that, despite containing errors, the current state of annotation is sufficient to allow CRE prediction. Finally, in experiments comparing mRNA abundance profiles between three different strains of a single species (*Ae. aegypti*), Bonizzoni et al. [22] illustrates the need to take precautions when comparing differences in transcript abundance profiles between species that are estimated to have last shared a common ancestor as long as 200 million years ago (Mya).

2.1 Comparative genomics allows the discovery of *cis*-regulatory elements in mosquitoes. [115]

The role CREs play in regulating gene expression during development is well-established [33], and the development of tools for their identification is an active area of research following the publication of genome sequence and associated genome-wide expression datasets. However, the discovery *in silico* of CREs is challenging because typically they are short, degenerate, and contained within vast amounts of intergenic genomic DNA. Despite these limitations, various computational approaches have been developed for their discovery [32, 57, 123, 130]. Comparative genomics represents a powerful extension to CRE discovery that diminishes these effects. Functional gene regulatory elements, including CREs, are proposed to diverge at much lower rates compared to neutral sequences because of selective pressures, and therefore may stand out from surrounding neutral DNA by virtue of their greater levels of conservation among orthologous sequences. Previous work has demonstrated the utility of this concept [38, 118, 138] and comparative genomics of insects has been applied successfully to map putative CREs in the genomes of relatively closely related *Drosophila* species, [divergence times estimated $\leq$ 50 Mya [118] (Figure 2.1)], in which orthologous intergenic sequences are aligned more easily. The work represented by Sieglaff et al. [115] includes mosquito species with phylogenetic relationships spanning 20 to 200 Mya with genomic comparisons complicated by large amounts of dispersed repetitive elements [92].

The MDOS algorithm [136] was designed to mitigate these problems by not requiring alignment of orthologous sequences, and incorporating features that account for the greater probability for the co-occurrence of a motif because of shared ancestry in orthologous sequences. The application of MDOS for computational discovery of puta-

tive CREs shared among mosquitoes resulted in the identification of mosquito-specific or enriched motifs. These motifs demonstrate that MDOS can identify related DNA sequences in diverged mosquitoes and may be useful in similar scenarios in which genome sequences are being analyzed for species that have no other closely related genomes available. Furthermore, the identification *in silico* of GATA-factor and other transcription factor binding sites (TFBSs) that are identical to those whose function was established experimentally [66, 28, 10, 4, 97, 45, 36, 87] serves as a powerful positive control for these analyses. The 5′-end regions of immune-related genes from *An. gambiae* and *D. melanogaster* were screened for conserved motifs, and AT-rich domains were found to associate with Nuclear Factor κ-light-chain-enhancer of activated B cells (NFκB) response elements [55]. These sequences were not identified in these analyses because the restrictions on the dataset used eliminated many of the immune-related genes where one-to-one orthology may be difficult to establish. In addition, most AT-rich domains discovered in their study were ≤ 6 nucleotides in length, and these analyses only addressed those recovered with lengths of 7 to 9 nucleotides.

The 3 mosquito species included in Sieglaff et al. [115] require a bloodmeal for successful reproduction. Although this trait is not unique to this group of arthropods, acquiring the necessary nutrients for egg development from ingested blood is a specialized adaptation in insects. There is debate on when and how hematophagy arose in mosquito evolution [101], but it is hypothesized that the occurrence of this trait in the larger clade is monophyletic [23, 26]. Most of the discovered mosquito motifs are associated with bloodmeal-regulated genes. Thus, these data support a common hematophagous ancestor for all mosquitoes, and indicate that hematophagy acts as a selective force for conservation of CREs.

The functionality of some of the discovered motifs was supported further in *An. gambiae* and *Ae. aegypti* by their association with genes displaying enriched transcrip-

tion product accumulation within tissues responsible for bloodmeal digestion and reproduction (Figure 2.2). These findings bolster the conclusion that these mosquitoes share a regulatory code controlling expression levels for some genes regulated after hematophagy. The availability of additional genome-wide studies of gene expression for all 3 mosquito species will facilitate discovery of other associations of the motifs with specific gene-regulation patterns.

Sieglaff et al. [115] and similar genome-wide approaches to identify putative CREs in mosquitoes [55] are furthering our understanding of the mechanisms involved in gene regulation in this group of vector insects. An expanded set of putative mosquito CREs will allow the definition of genome-wide motif-association maps and the identification of CRMs comprising multiple, linked CREs that convey specific patterns of gene expression. Experimental validation of the functionality of each discovered motif and regulatory module is necessary and will provide support for the development of mosquito synthetic promoters that deliver desired and predetermined expression patterns in transgenic mosquitoes. The availability of defined synthetic mosquito promoters that direct controlled, local gene expression in response to pathogens also would be a major advance. These promoters would allow engineering of mosquitoes with increased parasite or virus resistance. These and similar envisioned applications for mosquito control and the control of mosquito-borne disease transmission will benefit greatly from a better understanding of gene regulation mechanisms in these insects.

2.2 RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti* [20]

Bonizzoni and Dunn [20] provides a detailed examination of the changes in transcripts accumulation occurring at the whole-body level of *Ae. aegypti* females 5 hours PBM. The observed changes are consistent with the beginning of an intense physiological response to a bloodmeal. The majority of immunity-related transcripts tended to accumulate at lower levels in bloodfed mosquitoes. This finding supports the hypothesis that there may be a gap in immunity following a bloodmeal. Reduced expression of immune genes in bloodfed mosquitoes could favor the establishment of infections, especially considering that pathogens such as dengue viruses infect the midgut epithelial cells within minutes after the contact [108]. However, changes in transcript abundance observed at the whole-body level may mask changes in accumulation occurring primarily in the midgut. Different levels of activation of immunity genes after a blood feeding may be one of the factors contributing to the variability in vector competence for dengue viruses observed in different geographic populations of *Ae. aegypti* [16],[18]. The quantity and quality of data generated by RNA-seq technology makes it an ideal approach for comparative analyses of the transcriptome of *Ae. aegypti* strains with different vector competence and vectorial capacity.

Our analyses of the expression profiles of sugarfed (S) and bloodfed (B) mosquitoes allowed the identification of co-regulated genes and putative *cis*-regulatory elements and modules from the *Ae. aegypti* genome. Further knowledge of the mechanisms involved in regulation of gene expression in vector species is critical to the development of control strategies whereby the vector is modified genetically to express anti-pathogen effector molecules in tissue-specific and time-regulated manners [121]. Promoter and

other *cis*-acting regulatory DNA fragments are needed to regulate restricted expression of selected anti-pathogen effector molecules. Moreover, we described several examples of how the RNA-seq data generated can help improve the current annotation of the *Ae. aegypti* genome.

2.2.1 Transcripts found exclusively in bloodfed mosquitoes

Forty transcripts were found only in bloodfed mosquitoes, with the highest read-counts reaching 1000/transcript, after normalizing for different library sizes (data not shown). Functional parent attribution for these transcripts is consistent with a role in digestion and in the progression of the gonotrophic cycle. Specifically, two transcripts, Aa5G1 (AAE013712-RA) and AaSPVI (AAE010196-RA), correspond to the midgut serine proteases shown previously to be elicited by a bloodmeal in the midgut of *Ae. aegypti* females [24]. Seven other transcripts encode enzymes (i.e. decarboxylase, cathepsin b and trypsins), and two are implicated in trafficking. Transcripts AAE014815-RA and AAE005950-RB correspond to the vacuolar protein sorting 13B from yeast and the chloride channel protein 2, respectively. Ten transcripts are orthologous to the G12 gene of *An. gambiae* and share the Insect Allergen Repeat motif. This motif is hypothesized to be a novel, insect-specific detoxifying domain implicated in the co-evolution of herbivorous insects and their plant hosts and also has been linked to nitrile-specific detoxification [41]. Transcripts AAEL006126-RB and AAEL008921-RC are predicted orthologues of the *Culex quinquefasciatus* vitellogenin-A1 gene and the *Drosophila melanogaster* spaghetti squash (sqh) gene, respectively. The sqh gene product encodes the regulatory light-chain of non-muscle myosin II, which is required for cytoplasmic transport in nurse cells during oogenesis and also has been implicated in germline RNA interference (RNAi) processes [96].

2.2.2 *Cis*-regulatory element discovery

Tightly regulated and bloodmeal-induced expression profiles are of particular interest for designing transgenic mosquito-based control strategies to reduce transmission of dengue fever. *Cis*-regulatory sequences derived from bloodmeal-induced / up-regulated mosquito genes allow potentiating swift induction and effective levels of transcription of an associated effector gene, while likely inflicting the least fitness cost [83], [5]. We interpret the different levels of mRNA accumulation seen in this study to reflect changes in transcriptional activity of the corresponding genes, although it is possible that some levels may vary as a function of changing transcript stability or rates of turnover. With this in mind, we used Suite for Computational identification Of Promoter Elements (SCOPE) [27] to predict putative CREs that may provide the basis for rational identification and selection of new candidate promoter regions and for modification of the transcriptional profiles of current transgene constructs. We examined the 2000 base pairs (bp) flanking the 5'-boundaries of the 40 transcripts that were undetected in libraries from sugarfed mosquitoes but detected at significant levels in the RNA-seq libraries from bloodfed mosquitoes and identified a redundant list of 22 motifs that are enriched significantly in these sequences (Figure 2.3). A possible CRM constructed with the discovered CREs is represented by the motif consensus sequences, `cnatcnkcwgtt`, `gyactyvar`, and `tgakamga`, and is associated with *Ae. aegypti* orthologs of the G12 gene of *An. gambiae* (AGAP006187) (Figure 2.3). *Ae. aegypti* has 17 G12 genes, many more relative to other insects, which have 4.5 on average (according to OrthoDB; group EOG95TCTG) [68]. The transcripts of nine of the G12 orthologs are present in this co-regulated gene set (representing 25% of the 40).

Another putative CRM contains the consensus sequence `tgakamga`, `cnatcnkcwgtt`, `asttrccc` and `aarcttbd` (Figure 2.3). This CRM groups with the cathepsin b genes, AAEL015312-RA and AAEL007585-RA. Verification of these CRMs will require empirical testing,

however, the top 10 matches for `tgakamga`, which is present in both putative CRMs, align well to members of the mosquito-conserved GATA motifs correlated to transcriptional responses to blood feeding in *An. gambiae* [115].

2.2.3 RNA-seq identifies annotation corrections

Distance from annotation (bp)	Predicted junctions
0-1000	1170
1001-2000	120
2001-4000	149
4001-8000	245
8001-16000	486
16001-32000	517
> 32000	854
Total	3541

Table 2.1: Predicted novel junctions within various symmetric distance-windows from annotated transcripts:
(Gene build AaegL1.2)
Excerpted from [20]

RNA-seq also provides an opportunity to examine and improve the current annotation of the *Ae. aegypti* genome and examine the level of transcriptome plasticity in terms of alternative splicing. We used HMMSplicer [35] to compare junctions revealed by our data to the annotation provided by Vectorbase and Ensembl [72, 58]. HMMSplicer predicted 32,501 junctions supported by at least two RNA-seq reads using the combined data from sugar and bloodfed samples. Of these, 24,100 (74%) matched junctions present in the AaegL1.2 gene-build provided by VectorBase, leaving 8,401 predicted novel high-scoring splice sites supported by multiple RNA-seq reads. A total of 4500 (54%) of these occur within annotated gene boundaries and may represent unannotated alternatively-spliced transcripts. To estimate how many of the remaining splice junctions might be truly novel, we mapped them to increasingly larger DNA frag-

ments flanking the currently-annotated genes (Table 2.1). A total of 2687 (33%) junctions mapped within 32,000 bp of the 5'- or 3'-ends of annotated gene boundaries. Of these, 1439 mapped within 4000 bp, consistent with the interpretation that they may represent alternatively-spliced transcripts of the previously-identified genes. Those mapping beyond 4000 bp could be alternate junctions of the known genes, represent un-annotated transcription products, or be artifacts.

An accurate gene annotation, especially with respect to the transcriptional start sites (TSS), is considerably advantageous with respect to the accurate discovery of CREs because prediction tools must make the assumption that the sequences included are true regulatory regions, and their performance suffers when this is false. For the CRE predictions described in the previous section, 36 of the 40 transcript start sites were in close agreement to the Ensembl annotation [58]. Figure 2.4 highlights three determined amendments to the current annotation, all supported by EST data. Figure 2.4 (A and B) supports the conclusion that the current annotation has missed the putative first exons that extend the 5'-UTRs of some genes (AAEL006259, AAEL010818) and provides additional information for predicting accurate TSSs. In the case of AAEL010818, the TSS determined by RNA-seq data is 20 kb to the 5'-end of the annotated start site, far outside the distances commonly searched for CREs (Figure 2.4 B). In some cases, as was seen for AAEL001774, the first exon was annotated but included as a separate gene model, which also contains the likely 5'-UTR of AAEL001759 (Figure 2.4 C). AAEL001774 encodes a protein comprising 50 amino acids with no known functional domains aside from a predicted signal peptide that makes up 66% of its length.

2.3 Complex modulation of the *Aedes aegypti* transcriptome in response to dengue virus infection. [21]

The World Health Organization lists dengue as the most important arthropod-borne viral disease of humans [95]. The major vector of all four dengue virus serotypes (dengue virus (DENV)1–4) is the cosmopolitan mosquito, *Aedes aegypti*. Close association with human populations and increasing intercontinental travel favor the expansion of its geographic distribution. There are no effective prophylactic and therapeutic drugs specific for dengue, and vaccine development is hindered by potential antibody-dependent enhancement that could put people at greater risk of life-threatening, severe dengue [46]. As a consequence, vector control is currently the only practical and effective strategy for disease prevention. The *Ae. aegypti* genome was sequenced and knowledge of genome-wide changes in patterns of gene expression following DENV infection is expected to identify genes involved in vector competence, the intrinsic ability of the mosquito to host and transmit DENV [92], [67]. This knowledge, coupled with germline transformation technology and anti-viral effector molecules, can be applied to the development of genetically-modified mosquitoes incapable of arbovirus transmission [61], [43], [84].

Although vertical transmission of dengue viruses has been reported, mosquitoes become infected mainly following ingestion of an infectious-bloodmeal [49], [8]. Viruses are transmitted to new human hosts during a subsequent bloodmeal following an extrinsic incubation period (EIP) of 7–14 days. The duration of the EIP depends on the mosquito strain, virus genotype and environmental factors [133], [18], [7], [108], [70]. During the first 1–2 days post infection (dpi), DENVs invade midgut epithelial cells through receptor-mediated endocytosis and initiate replication [16], [54], [103], [86].

These processes involve both viral and host cellular factors [109]. Infection spreads laterally in the midgut epithelium to cells adjacent to those infected originally [108]. Virus titers peak in the midgut usually between 7–10 dpi and are followed by a decline [108], [137]. DENV infection disseminates from the midgut throughout the body, presumably through the tracheal system, reaching the salivary glands as early as 3 dpi [108]. Maximum virus titers in the salivary glands are reached 12–18 dpi. The saliva of an infected mosquito containing DENVs is injected in a human host during feeding to complete the transmission cycle.

Aedes aegypti populations of different geographic origin vary in their vector competence. Phenotypes include those in which DENVs either cannot establish a midgut infection (Midgut Infection Barrier) or cannot disseminate to other tissues (Midgut Escape Barrier) [18], [48], [31]. Other phenotypes manifest in the absence of virions in the saliva (Transmission Barrier). Differences in intensity of infection (peak titers) and duration of EIP also are observed among mosquito strains [108].

Multiple quantitative trait loci are associated with MIB and MEB, but specific genes have yet to be identified [18]. Mosquito genes and physiological pathways related to innate immunity, redox activity, energy production and metabolism are modulated in response to DENV infection [137], [110], [116], [120], [76], [14], [117], [30]. These observations come from multiple studies of specific mosquito tissues and time points following infection, and used different combinations of DENV2 genotypes and *Ae. aegypti* strains.

We investigated genome-wide changes in transcript accumulation in mosquito midguts, carcasses and salivary glands at 1, 4 and 14 dpi during the course of DENV2 infection. This analysis was conducted by RNA-seq with the Chetumal (CTM) strain of *Ae. aegypti* and DENV2-Jam1409. Chetumal (CTM) was colonized recently (2005) from mosquitoes from the Yucatan Peninsula and is well-characterized for its response

to non-infectious bloodmeals and for the kinetics of DENV2-Jam1409 infection [108], [17], [22]. A total of 397 genes had transcripts that showed statistically-significant differential accumulation following DENV infection, comprising both those found previously and those that are novel to this study, emphasizing the complex interaction between *Ae. aegypti* and DENVs [137], [116], [120], [76], [14], [117], [30].

2.3.1 Results

2.3.1.1 RNA Sequencing and Mapping Summary

Illumina RNA-seq technology was applied to study the accumulation levels of poly-adenylated RNAs at 1, 4 and 14 dpi in the carcasses and corresponding midguts of CTM females fed either a non-infectious (B) or DENV2-infected (dengue virus infected (DENVI)) bloodmeal. Accumulation levels also were assessed in the salivary glands at 14 dpi. Single-end RNA-seq libraries were constructed starting from pools of 20–40 mosquitoes. Each RNA-seq library generated between 14 and 45 million 40 bp reads (not shown). Sequenced reads were mapped by TopHat [124] to the *Ae. aegypti* transcriptome. The accumulation levels of specific poly-adenylated RNAs were compared between B and DENVI samples at each time point/conditions using DESeq [6] and Cufflinks [126] (not shown). Genes whose products were identified as significantly differentially accumulated by DESeq are contained mostly within the larger number designated similarly by Cufflinks (not shown). A total of 397 unique genes have poly-adenylated RNAs identified by both DESseq and Cufflinks as accumulated differentially and significantly between B and DENVI mosquitoes

2.3.1.2 Discovery of cis-regulatory Elements

Genes with similar expression profiles may share common regulatory mechanisms, including the occurrence of similar CREs [115]. Evidence of this shared regulation may be evident in the control DNA sequences of mosquito genes that are expressed exclusively or highly-induced following dengue virus infection. Some of these genes also may have transcripts that accumulate to high or higher levels following an uninfected bloodmeal (B samples). The preferred expression profile for an antiviral effector gene would be one that is induced highly after blood feeding and is either further elevated or not affected by virus infection.

A total of 2012 genes showed read-coverage in salivary glands of DENVI mosquitoes but not in the salivary glands of B mosquitoes. While fifteen of these genes were associated with lipid metabolism, the majority were related to transcription and translation and/or chromatin structure and dynamics (not shown). Forty-one of these had $\text{FPKM}_{DENVI} \geq 15$, and only one, AAEL005034, showed tissue-specific expression (not shown). As many as 762 and 1324 genes had transcripts that were detected in infected but not in the corresponding uninfected carcasses and midguts samples, respectively. The vast majority of these had $\text{FPKM}_{DENVI} < 4$. The exceptions were AAEL011066, AAEL010034 and AAEL002899, all encoding hypothetical proteins. At 1dpi, AAEL011066 had read coverage only in carcass of DENVI ($\text{FPKM}_{DENVI} = 14.97$); AAEL010034 and AAEL002899 in midguts of DENVI ($\text{FPKM}_{DENVI} = 8.11$ and 6.56, respectively) (not shown). The putative promoter regions of 41 genes with $\text{FPKM}_{DENVI} \geq 15$ in salivary glands were analyzed using MEME [12] (Figure **??**).

We refined the search for common, co-occurring CREs by considering the promoters of genes whose products accumulated highly ($\text{FPKM}_{DENVI} \geq 100$) in DENVI mosquitoes and were at the same or lower levels ($\text{FPKM}_{B} \leq 100$) in B mosquitoes (not shown).

The latter criterion eliminates all genes whose accumulation levels are lowered in the presence of a dengue infection. In addition, this grouping also contains genes whose transcripts do not have a significant differential accumulation between infected and uninfected mosquitoes. The results support the presence of a module composed of six motifs (Figure **??**: motifs 2, 4, 5, 6, 8, 10) in 4/51 genes with FPKM_{DENVI} > 100 in midguts at 1 and 4 dpi. The four genes sharing this module (AAEL007162 [APG8], AAEL001593 [glycerol-3-phosphate dehydrogenase (G3PDH)], AAEL003046 [saponin], AAEL000291 [V-type proton ATPase 16 kD proteolipid subunit (V-ATPase)]) had transcripts that were more abundant (1.03–1.64 fold) in DENVI than B mosquitoes in midguts samples at 1 and 4 dpi. These genes are not related functionally, but are associated with lipid metabolism, which has been shown to play an important role during DENV life cycle [109]. Furthermore, G3PDH links carbohydrate and lipid metabolism. The V-ATPase may be required to maintain transmembrane charge differential for virus entry. Saponin is a regulator of lipid degrading enzymes [74]. These genes, with the exception of that encoding G3PDH, had transcripts with FPKM_{DENVI} > 100 also in carcasses and salivary glands, but were not consistently higher or equally-abundant in DENVI than B mosquitoes. Transcripts encoding the V-ATPase were more abundant in DENVI mosquitoes of the susceptible MOYO-S strain three hours post infection (hpi) with DENV2 Jam1409 [14]. Matches to transcription factors in the TRASFAC- database [85] and supported by an e-value $\leq 10^{-5}$ were identified for motifs 4, 5, 6 and 10. The highest match of Motif 4 is Blimp1, which is an ecdysone-inducible transcription factor involved in Drosophila development, metamorphosis and oogenesis [3]. Motif 5 matched highest with Rel. The Rel or NFκB superfamily of conserved eukaryotic proteins is involved in the control of immune and inflammatory responses, developmental processes, cellular growth and apoptosis. Motif 6 has high matches with High Mobility Group (HMG), Broad-Complex (BRC) and Forkhead transcription factors. HMG transcription factors are implicated in replication, recombination and

DNA repair [102]. BRC is an ecdysone-regulated transcription factor implicated in developmental processes in *D. melanogaster* [111]. The Forkhead family of transcription factors includes seventeen sub-classes regulating development, homeostasis and reproduction in insects [53]. Motif 10 matches best with the NK-2-Nkx_TTF1 transcription factor, a homeodomain-containing transcription factor implicated in morphogenesis, differentiation and tissue-specific maintenance [19].

MEME analyses on the putative promoter regions of the 83 genes that had FPKM$_{DENVI}$ $\geq$ 100 in carcasses consistently from 1 to 14 dpi showed the presence of three to five motifs (Figure **??**: motifs 1, 2, 5, 7, 8) grouped tightly in eight genes (AAEL010097 [hypothetical protein], AAEL012629 [deoxyuridine 5′-triphosphate nucleotidohydrolase], AAEL008041 [bleomycin hydrolase], AAEL013068 [protein phsophatase-2a], AAEL017269 [novel protein coding], AAEL009604 [hypothetical protein], AAEL003552 [DNA-directed RNA polymerase subunit rpb6] and AAEL000739 [hypothetical protein]). Different combinations of these motifs are found in seven additional genes: (AAEL008849 [selenophosphate synthase], AAEL014903 [40S ribosomal protein S24], AAEL004484 [predicted protein], AAEL008768 [multiprotein bridging factor], AAEL009274 [hypothetical protein], AAEL001061 [glutathione transferase], AAEL012279 [eIF3j] and AAEL009320 [chaperonin]). This motif group is designated provisionally the "carcass module". All 15 of these genes also had read coverage in midguts and salivary glands, but the abundance of the corresponding transcripts was not always as high (FPKM$_{DENVI}$ = $\leq$100), and in some cases they are more abundant in B than DENVI mosquitoes. AAEL017269 and AAEL000739 had read coverage only in salivary gland samples of DENVI mosquitoes.

Motif 1 could be the binding site for the transcription factor MADS_MCM1+SFF_M01051 (e-value = $7.36E-06$). The MADS-box family of transcription factors is conserved among yeasts, plants, insects, amphibians and mammals. It includes proteins asso-

ciated with different biological roles (pheromone response, muscle-specific gene regulation, development) that operate generally by specifically recruiting other transcription factors into multi-component regulatory complexes [114]. Motif 2 has its best match to bZIP-type transcription factors (e-value = $2.27E-06$). bZIP proteins belong to the largest and most conserved superfamily of transcription factors, the basic region leucine zipper transcription factors, involved in regulation of development, metabolism, and other cell functions such as secretion, oxidative stress and response to pathogens [89], [1], [50]. Motifs 5 and 7 match the *Arabidopsis thaliana* transcription factor trp_AtMYB-84_M00970 (e-value = $1.35E-10$ – $7.08E-10$) and the Ras responsive element binding protein-1 (RREB-1) (e-value = $5.09E-9$ – $1.38E-6$), respectively, the latter of which regulates immunity and cancer-related gene expression in humans [42], [75].

A total of 94 genes had FPKM $\geq$ 100 in DENVI mosquitoes in both carcass and salivary gland samples at 14 dpi, with corresponding values in B mosquitoes $\leq$ 100. The putative promoter regions of these genes analyzed by MEME revealed the presence of two groups of motifs (Figure **??**: motifs 1, 2, 3, 4, 5, 7, 9, 10) co-occurring or alternating in 14 genes associated with diverse functions: AAEL001759 [40S ribosomal protein S9], AAEL005069 [ras-related protein Rab-1A], AAEL002372 [40S ribosomal protein S11], AAEL005165 [chaperone protein DNAj], AAEL000657 [hypothetical protein], AAEL005471 [Sec61 protein complex gamma subunit], AAEL001432 [protein disulfide isomerase], AAEL010059 [bacterial-type ABC transport ATP-binding subunit or RNAse l inhibitor], AAEL013407 [Catalase], AAEL007945 [eIF3 h], AAEL010169 [hypothetical protein], AAEL012827 [endoplasmin], AAEL011773 [calreticulin], AAEL010777 [TRX, putative]. These genes, except for AAEL002372, had high read coverage also in midgut samples, but not necessarily higher or equal accumulation in DENVI versus B mosquitoes. Some of these genes were detected previously as significantly differentially accumulated in the salivary glands of mosquitoes of the Liverpool (LVP) strain

infected with DENV2 Thailand 16681 (Figure 2.5; [76]). Specifically, the transcription products of AAEL001432, AAEL010059, AAEL007945 and AAEL011773 were more abundant in salivary glands of infected LVP, while the opposite was observed for AAEL001759, AAEL005069, AAEL002372 and AAEL012827, which were more abundant in uninfected samples. AAEL00657 was listed as both up- and down-regulated following DENV infection, most likely at different time points, but this is not explicit in the report [76]. MADS_MCM1+SFF_M01051 and bZIP transcription factors were the best matches to motifs 1, 2, 5 and 7 (e-value $\leq 10^{-5}$). Motif 3 has sequence similarity with motifs 5 and 7 of the previously designated "carcass module" and matches to trp_AtMYB-84_M00970 (e-value = $2.25E-10$) and RREB-1 (e-value = $2.16E-7$). The best match of motif 9 is to the HMG transcription factor ($1.87E-5$).

2.3.2 Discussion

The dengue-susceptible strain CTM is well-documented in its ability to disseminate quickly viruses to peripheral tissues when compared to long-established laboratory strains [108]. We describe in Bonizzoni et al. [21] changes in transcript accumulation levels in CTM during the course of a Jam 1409 DENV2 infection. Our analyses include multiple post-infection time-points and three tissues important for the viral transmission cycle: midguts, salivary glands and carcasses, the latter of which includes the fat body. The RNA-seq technology used in this study interrogates the whole *Ae. aegypti* transcriptome and allows for absolute quantification of poly-adenylated RNA levels. In general, more depleted rather than enriched transcript accumulation was observed. Importantly, hundreds of genes had exclusive read-coverage only in DENVI mosquitoes, but their accumulation levels were generally low ($FPKM_{DENVI} < 4$ in carcass and midgut samples and < 15 in salivary gland samples).

The "DENV down-regulation" trend is consistent to that observed with the Rockefeller (ROCK) strain infected by DENV2 New Guinea in transcriptional responses examined in whole-body (1–2, 7 and 14 dpi) or midgut samples (10 dpi) and in salivary glands of the LVP strain infected with DENV2 Thailand 16681 [137], [76], [117], [30]. These results support the hypothesis that dengue infection has a negative impact on fitness. However, a larger number of genes had products that increased instead of decreased in abundance in carcass samples 10 dpi and in salivary glands at 14 dpi in ROCK mosquitoes infected with DENV2-New Guinea [137], [117]. This trend also was evident in *Ae. aegypti* Aag2 cultured cells infected by DENV2 New Guinea and in both susceptible (Moyo-S) and resistant (Moyo-R) strains infected with DENV2 Jam1409 [116], [14]. These results emphasize the complex relationship among *Ae. aegypti* and dengue viruses.

2.3.2.1 *Cis*-regulatory Elements

The preferred promoter for an anti-dengue effector gene would be one that is induced following a bloodmeal and is not affected by viral infection. Furthermore, we hypothesize that genes with this preferred expression profile may share common regulatory mechanisms based on common CREs. Analysis of the 5′-end of genes whose transcripts exhibited accumulation profiles exclusively in salivary gland following DENV infection did not reveal substantial co-occurrence of CREs. The regulation of these genes is likely not coordinated through common promoter elements. In contrast, co-occurring putative CREs were identified for genes whose transcripts were represented and modulated highly in response to dengue infection in midguts (1–4 dpi), carcasses (all time points studied) and salivary glands and carcasses (14 dpi).

Several of the identified motifs showed high-quality matches to known transcription factors. The CREs likely are not related specifically to dengue infection since the ma-

jority of the genes included in the analyses also had read coverage in B mosquitoes. This observation does not affect the potential utility of the CREs as components of synthetic promoters to direct expression of anti-dengue effector molecules. Indeed, driving effector gene expression with control DNA that responds to an uninfected bloodmeal and either is unaffected or is enhanced by an infected bloodmeal could elicit a protective antiviral response in the mosquitoes. Remarkably, 10 genes whose transcripts were modulated following DENV2 16881 infection in LVP salivary gland samples were among the 14 genes with $\text{FPKM}_{DENVI} \geq 100$ in our carcass and salivary gland samples that have a common group of CRE motifs [76]. However, not all of these ten genes had transcripts that were consistently more abundant following infection, and one gene, (AAEL00657), had transcripts reported as more and less abundant [76]. These results and the findings that several motifs are putative binding sites for transcription factors acting through repressors and activators support a complex model of transcription modulation that requires further investigation.

2.3.3 Conclusions

Dengue fever is the most important arboviral disease world-wide, with *Aedes aegypti* being the major vector. Interactions between the mosquito host and dengue viruses (DENV) are complex and vector competence varies among geographically-distinct *Ae. aegypti* populations. Additionally, dengue is caused by four antigenically-distinct viral serotypes (DENV1–4), each with multiple genotypes. Each virus genotype interacts differently with vertebrate and invertebrate hosts. Analyses of alterations in mosquito transcriptional profiles during DENV infection are expected to provide the basis for identifying networks of genes involved in responses to viruses and contribute to the molecular-genetic understanding of vector competence. In addition, this knowledge is anticipated to support the development of novel disease-control strategies. RNA-

seq technology was used to assess genome-wide changes in transcript abundance at 1, 4 and 14 days following DENV2 infection in carcasses, midguts and salivary glands of the *Ae. aegypti* Chetumal strain. DENV2 affected the expression of 397 *Ae. aegypti* genes, most of which were down-regulated by viral infection. Differential accumulation of transcripts was mainly tissue- and time-specific. Comparisons of our data with other published reports reveal conservation of functional classes, but limited concordance of specific mosquito genes responsive to DENV2 infection. These results indicate the necessity of additional studies of mosquito-DENV interactions, specifically those focused on recently-derived mosquito strains with multiple dengue virus serotypes and genotypes.

The study of genome-wide changes in transcript abundance in *Ae. aegypti* following dengue infection is anticipated to identify genes and control DNA elements involved in vector competence. This knowledge is expected to contribute to development of novel vector control strategies. The interactions between the mosquito host and the virus are complex at both the individual and population levels. Many mosquito tissues are affected by the virus during the course of the infection and *Ae. aegypti* strains can show unique responses of the transcriptome in response to a bloodmeal and susceptibility to DENV infection [18], [22]. Additionally, the variation in DENV serotypes and genotypes contributes to genetically-distinct differences in vector competence [7], [134]. While we identified modulations in transcript abundance following DENV infection that represent genes encoding similar functional categories, a limited number of specific genes were concordant across the mosquito strains and DENV2 genotypes following comparison of our data with previous studies [137], [76], [14], [117], [30]. Furthermore, the direction of changes in transcript abundance in samples from infected and uninfected mosquitoes was not always conserved among the various studies, an observation that we consider independent of the methodology used to assess differential transcript accumulation. These results support the need for a comprehen-

sive analysis of dengue infection focusing on recently-colonized laboratory strains or wild-caught mosquitoes to capture most of the genetic variability at the host level and different DENV serotypes/genotypes [9]. This analysis also should account for possible variation in the progression of the viral infection and timing of host gene expression.

2.4 Strain variation in the transcriptome of the dengue fever vector, *Aedes aegypti*. [22]

Studies of transcriptome dynamics provide a basis for understanding functional elements of the genome and the complexity of gene regulation. The dengue vector mosquito, *Aedes aegypti*, exhibits great adaptability to diverse ecological conditions, is phenotypically polymorphic, and shows variation in vectorial capacity to arboviruses. Previous genome sequencing showed richness in repetitive DNA and transposable elements that can contribute to genome plasticity. Population genetic studies revealed a varying degree of worldwide genetic polymorphism. However, the extent of functional genetic polymorphism across strains is unknown. The transcriptomes of three *Ae. aegypti* strains, CTM, Rexville D Puerto Rico (Rex-D) and LVP, were compared. CTM is more susceptible than Rex-D to infection by dengue virus serotype 2. A total of 4188 transcripts exhibit either no or small variation (< 2-fold) among sugarfed samples of the three strains and between sugarfed and bloodfed samples within each strain, corresponding most likely to genes encoding products necessary for vital functions. Transcripts enriched in bloodfed mosquitoes encode proteins associated with catalytic activities, molecular transport, metabolism of lipids, carbohydrates and amino acids, and functions related to blood digestion and the progression of the gonotropic cycle. Significant qualitative and quantitative differences were found in individual transcripts among strains including differential representation of paralogous gene products. The majority of immunity-associated transcripts decreased in accumulation after a bloodmeal and the results are discussed in relation to the different susceptibility of CTM and Rex-D mosquitoes to dengue virus serotype 2 (DENV2) infection.

This mosquito species is phenotypically polymorphic, has great adaptability to diverse ecological conditions, and shows variation in vectorial capacity for arboviruses [16,

18, 69]. Genetic polymorphism among geographically distinct *Ae. aegypti* populations is documented [127]; however, the extent of genome sequence polymorphism and its effects on transcriptional activity are not known.

The transcriptional profiles of three *Ae. aegypti* strains, LVP, CTM, Rex-D, were investigated. LVP originated in West Africa in the 1930s, and its genome is sequenced [92], CTM was derived from the Yucatan Peninsula in Mexico in the early 2000s [16, 47, 105], whereas Rex-D was established from mosquitoes captured in Puerto Rico in the early 1990s [88]. CTM supports a faster and more intense dissemination of DENV2 than Rex-D [16, 108]. Our data show significant differences in transcript accumulation among strains, and the results may account for the differing susceptibility of CTM and Rex-D mosquitoes to virus infection.

2.4.1 Results

2.4.1.1 RNA-seq mapping summary

Transcriptional profiles of LVP, CTM, and Rex-D strains before and after a bloodmeal were compared. RNA-seq libraries were generated with the use of RNA extracted from females collected 3 to 5 days after eclosion and kept either on a sugar diet (S) or harvested 5 h PBM (B) (not shown). Two RNA-seq libraries were constructed for each RNA sample. The reproducibility of the parallel libraries was verified (not shown), and the data from the two technical replicates were merged for further analyses. Reads mapping to multiple genomic locations were discarded during Bowtie alignment [71]. As a consequence, the accumulation levels of transcripts encoded by highly conserved gene families are under-estimated. This bias is expected to affect all samples equally; there-

fore, if the absolute accumulation level of a transcript is underestimated, differences across conditions still can be calculated.

2.4.1.2 Constitutively accumulated transcripts

A total of 4188 transcripts exhibit no or small variation ($\leq$ 2-fold) among S samples of the three strains and between B and S samples within each strain and correspond most likely to genes that encode products necessary for vital and/or basal metabolic functions (not shown). Function−parent attributions for this group of transcripts revealed a predominance of those encoding proteins with binding activity over transcripts with molecular function and catalytic activity (not shown). Functional description is available for 2262 of these 4188 transcripts [72] (not shown). More than 10 transcripts were found associated with each of the following classes: protein recycling (ubiquitin and proteasome), zinc finger proteins, WD-repeat proteins associated with diverse functions (i.e. RNA-processing complexes, transcription, cytoskeleton assembly), ribosome proteins, signal transduction (GTP binding proteins, G coupled proteins), redox processes (cytochrome), translation (eukaryotic translation initiation factor), transcription (transcription factors, mediator complex), membrane trafficking (rab), cell growth, differentiation and survival (ras), and immunity (PIWIs, autophagy proteins, galectins, TOLL, IMD, and JACK-STAT pathway signaling members).

Approximately 25% of all analyzed transcripts accumulated differentially in S mosquitoes of different strains. However, in the majority of cases the differences observed were lower than 2-fold (not shown). Here is described only differences between the DEN-refractory Rex-D and the more permissive CTM strain. Totals of 2992 and 2080 transcripts accumulated more in Rex-D or CTM, respectively. Six were accumulated more than 5-fold in CTM with respect to Rex-D, including four of unknown function, one (AAEL005534-RA) annotated as a gonadotropin-inducible transcription factor, and

one (AAEL014188-RA) annotated as a serine-type endopeptidase. In contrast, 310 transcripts accumulated $\geq$ 5-fold in Rex-D with respect to CTM. Functions could be attributed to 51% of these and they include aldehyde oxidases, cytochrome P450s, G-coupled receptors, N-acetylgalactosaminyltransferases, and nicotinic acetylcholine receptors. Immunity-associated transcripts accumulated differently in CTM and Rex-D S mosquitoes include five C-type lectins (AAEL008681-RA [CTL12], AAEL012353-RA [CTL15], AAEL005482-RA [CTL18], AAEL013566-RB [CTLGA2], AAEL017484-RA [CTLGA4]), two heme peroxidases (AAEL00376-RA [HPX4], AAEL002354-RA), one class B scavenger receptor (AAEL010655-RA [SCRBSP2]), a fibrinogen-related protein (AAEL008646-RA [FREP10]), and a Toll-like receptor (AAEL000671-RA [TOLL6]).

Quantitative reverse transcriptase, gene amplification, i.e., polymerase chain reaction (qRT-PCR) was performed on a selection of 13 transcripts and RNA from CTM, Rex-D, and LVP mosquitoes. The qRT-PCR results validated the RNA-seq data, with only two exceptions (transcript AAEL011871-RA in CTM and AAEL008848-RA in Rex-D; not shown). To assess the similarity of the RNA-seq data with the qRT-PCR measurements, we calculated Pearson correlations of 0.97 ($p < 0.001$), 0.85 ($p < 0.001$), and 0.90 ($p < 0.01$) for LVP, CTM, and Rex-D, respectively. T-test values were 2.18 ($p = 0.148$), 2.18 ($p = 0.453$), and 2.18 ($p = 0.188$), respectively, for the three strains (not shown).

The accumulation of 1845 and 981 transcripts consistently increased and decreased, respectively, in all three strains after a bloodmeal (not shown). In general, Rex-D exhibited the lowest fold-changes and the smallest number of transcripts accumulated highly after the bloodmeal (not shown). This observation supports the interpretation that CTM mosquitoes respond quicker or more intensively to a bloodmeal than LVP and Rex-D. To test this hypothesis, the expression profiles of five transcripts were analyzed by qRT-PCR at 5, 8, 12, and 24 h PBM. Results for four of the five transcripts

tested showed a greater response of CTM to a bloodmeal compared with mosquitoes of the other two strains (not shown).

Despite the observation that the same functional classes are represented within groups of transcripts that increased and decreased in accumulation after a bloodmeal, the numbers of transcripts associated with the various functions differed significantly (not shown). The transcripts accumulated after a bloodmeal were enriched in functions related to protein-turnover and chaperones, transcription, translation, posttranslational modification, and RNA processing. These transcripts encode proteins associated with catalytic activities, in transport and metabolism of lipids, carbohydrates and amino acids, and related to blood digestion and the progression of the gonotropic cycle. Many of the transcripts showing the greatest decrease in accumulation after a bloodmeal in all three strains encode structural proteins.

The number of strain-specific transcripts varying highly (< 10-fold) in abundance between B and S mosquitoes is smaller in general than those found in all three strains (not shown). CTM mosquitoes show both the greatest number of strain-specific, differentially accumulated transcripts (not shown) and the largest range of changes in transcript accumulation levels. Overall, the largest proportion of transcripts showing strain-specific increased accumulation after bloodmeal are linked to binding and catalytic functions and associated with transcription, translation, and posttranslational modification, whereas the largest proportion of strain-specific decreased transcripts are linked to transcription, translation, posttranscriptional modification, and RNA processing in LVP and CTM and to transport and metabolism of lipids, carbohydrates, and amino acids in Rex-D (not shown).

2.4.1.3 Protein network analysis of bloodmeal-induced changes in transcript accumulation

By using a Markov Cluster algorithm, [51] developed a protein interaction network in which 3500 *Ae. aegypti* proteins are organized into 494 functional modules. A total of 2465 of these proteins are encoded by transcripts exhibiting significant differential accumulation between B and S mosquitoes in at least one of the three strains analyzed (not shown). Modules enriched in Gene Ontology terms for cytoplasm organization and biogenesis, ribosome biogenesis and assembly, transcription, DNA metabolic processes, response to endogenous stimuli, and cell cycle were enriched significantly with proteins encoded by transcripts increased in accumulation in all three strains after a bloodmeal (not shown). A module enriched in Gene Ontology terms for cytoskeleton organization and biogenesis, reproduction, and cell differentiation was enriched significantly with proteins corresponding to transcripts consistently decreased in accumulation following a bloodmeal. Results from the protein network analyses are consistent with those from functional attribution of the differentially accumulated transcripts.

2.4.1.4 Metabolic pathways involving transcripts differentially accumulated post-bloodmeal

Metabolic pathways correlated with transcripts increased in accumulation after a bloodmeal in all three strains were visualized in LinkinPath [59] alongside pathways associated with transcripts accumulated in a strain-specific manner and transcripts decreased in accumulation after a bloodmeal (not shown). As seen with the function−parent attribution analyses, the transcripts increased and decreased in accumulation are associated with proteins involved in similar pathways. However, differential expression of paralogous genes among the strains was observed that could

account for the intensity and/or the rate of pathway activation/deactivation after a bloodmeal. For example, biosynthesis of fatty acids associates predominantly with transcripts decreased in accumulation in all three strains after a bloodmeal, but the elongation of fatty acids in mitochondria was exclusively associated with transcripts increased in accumulation (Figure 2.6). Proteins corresponding to transcripts accumulated differentially in a strain-specific manner were identified in both cases, with the same protein function encoded by paralogous transcripts in the different strains. Examples include FabI (enoyl-[acyl carrier protein] reductase) of the fatty acid biosynthesis pathway linked to transcripts AAEL003961-RA and AAEL002493-RA, which significantly increased in all three strains (fold-changes: 1.4-2.4); AAEL009634-RC and AAEL014840-RA, both of which increased 2-fold in CTM; AAEL000705-RA, AAEL000689-RA and AAEL004273-RA, which decreased in CTM (fold-changes: 2.5-4.4); AAEL009685-RB, which decreased 5.4-fold in Rex-D; and AAEL017302-RA, AAEL007669-RA, AAEL008227-RA, AAEL009625-RA, AAEL013491-RA, AAEL017320-RA, AAEL000690-RA, AAEL002228-RA, AAEL003148-RA, AAEL003183-RA, AAEL010075-RA, AAEL012400-RA, and AAEL008016-RA, which decreased in all three strains, with fold-changes in the range of 0.6-11.3. All of these transcripts are annotated currently as short-chain and steroid dehydrogenases, oxidoreductases, or fatty acid synthases [104] and match to short-chain dehydrogenases in the PFAM database [40], with the exception of AAEL002228-RA, whose best match is ketoacyl synthase. They all have the PKS_KR enzymatic domain as a best match in the SMART database [73], except for AAEL012400-RA, which matches the SEP enzymatic domain. The protein HADHA (4.2.1.17 in the fatty acid elongation pathway), functioning as an aldehyde reductase and enoyl-coA hydratase, is associated with transcripts AAEL010146-RB (annotated as 3-hydroxyacyl-coa dehydrogenase) and increased 2.9-fold in LVP and AAEL003993-RA (putative cyclohex-1-ene-1-carboxyl-CoA hydratase), which increased in all three strains with fold-changes ranging between 1.6 and 2.3.

Another example of differential accumulation of transcripts representing paralogous genes is seen in the lipase gene family. Three genes (AAEL001837, AAEL007055, AAEL14551) of the 72 annotated currently as lipases correspond to transcripts accumulated differentially after a bloodmeal in all three strains (fold-changes: -7.5 to 4.3) and eight to transcripts accumulated differentially in a strain-specific manner. Two (AAEL002909-RA, AAEL001076-RA) of these latter transcripts were found exclusively in S and one (AAEL006970-RA) only in B CTM mosquitoes. The remaining had fold-changes ranging from -2.2 to 2.6. No reads were mapped in any sample for 21 lipase genes, which may indicate that they are expressed exclusively in adult males or during pre-adult stages of *Ae. aegypti* development. Alternatively, these predicted genes could be pseudogenes.

2.4.2 Discussion

The results reported here reveal that the transcriptomes of *Ae. aegypti* mosquitoes from distinct strains vary significantly in complexity and abundance of specific transcripts. This variation is evident in non-bloodfed mosquitoes and is enhanced after a bloodmeal. CTM showed a larger number of differentially abundant transcripts after a bloodmeal and a wider range of fold-changes than Rex-D and LVP.

Transcripts associated with metabolism and transport of lipids, carbohydrate and amino acids, and the progression of the gonotropic cycles were among the most increased in accumulation after bloodmeal. Transcripts decreased most in accumulation after a bloodmeal are associated with structural components. Transcripts attributed functions in transcription, translation, and posttranslational protein modification also were highly differentially-accumulated, but many of these show strain-specific regulation. These observations support the hypothesis that important differences among

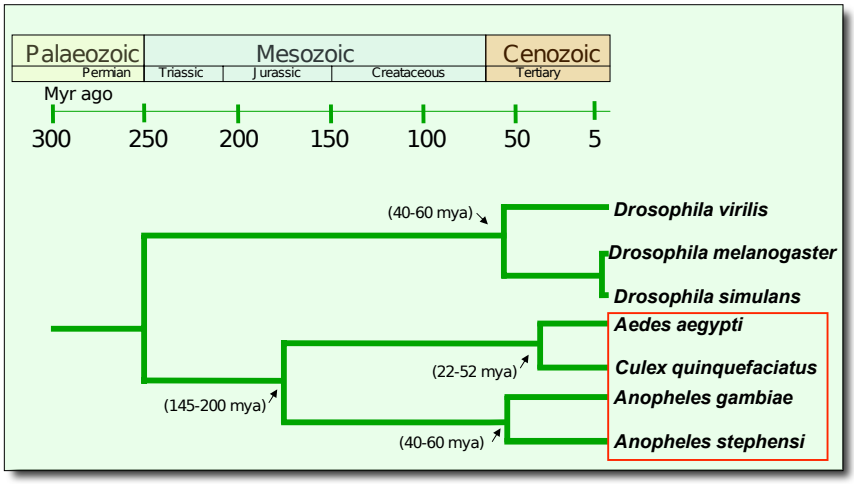
the strains are conferred by distinct patterns of gene expression, protein synthesis and modifications. Strain differences also may result from selective expression of paralogous transcripts.

Digestion and immunity share bio-products such as reactive oxygen species (ROS) [90], and these processes are linked at the protein network level [51]. The majority of immunity-related genes were decreased in accumulation in all three strains after a bloodmeal. The most pronounced decrease in accumulation was observed for the antimicrobial peptide attacin (AMP) (AAEL003389-RA). Indeed, all AMPs decreased in accumulation with the exception of CecF (AAEL000625-RA), which significantly increased in accumulation exclusively in CTM. The overall AMP decrease may be associated with an increase in bacterial proliferation observed after a bloodmeal [94]. The most common midgut bacteria do not show proteolysis activity but are implicated in the lysis of red blood cells, which release hemoglobin [44]. At the same time, hemoglobin negatively affects ROS production in the midgut, which triggers bacteria proliferation [94]. Although ROS reduction potentially favors DEN infection [94], bacteria proliferation antagonizes it [137]. Antioxidant activity by peroxidase and superoxide-dismutase varied across strains. CTM showed the greatest number of transcripts associated with antioxidant activity accumulated after a bloodmeal among the strains analyzed. In addition, S CTM mosquitoes accumulate greater levels of transcripts associated with antioxidant activity than Rex-D. CTM also had the largest number of P450 and glutathione-s transferase transcripts accumulated differentially after a bloodmeal, with increases in accumulation up to 15-fold (AAEL000325-RA). These results support a model of more intense antioxidant activity in B CTM than in Rex-D, which may facilitate DENV infection in the former.

Transcripts encoding proteins associated with autophagy, IAPs, IMD pathway, and SRRP members represent immunity genes that showed an exception to the overall

decline in corresponding-transcript accumulation after a bloodmeal. Increases were overall within 2-fold. A notable exception is the IMD pathway member Caspar1 (AAEL014734-RA) that increased 4.7- to 5.8-fold in all strains. Caspar is a negative regulator of the IMD pathway [65]; therefore, its up-regulation is consistent with the observed decrease in accumulation of AMPs.

In summary, the *Ae. aegypti* strains analyzed demonstrated variability in their transcriptomes before and after bloodmeal. Profiles differ in the number of transcripts detected, their level of accumulation in S mosquitoes, and in the changes in accumulation following the ingestion of a bloodmeal. Although these differences may result from distinct RNA turnover rates among strains, it is most likely a result of differential gene regulation. These data indicate the need for caution in making generalizations about individual gene expression profiles across different strains of *Ae. aegypti*. For example, constructs used in genetics-based strategies of vector control would require the previous analyses of cross-strain promoter activity [62]. The differences observed encompass several transcripts associated previously with vectorial capacity to DENV. Future studies will investigate the transcriptomes of CTM and Rex-D mosquitoes infected with DENV2. Also, it will be necessary to assess the susceptibility of CTM and Rex-D to other DENV serotypes to determine whether or how their distinct transcriptional responses to bloodmeals described herein influence vector competence.



Modified from: Sieglaff et al. PNAS. 2009 Mar 3;106(9):3053-8

Figure 2.1: Phylogenetic relationships between four vector mosquitoes compared with representative Drosophila species.
Adapted from [115]

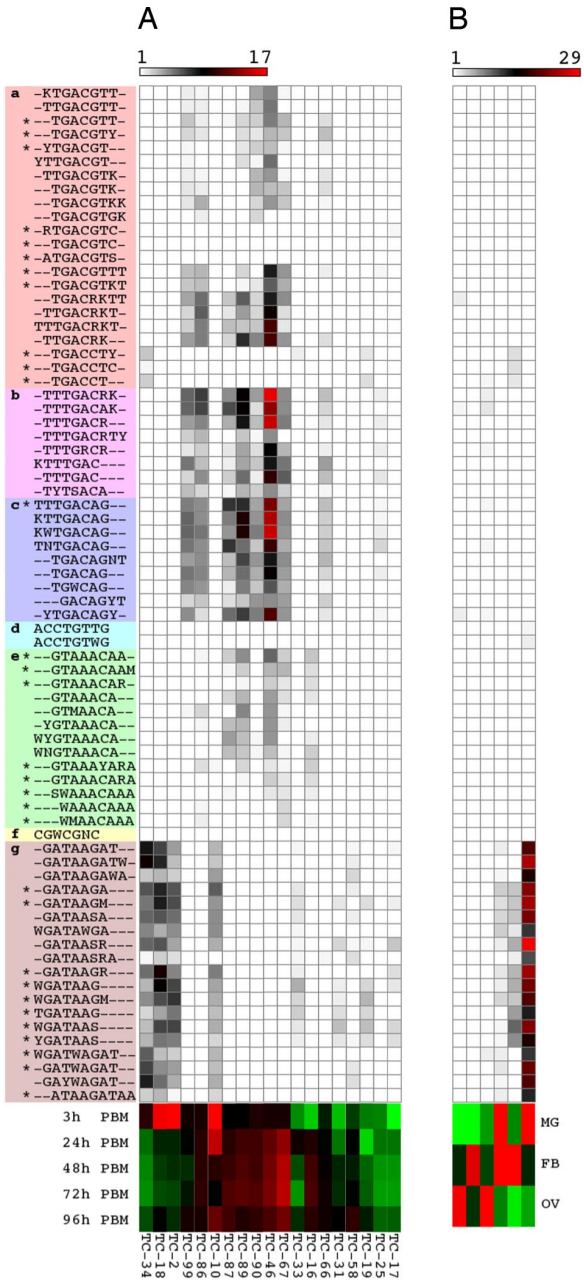


Figure 2.2: Associations of mosquito motifs with gene expression profiles in An. gambiae: Motif enrichment within (A) 5′-end flanking regions of genes in clusters responsive to blood meal ingestion, and in (B) 5′-end flanking regions of genes in clusters enriched in selected tissues. The significance of motif enrichment is indicated by pseudocolor of -log10 (P-value) determined through hypergeometric statistics, and the median expression profile of each gene cluster is shown below each respective column. Red and green colors represent higher and lower relative mRNA accumulation, respectively. Asterisks (*) indicate a match to a previously described mosquito TFBS. Heatmaps were created with Matrix2png (CITEME:local-58). FB, fat body; hPBM, hours post blood meal; MG, midgut; OV, ovaries; TC, time course clusters.
Adapted from [115]

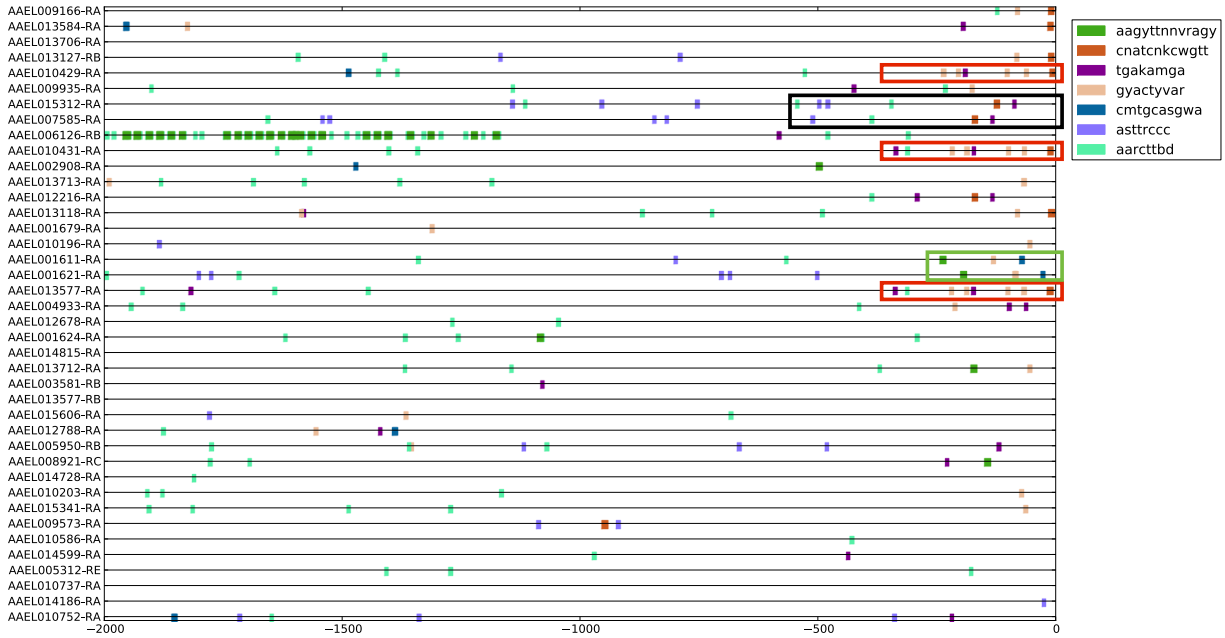


Figure 2.3: Motif map of putative CREs discovered by SCOPE using transcripts detected significantly only in bloodfed female Ae. aegypti: Locations of representative SCOPE-derived CRE motifs in the 2000 bp upstream of the annotated translational start site in the 40 transcripts detected significantly only in bloodfed females. Transcript names on the left are ordered from most (top) to least (bottom) abundant. Candidate CRMs are highlighted in like-colored rectangles. Adapted from [20]

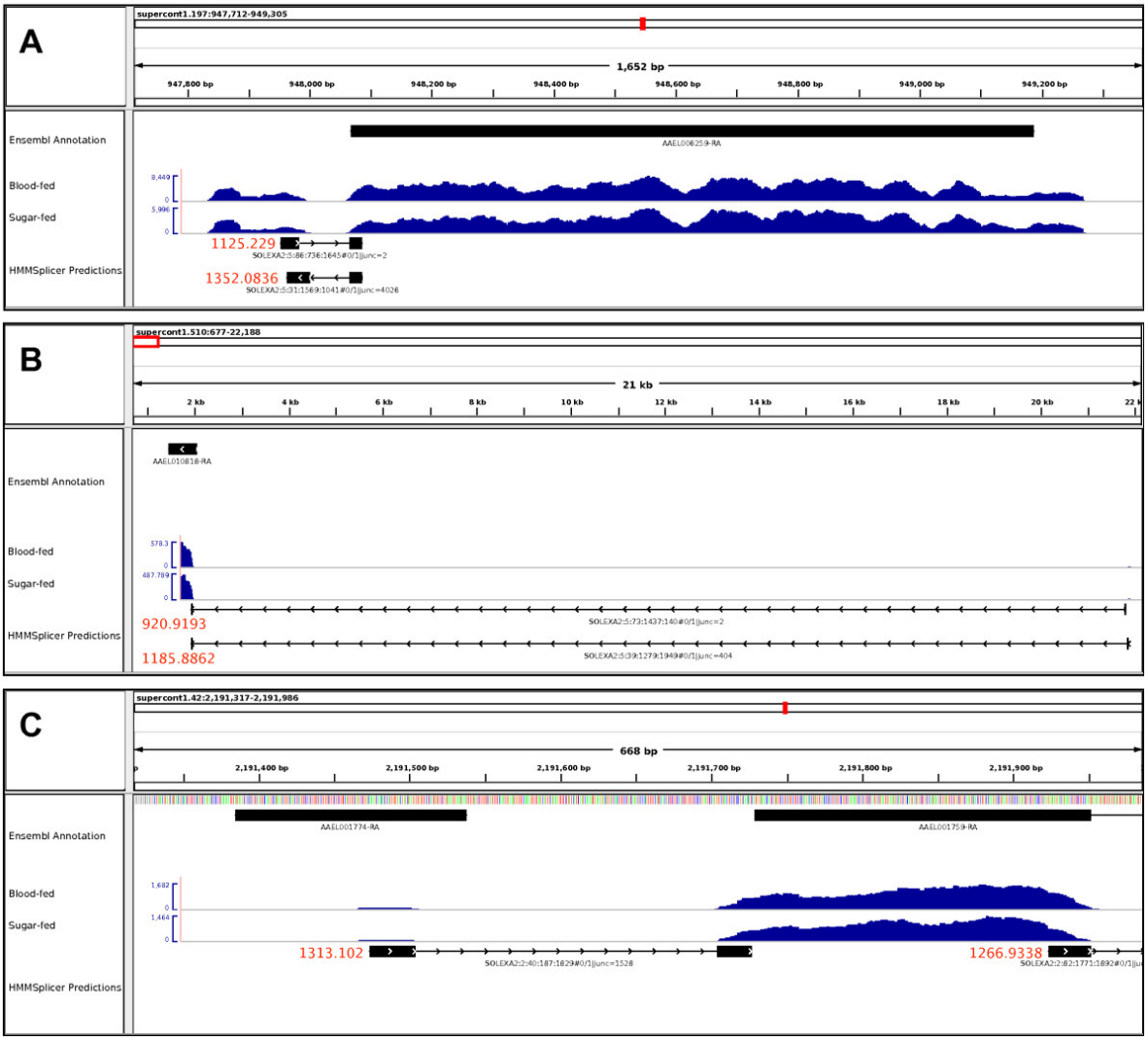


Figure 2.4: Examples of amendments to the Ae. aegypti annotation supported by HMMSplicer results: Black bars in the top tracks represent the current gene annotations. Blue histograms in the second track represent the non-normalized coverage of RNA-seq reads at each position. The range of the histogram values shown in each view is depicted on the labeled y-axis of each RNA-seq track. Black boxes in the lower track represent splice-site predictions based on the RNA-seq reads using HMMSplicer determined in this study. Each function has a unique identifier listed below and its HMMSplicer score is listed in red. If multiple reads support a single junction, "junc = x" lists the number of supporting reads. This information provides evidence to link two islands of transcription as a single transcription event, therefore, exons of a common mRNA. All predicted junctions shown here also are supported by EST alignments. Genes are (A) AAEL006259; (B) AAEL010818; and (C) AAEL001774 and AAEL001759.
Excerpted from [20]

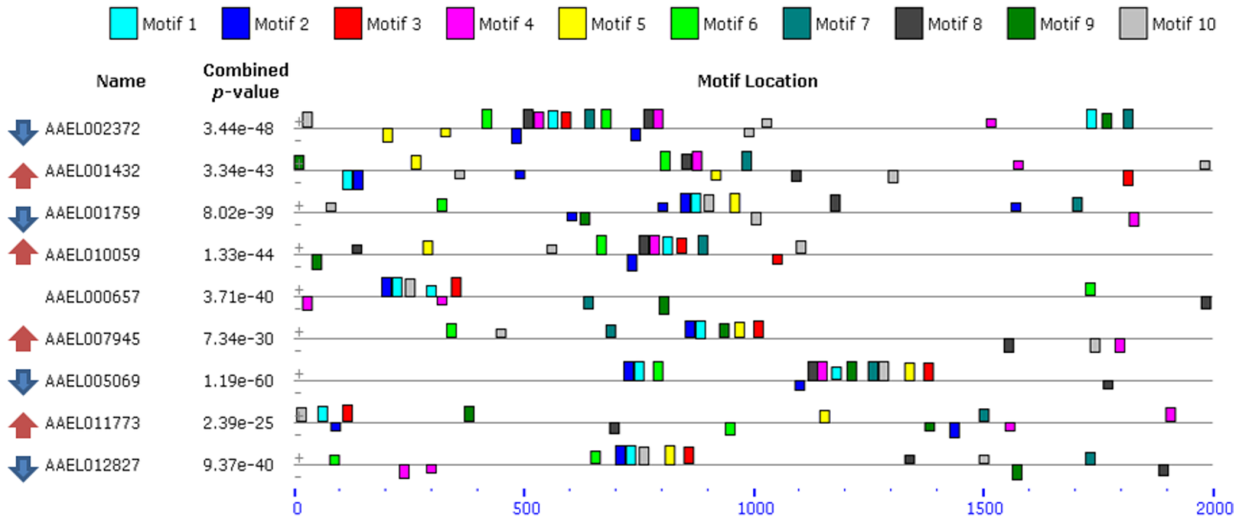


Figure 2.5: MEME analysis of nine genes with $FPKM_{DENVI} \geq 100$ in carcasses and salivary glands at 14 dpi These genes also were identified with transcripts exhibiting significant differential accumulation in analyses of salivary gland samples of the Liverpool strain infected with DENV2 Thailand 16881 [76]. Colored boxes represent individual putative CREs and their locations in promoters of each gene. Red and blue arrows adjacent to Ensembl Gene ID indicate those genes whose transcripts were detected previously as more or less abundant following DENV infection [76]. Distances in base-pairs are provided below the schematic of each gene. doi:10.1371/journal.pone.0050512.g004
Excerpted from [21]

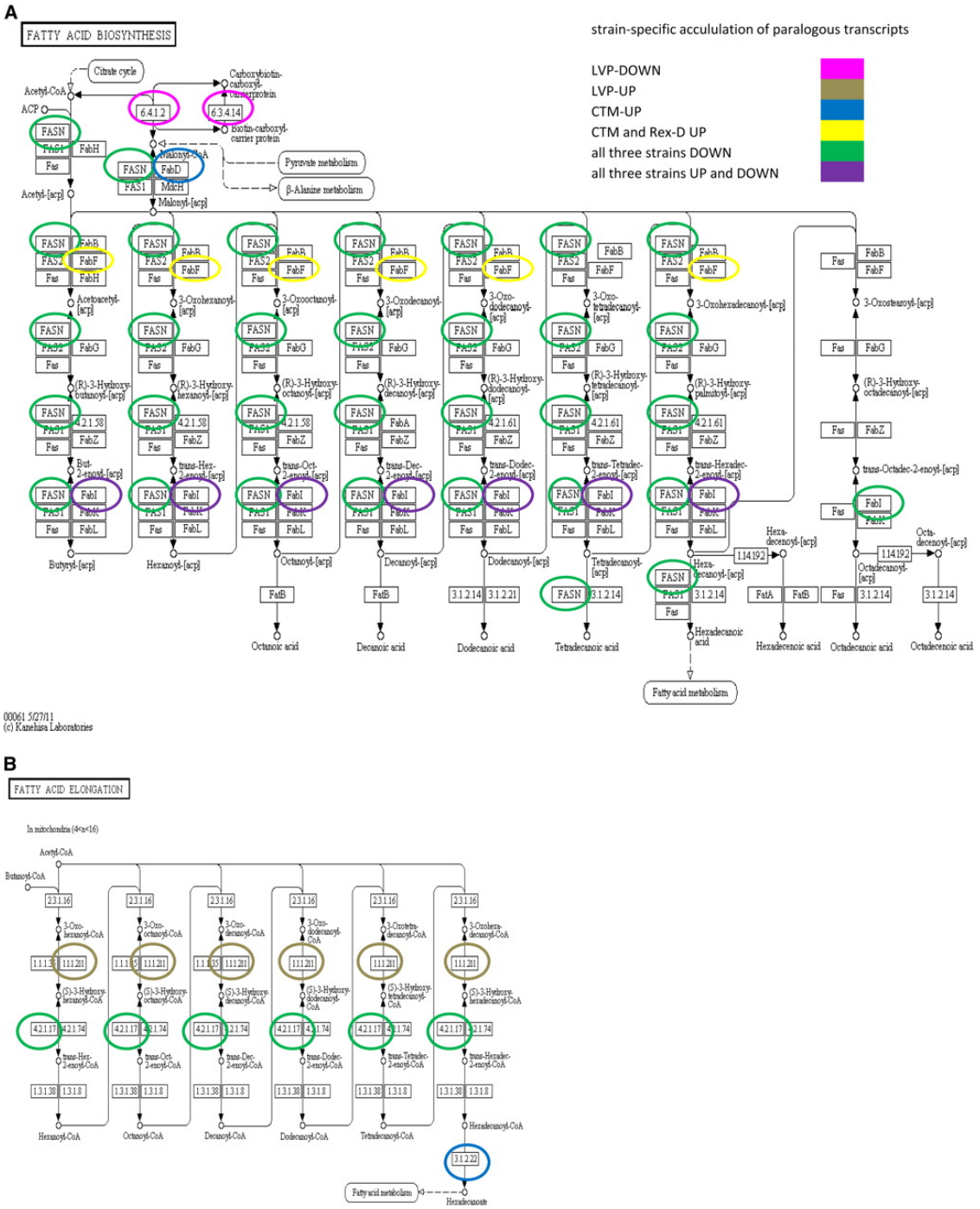


Figure 2.6: Fatty acid biosynthesis and elongation in mitochondria: Proteins corresponding to transcripts accumulated in a strain-specific manner are circled in a color corresponding to the strain and condition defined in the legend (panels A and B).
Excerpted from [22]

Chapter 3

Approach and Custom Software solutions

Functional Genomics can be thought of as a problem of generating meaningful gene-sets. Signal-to-noise is major challenge in generating useful gene-sets. This project aims to generate meaningful gene-sets through integration of multiple supporting data-types

- comparative genomics (N-way 1-to-1 orthologs)
- comparative transcriptomics (identical RNA-seq experiments on divergent Mosquito species with shared trait: bloodfeeding)
- phylogenetics (estimated divergence since last common ancestor)
- existing knowledge about certain bloodmeal related transcriptional control programs (20-hydroxyecdysone (20E) interacting TFBS motifs)

Comparative Genomics: The initial gene-set for this project was limited to genes that have exactly one matching 1-to-1 ortholog in *Ae. aegypti*, *An. gambiae*, *An. stephensi*, and *Cx. quinquefasciatus*. The purpose of this decision is to maximize the probability of conserved function between genes analyzed.

This preliminary gene set was obtained through the orthoDB7 http://cegg.unige.ch/orthodb7 website by first selecting the node representing only the mosquito branch of the phylogeny [132]. Then the following requirements were enforced when retrieving ortholog relationship results

```
co-ortholog copy-number:  AAEGY=1, AGAMB=1, ASTEP=1, CQUIN=1
```

The results of the query are provided as file `orthodb7_mosqs_20130918.txt` in the supplementary data files associated with this dissertation http://dx.doi.org/10.6084/m9.figshare.810442 [37].

Custom python code was used to extract the 1-to-1 orthologs for *Ae. aegypti*, *An. gambiae*, and *Cx. quinquefasciatus*. The coded process is provided as an executable IPython notebook located at http://goo.gl/EM1pHL for in-depth inspection as well as for the purpose of openness and reproducibility.

Briefly, the results table was filtered for data pertaining to *Ae. aegypti*, *An. gambiae*, or *Cx. quinquefasciatus*. Then I built a graph-representation of the data such that each orthoDB ortholog group could be queried to ensure that exactly one gene was represented for each species. These genes represent gene-set that formed the starting point for the rest of my approach. The filtered set is available as supplementary file 'AaAgCq_1to1_orthos.tsv'.

It should be noted that while this set contains only genes from the three species included in this document, each group should also have exactly one ortholog from *An. stephensi* as well.

Comparative Transcriptomics: One of this study's aims is to identify and analyze orthologous genes that exhibit correlated mRNA accumulation profiles (mAPs) in the midguts of females directly following the ingestion of a bloodmeal. To this end, whole RNA was extracted from dissected midguts of 16 female individuals of each species among the following time points: non-bloodfed (NBF) (0), 2, 4, 6, 8, 10 h PBM (Protocol **??**).

Phylogenetics: The phylogenetics aspect of this approach consists of using the estimated phylogenetic distances between each species to weight the comparisons between data correlations calculated for each pairwise 1-to-1 ortholog duo. This is encoded as d in the Phylogenetic Transcriptional Correlation Index (PTCI) expression defined in Equation 3.1.

The range of d for each species-pair has been estimated previously (compiled in [115]) and were as follows: - *Ae. aegypti*:*An. gambiae*= 145 MYA - *An. gambiae*:*Cx. quinquefasciatus*= 145 MYA - *Ae. aegypti*:*Cx. quinquefasciatus*= 22 MYA

Targeting Specific Transcriptional Control Mechanisms: Part of the glsPTCI includes correlation of a glsTFBS scanning signature between ortholog pairs. If the glsPTCI is calculated using a set of glsTFBS that are known to interact with the 20E signaling cascade, then this portion of the glsPTCI acts as a rough proxy for correlation with putative activity of the 20E cascade.

This is an imperfect solution as ChIP-Seq using antibodies to these TFs in each species is preferable. However this was impractical for this work due to the absence of suitable antibodies and the cost of developing and executing all assays in each species.

3.0.2.0.1 Custom Software Solutions: Two core bodies of software were developed for this project. Blacktie is a python package that automates data flow between and record keeping for analysis runs of the components of the third-party analysis pipeline dubbed the Tuxedo Protocol [125]. The purpose of the second software project graph-based functional genomics framework (gFunc) is to provide a framework to integrate many, disparate data types to facilitate the synthesis of functional genomics analyses that make use of as many types of data as may be useful.

Blacktie's code and documentation have been made available to the public. It is hosted as a git project with the complete history of commits included and available for easy open source improvement and collaboration. It is also available as an easily installable package from the Python Package Index (PyPI) for general users. It has been downloaded and installed over 1000 times as of this time with its latest version being downloaded 284 times.

glsgFunc is based around an existing, well known, and well maintained network-graph framework called networkx [52]. glsgFunc uses this code base as a database backbone and provides a framework for installing common genomic data-types into the graph structure as well as a convention for extending the core objects of glsgFunc to handle new types of data. The project is not formally released. After using it for my own work I have learned more about how best to organize the code. Once it has been re-factored to incorporate the lessons learned here (mainly that the core objects should be simplified as much as possible and more of the data management should be coded directly

with networkx rather than re-implementing certain functions) it will be release to the public.

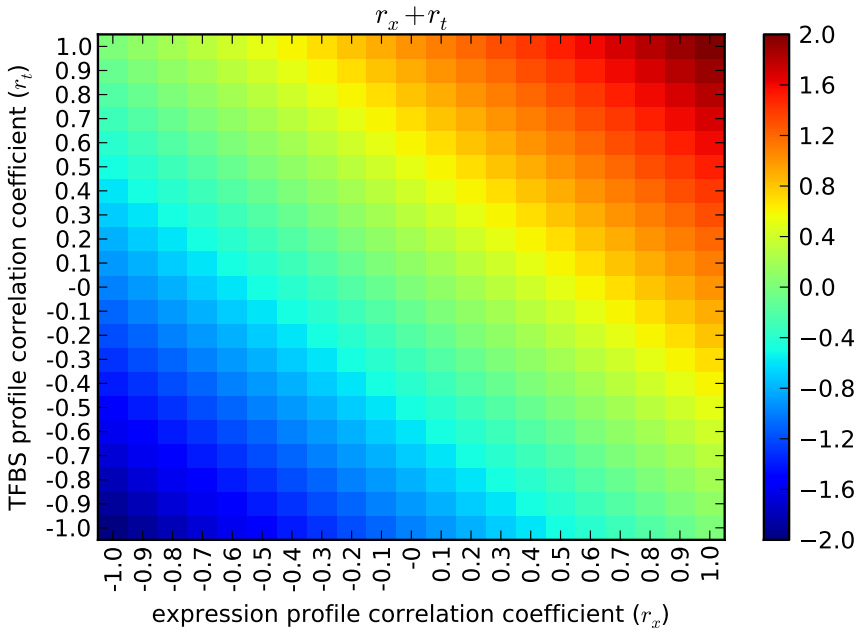
3.1 The phylogenetic transcriptional correspondence index (PTCI)

Look at the parameter space of the PTCI (Figure 3.1).

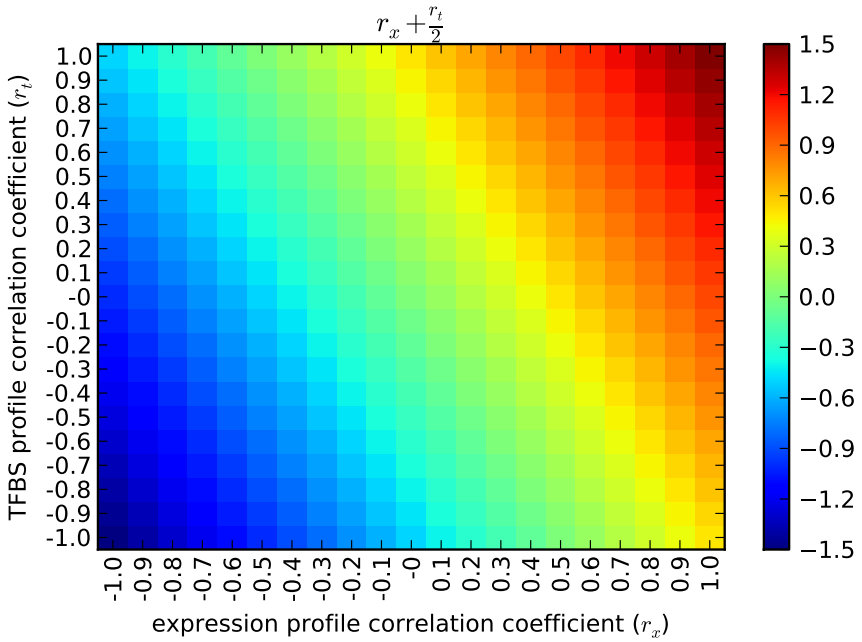
Here is the glsPTCI calculation:

$$PTCI = r_x + \frac{r_t}{2} \cdot w(d) \tag{3.1}$$

The graph model is useful (Figure 3.2)

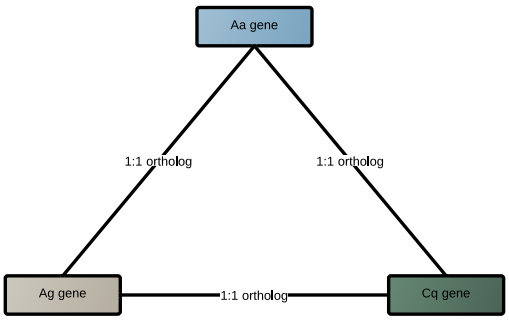


(A)

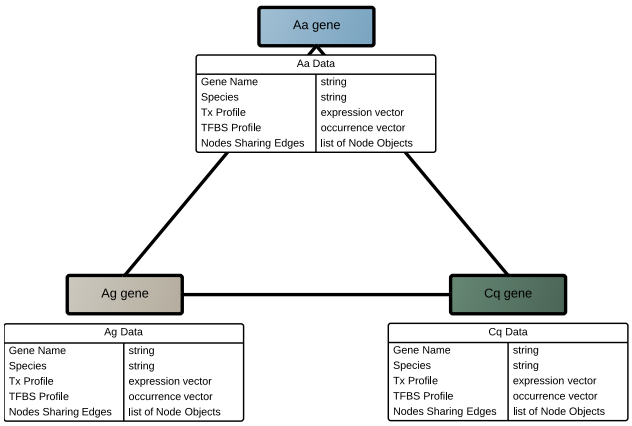


(B)

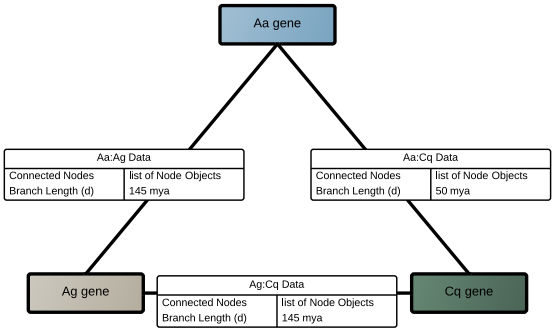
Figure 3.1: Exploring the parameter space of the expression and TFBS components of the PTCI:
(A) The expression and TFBS profile correlations each carry the same weight. TFBS data is not strictly empirical, and position weight matrix models inherently are prone to produce false positive predictions. Here, each correlation-type exerts equal influence on the final score assigned to the ortholog relationship. (B) The TFBS profile correlation is penalized due to the non-empirical nature and expected false positive information it contains. Here, TFBS profile data contributes to the final score at most 50% as much as expression data.



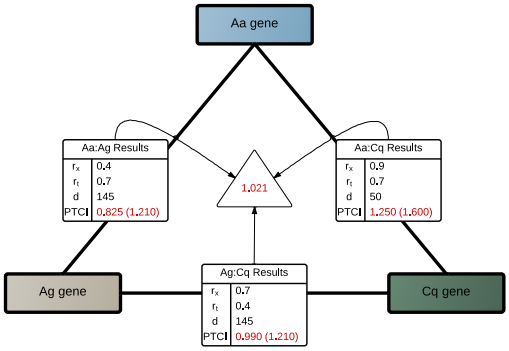
(A) N-way 1:1 ortholog graph structure



(B) Node data model



(C) Edge data model



(D) Example PTCI data

Figure 3.2: Graph model used to integrate data types

3.2 Blacktie RNA-seq pipeline

See Figure 3.3

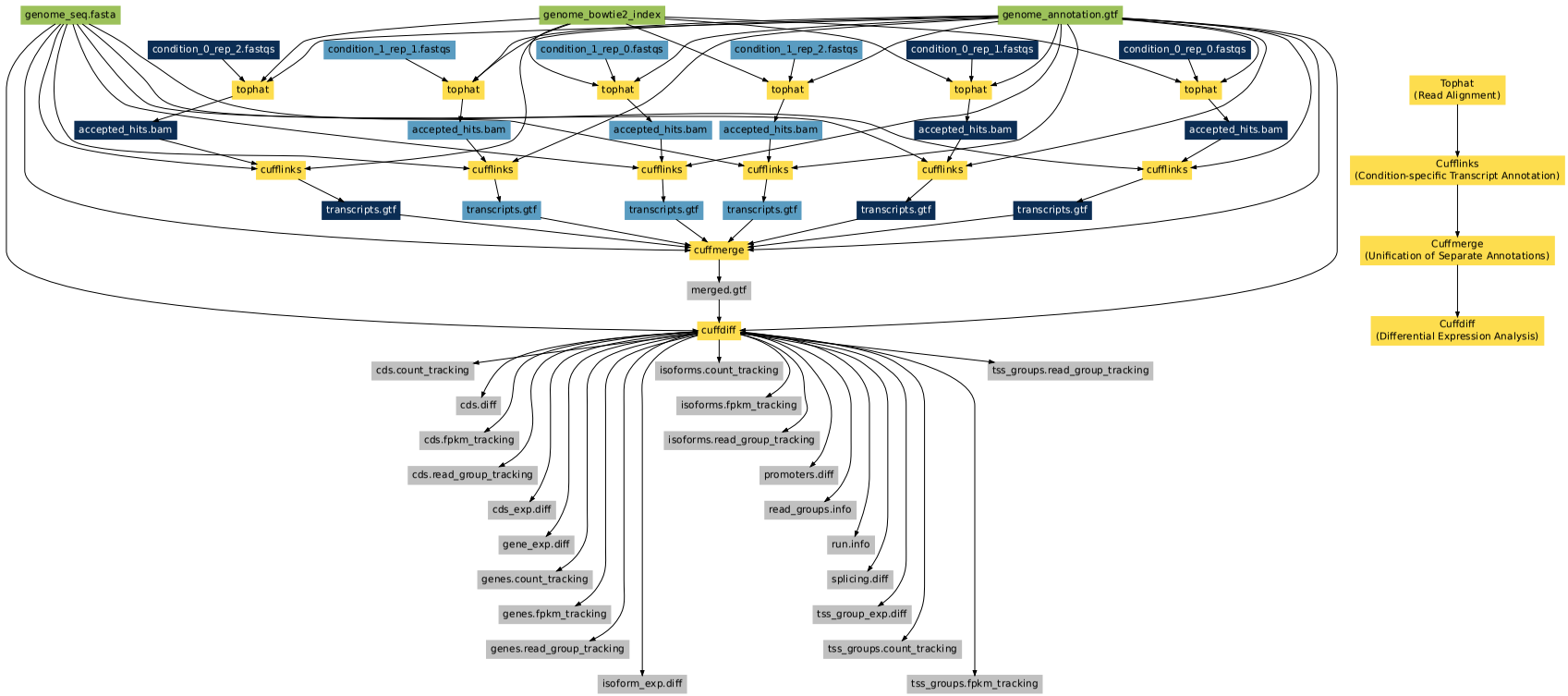


Figure 3.3: Diagram of Abbreviated Tophat/Cufflinks Inputs and Outputs:
This figure demonstrates the complexities of a typical RNA-seq experiment as analyzed with the "Tuxedo Protocol". It models a fairly **simple** two condition experiment with each condition having three replications. It is "abbreviated" in that it only displays the output files that will be used in the next step for all steps except the `cuffdiff` step. The complete output of `cuffdiff` is included to demonstrate the final challenge of integrating the data which exists in multiple cross-referenced files.
(Dark Blue - Inputs/Outputs associated with Condition Zero; Light Blue - Inputs/Outputs associated with Condition One; Grey - Inputs/Outputs associated with Condition Zero **AND** Condition One; Green - Inputs Specific to the Reference Genome; Gold - Program Calls)

Chapter 4

Analysis of Conservation of Midgut Transcriptional Response to Bloodfeeding in Four Vector Mosquitoes

4.1 Background

1. Biological importance of bloodmeal to mosquito
2. Biological importance of bloodmeal to transmission cyc

Figure 4.1

Figure 4.2

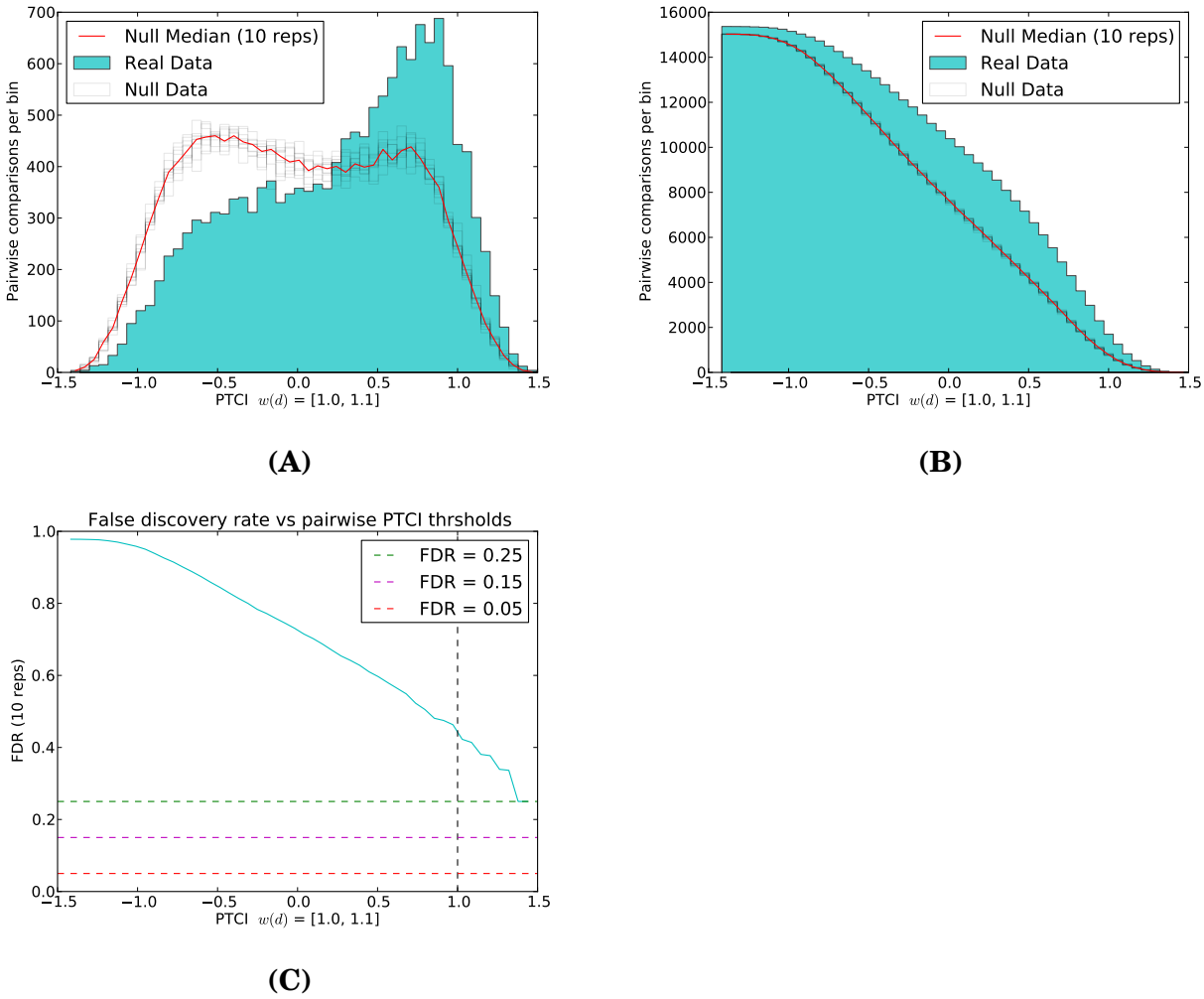


Figure 4.1: Pairwise PTCI results:
(A) Histogram of pairwise PTCI results vs null distributions. (B) Reverse Cumulative histogram of pairwise PTCI results vs null distributions. (C) False discover rate vs PTCI threshold.

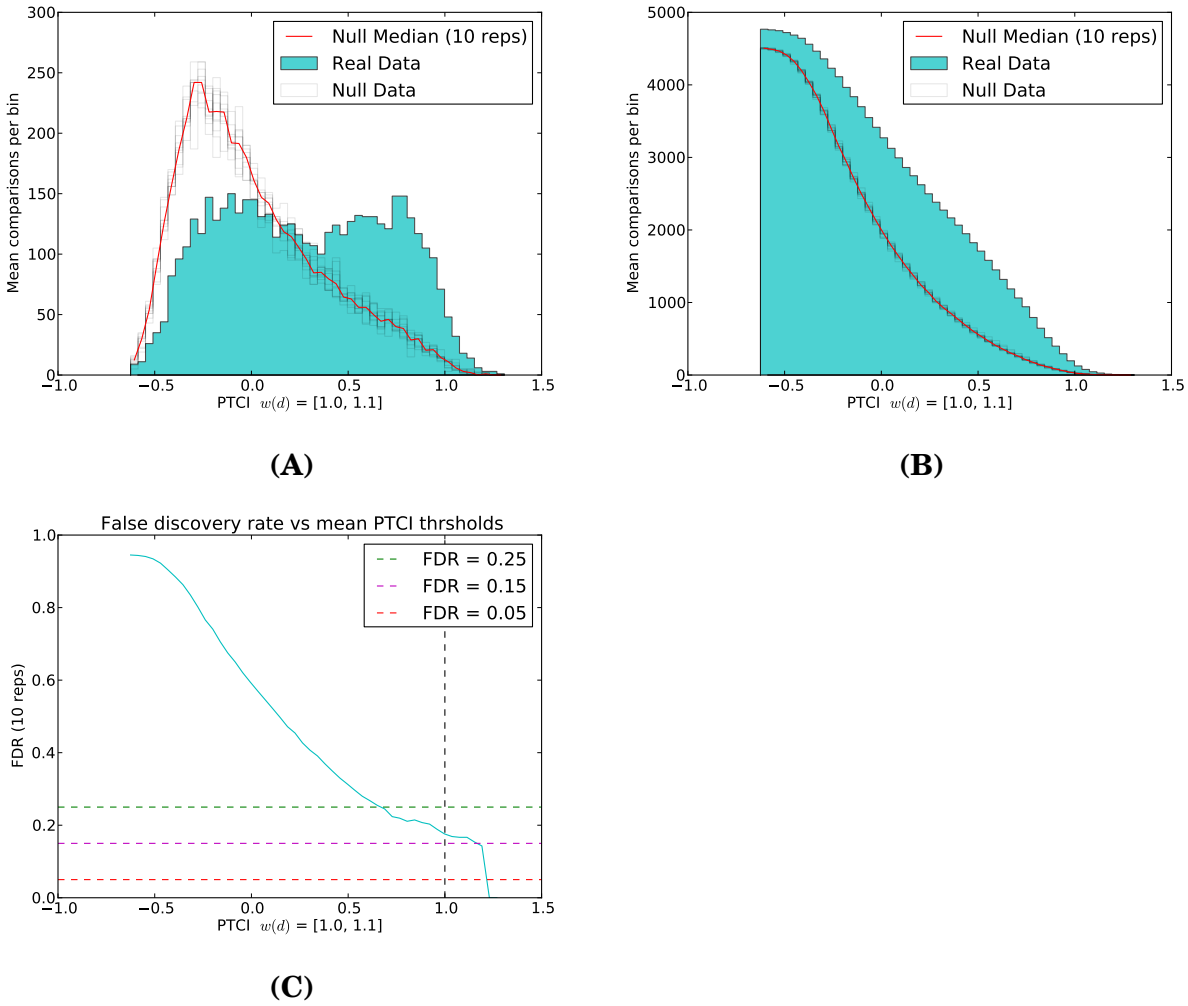


Figure 4.2: Mean PTCI results:
(A) Histogram of mean PTCI results vs null distributions. (B) Reverse Cumulative histogram of mean PTCI results vs null distributions. (C) False discover rate vs PTCI threshold.

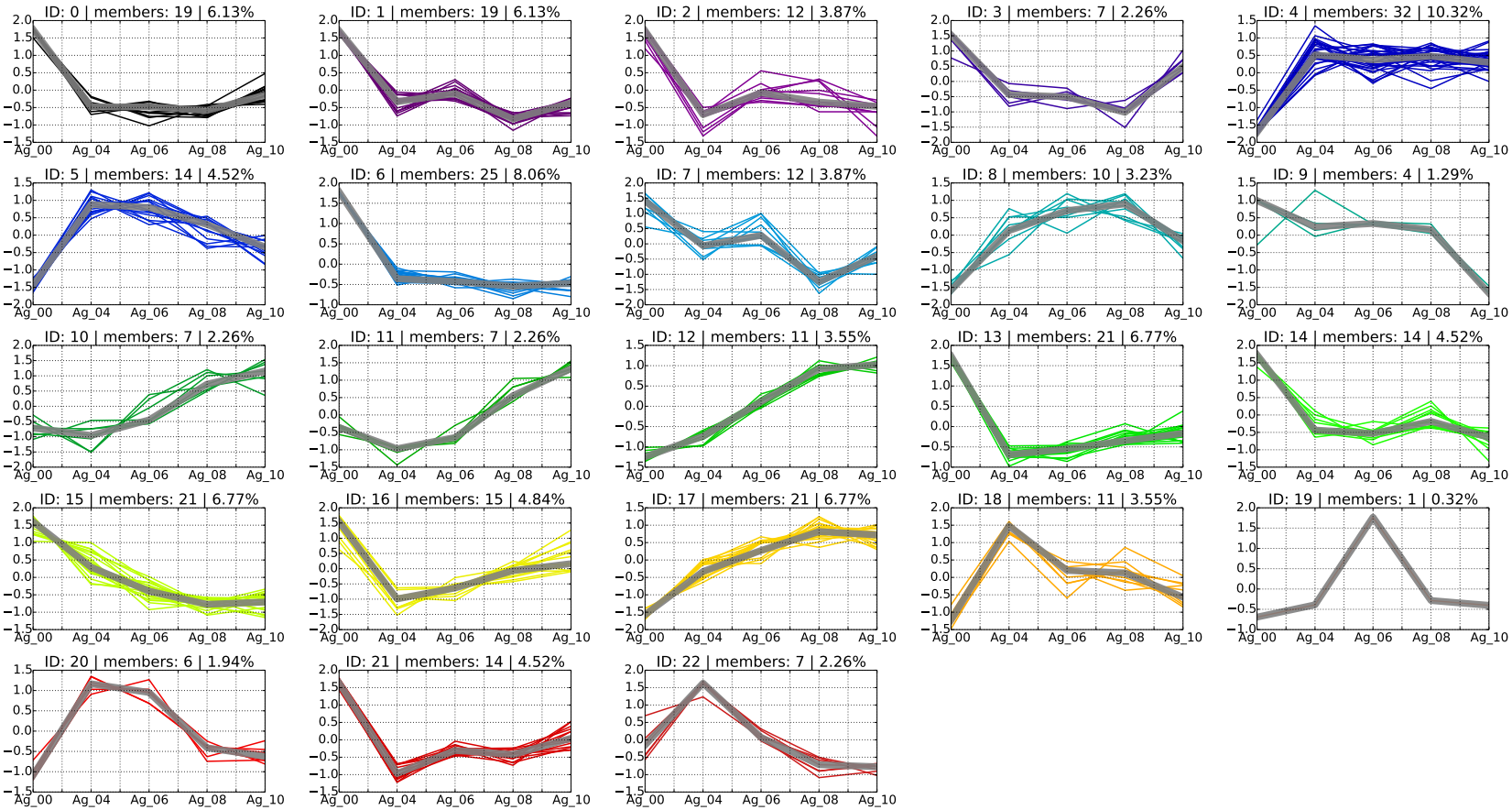


Figure 4.3: An. gambiae clustered abundance profiles.
Clusters were generated using k-means clustering as implemented in Biopython version 1.62 [29]. Abundance profile data was log transformed after one FPKM was added to all data to remove zeros ($log_{10}(\mathrm{FPKM}+1)$). K-means was then applied using the arithmetic mean as the center for cluster definition. The data displayed here is the gene-wise standardization of the raw FPKM data such that each profile has mean = 0 and standard deviation = 1. Each median abundance profile is marked by a thick gray line. Panel Titles - ID: cluster identifier | members: number of genes in cluster | percentage of genes represented in all clusters. Time Points - NBF, 4, 6, 8, 10 h PBM.

Figure 4.3

Chapter 5

Contributions to Literature Not Otherwise Mentioned

5.1 Purpose

xxxxxxxxxxxxxxxxxxxxxxxxxxx

1. xxx

Chapter 6

Conclusions

6.1 Purpose

xxxxxxxxxxxxxxxxxxxxxxxxxxx

1. xxx

Bibliography

[1] Abrams EW, Andrew DJ (2005) CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. Development (Cambridge, England) 132: 2743–58.

[2] Adelman ZN, Jasinskiene N, Vally KJM, Peek C, Travanty EA, et al. (2004) Formation and loss of large, unstable tandem arrays of the piggyBac transposable element in the yellow fever mosquito, Aedes aegypti. Transgenic research 13: 411–25.

[3] Agawa Y, Sarhan M, Kageyama Y, Akagi K, Takai M, et al. (2007) Drosophila Blimp-1 is a transient transcriptional repressor that controls timing of the ecdysone-induced developmental pathway. Molecular and cellular biology 27: 8739–47.

[4] Ahmed a, Martín D, Manetti aG, Han SJ, Lee WJ, et al. (1999) Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of America 96: 14795–800.

[5] Amenya DA, Bonizzoni M, Isaacs AT, Jasinskiene N, Chen H, et al. (2010) Comparative fitness assessment of Anopheles stephensi transgenic lines receptive to site-specific integration. Insect molecular biology 19: 263–9.

[6] Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome biology 11: R106.

[7] Anderson JR, Rico-Hesse R (2006) Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus. The American journal of tropical medicine and hygiene 75: 886–92.

[8] Angel B, Joshi V (2008) Distribution and seasonality of vertically transmitted dengue viruses in Aedes mosquitoes in arid and semi-arid areas of Rajasthan, India. Journal of vector borne diseases 45: 56–9.

[9] Armstrong PM, Rico-Hesse R (2001) Differential Susceptibility of Aedes aegypti to Infection by the American and Southeast Asian Genotypes of Dengue Type 2 Virus. Vector-Borne and Zoonotic Diseases 1: 159–168.

[10] Attardo GM, Higgs S, Klingler Ka, Vanlandingham DL, Raikhel AS (2003) RNA interference-mediated knockdown of a GATA factor reveals a link to anautogeny in the mosquito Aedes aegypti. Proceedings of the National Academy of Sciences of the United States of America 100: 13374–9.

[11] Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB International Conference on Intelligent Systems for Molecular Biology 2: 28–36.

[12] Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic acids research 34: W369–73.

[13] Bargielowski I, Kaufmann C, Alphey L, Reiter P, Koella J (2012) Flight performance and teneral energy reserves of two genetically-modified and one wild-type strain of the yellow fever mosquito Aedes aegypti. Vector borne and zoonotic diseases (Larchmont, NY) 12: 1053–8.

[14] Behura SK, Gomez-Machorro C, Harker BW, DeBruyn B, Lovin DD, et al. (2011) Global cross-talk of genes of the mosquito Aedes aegypti in response to dengue virus infection. PLoS neglected tropical diseases 5: e1385.

[15] Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57: pp. 289–300.

[16] Bennett KE, Olson KE, Muñoz MdL, Fernandez-Salas I, Farfan-Ale JA, et al. (2002) Variation in vector competence for dengue 2 virus among 24 collections of Aedes aegypti from Mexico and the United States. The American journal of tropical medicine and hygiene 67: 85–92.

[17] Bernhardt SA, Simmons MP, Olson KE, Beaty BJ, Blair CD, et al. (2012) Rapid intraspecific evolution of miRNA and siRNA genes in the mosquito Aedes aegypti. PloS one 7: e44198.

[18] Black WC, Bennett KE, Gorrochótegui-Escalante N, Barillas-Mury CV, Fernández-Salas I, et al. (2002) Flavivirus susceptibility in Aedes aegypti. Archives of medical research 33: 379–88.

[19] Boggaram V (2009) Thyroid transcription factor-1 (TTF-1/Nkx2.1/TITF1) gene regulation in the lung. Clinical science (London, England : 1979) 116: 27–35.

[20] Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Dimon MT, et al. (2011) RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, Aedes aegypti. BMC Genomics 12: 82.

[21] Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Marinotti O, et al. (2012) Complex modulation of the Aedes aegypti transcriptome in response to dengue virus infection. PloS one 7: e50512.

[22] Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Marinotti O, et al. (2012) Strain Variation in the Transcriptome of the Dengue Fever Vector, Aedes aegypti. G3: Genes | Genomes | Genetics 2: 103–114.

[23] Borkent A, Grimaldi DA (2004) The Earliest Fossil Mosquito (Diptera: Culicidae), in Mid-Cretaceous Burmese Amber. Annals of the Entomological Society of America 97: 882–888.

[24] Brackney DE, Isoe J, W C B, Zamora J, Foy BD, et al. (2010) Expression profiling and comparative analyses of seven midgut serine proteases from the yellow fever mosquito, Aedes aegypti. Journal of insect physiology 56: 736–44.

[25] Bushland RC, Lindquist AW, Knipling EF (1955) Eradication of Screw-Worms through Release of Sterilized Males. Science 122: 287–8.

[26] Calvo E, Mans BJ, Andersen JF, Ribeiro JMC (2006) Function and evolution of a mosquito salivary protein family. The Journal of biological chemistry 281: 1935–42.

[27] Carlson JM, Chakravarty A, DeZiel CE, Gross RH (2007) SCOPE: a web server for practical de novo motif discovery. Nucleic acids research 35: W259–64.

[28] Cho KH, Cheon HM, Kokoza V, Raikhel AS (2006) Regulatory region of the vitellogenin receptor gene sufficient for high-level, germ line cell-specific ovarian expression in transgenic Aedes aegypti mosquitoes. Insect biochemistry and molecular biology 36: 273–81.

[29] Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics (Oxford, England) 25: 1422–3.

[30] Colpitts TM, Cox J, Vanlandingham DL, Feitosa FM, Cheng G, et al. (2011) Alterations in the Aedes aegypti transcriptome during infection with West Nile, dengue and yellow fever viruses. PLoS pathogens 7: e1002189.

[31] Cox J, Brown HE, Rico-Hesse R (2011) Variation in vector competence for dengue viruses does not depend on mosquito midgut binding affinity. PLoS neglected tropical diseases 5: e1172.

[32] Das MK, Dai HK (2007) A survey of DNA motif finding algorithms. BMC bioinformatics 8 Suppl 7: S21.

[33] Davidson EH (2010) The Regulatory Genome: Gene Regulatory Networks In Development And Evolution (Google eBook). Academic Press, 304 pp.

[34] Deubel V, Kinney RM, Trent DW (1986) Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. Virology 155: 365–77.

[35] Dimon MT, Sorber K, DeRisi JL (2010) HMMSplicer: A Tool for Efficient and Sensitive Discovery of Known and Novel Splice Junctions in RNA-Seq Data. PLoS ONE 5: 16.

[36] Dittmer NT, Sun G, Wang Sf, Raikhel AS (2003) CREB isoform represses yolk protein gene expression in the mosquito fat body. Molecular and cellular endocrinology 210: 39–49.

[37] Dunn A (2013) Dissertation Supplementary Files .

[38] Elemento O, Tavazoie S (2005) Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. Genome biology 6: R18.

[39] Facchinelli L, Valerio L, Ramsey JM, Gould F, Walsh RK, et al. (2013) Field cage studies and progressive evaluation of genetically-engineered mosquitoes. PLoS neglected tropical diseases 7: e2001.

[40] Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic acids research 36: D281–8.

[41] Fischer HM, Wheat CW, Heckel DG, Vogel H (2008) Evolutionary origins of a novel host plant detoxification gene in butterflies. Molecular biology and evolution 25: 809–20.

[42] Flajollet S, Poras I, Carosella ED, Moreau P (2009) RREB-1 is a transcriptional repressor of HLA-G. Journal of immunology (Baltimore, Md : 1950) 183: 6948–59.

[43] Franz AWE, Sanchez-Vargas I, Adelman ZN, Blair CD, Beaty BJ, et al. (2006) Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. Proceedings of the National Academy of Sciences of the United States of America 103: 4198–203.

[44] Gaio AdO, Gusmão DS, Santos AaV, Berbert-Molina MA, Pimenta PFP, et al. (2011) Contribution of midgut bacteria to blood digestion and egg production in aedes aegypti (diptera: culicidae) (L.). Parasites & vectors 4: 105.

[45] Giannoni F, Müller HM, Vizioli J, Catteruccia F, Kafatos FC, et al. (2001) Nuclear factors bind to a conserved DNA element that modulates transcription of Anopheles gambiae trypsin genes. The Journal of biological chemistry 276: 700–7.

[46] Gubler DJ (2002) Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends in microbiology 10: 100–3.

[47] Gubler DJ, Novak RJ, Vergne E, Colon NA, Velez M, et al. (1985) Aedes (Gymnometopa) mediovittatus (Diptera: Culicidae), a potential maintenance vector of dengue viruses in Puerto Rico. Journal of medical entomology 22: 469–75.

[48] Gubler DJ, Rosen L (1976) A simple technique for demonstrating transmission of dengue virus by mosquitoes without the use of vertebrate hosts. The American journal of tropical medicine and hygiene 25: 146–50.

[49] Günther J, Martínez-Muñoz JP, Pérez-Ishiwara DG, Salas-Benito J (2007) Evidence of vertical transmission of dengue virus in two endemic localities in the state of Oaxaca, Mexico. Intervirology 50: 347–52.

[50] Guo M, Guo W, Chen Y, Dong S, Zhang X, et al. (2010) The basic leucine zipper transcription factor Moatf1 mediates oxidative stress responses and is necessary for full virulence of the rice blast fungus Magnaporthe oryzae. Molecular plant-microbe interactions : MPMI 23: 1053–68.

[51] Guo X, Xu Y, Bian G, Pike AD, Xie Y, et al. (2010) Response of the mosquito protein interaction network to dengue infection. BMC genomics 11: 380.

[52] Hagberg AA, National LA, Alamos L (2008) Exploring Network Structure , Dynamics , and Function using NetworkX pp. 11–15.

[53] Hansen Ia, Sieglaff DH, Munro JB, Shiao SH, Cruz J, et al. (2007) Forkhead transcription factors regulate mosquito reproduction. Insect biochemistry and molecular biology 37: 985–97.

[54] Heinz FX, Allison SL (2003) Flavivirus structure and membrane fusion. Advances in virus research 59: 63–97.

[55] Hernandez-Romano J, Carlos-Rivera FJ, Salgado H, Lamadrid-Figueroa H, Valverde-Garduño V, et al. (2008) Immunity related genes in dipterans share common enrichment of AT-rich motifs in their 5' regulatory regions that are potentially involved in nucleosome formation. BMC genomics 9: 326.

[56] Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, et al. (2011) Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC microbiology 11: 45.

[57] Hu J, Li B, Kihara D (2005) Limitations and potentials of current motif discovery algorithms. Nucleic acids research 33: 4899–913.

[58] Hubbard T, Barker D, Birney E, Cameron G, Chen Y, et al. (2002) The Ensembl genome database project. Nucleic acids research 30: 38–41.

[59] Ingsriswang S, Yokwai S, Wichadakul D (2011) LinkinPath: from sequence to interconnected pathway. Bioinformatics (Oxford, England) 27: 2015–7.

[60] Isaacs AT, Jasinskiene N, Tretiakov M, Thiery I, Zettor A, et al. (2012) Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development. Proceedings of the National Academy of Sciences of the United States of America 109: E1922–30.

[61] James AA (2007) Preventing the spread of malaria and dengue fever using genetically modified mosquitoes. Journal of visualized experiments : JoVE p. 231.

[62] James S, Simmons CP, James AA (2011) Ecology. Mosquito trials. Science (New York, NY) 334: 771–2.

[63] Juhn J, James AA (2006) oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti. Insect molecular biology 15: 363–72.

[64] Kersey PJ, Staines DM, Lawson D, Kulesha E, Derwent P, et al. (2012) Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. Nucleic acids research 40: D91–7.

[65] Kim M, Lee JH, Lee SY, Kim E, Chung J (2006) Caspar, a suppressor of antibacterial immunity in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 103: 16358–63.

[66] Kokoza Va, Martin D, Mienaltowski MJ, Ahmed A, Morton CM, et al. (2001) Transcriptional regulation of the mosquito vitellogenin gene via a blood meal-triggered cascade. Gene 274: 47–65.

[67] Kramer LD, Ebel GD (2003) Dynamics of flavivirus infection in mosquitoes. Advances in virus research 60: 187–232.

[68] Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM (2008) OrthoDB: the hierarchical catalog of eukaryotic orthologs. Nucleic acids research 36: D271–5.

[69] Kuno G (2010) Early history of laboratory breeding of Aedes aegypti (Diptera: Culicidae) focusing on the origins and use of selected strains. Journal of Medical Entomology 47: 957–71.

[70] Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD, et al. (2011) Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti. Proceedings of the National Academy of Sciences of the United States of America 108: 7460–5.

[71] Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.

[72] Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, et al. (2009) VectorBase: a data resource for invertebrate vector genomics. Nucleic acids research 37: D583–7.

[73] Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic acids research 37: D229–32.

[74] Lindholm P, Saarma M (2010) Novel CDNF/MANF family of neurotrophic factors. Developmental neurobiology 70: 360–71.

[75] Liu H, Hew HC, Lu ZG, Yamaguchi T, Miki Y, et al. (2009) DNA damage signalling recruits RREB-1 to the p53 tumour suppressor promoter. The Biochemical journal 422: 543–51.

[76] Luplertlop N, Surasombatpattana P, Patramool S, Dumas E, Wasinpiyamongkol L, et al. (2011) Induction of a peptide with activity against a broad spectrum of pathogens in the Aedes aegypti salivary gland, following Infection with Dengue Virus. PLoS pathogens 7: e1001252.

[77] Macdonald WW (1963) Further Studies On A Strain Of Aedes Aegypti Susceptible To Infection With Sub-periodic Brugia Malayi. Annals of tropical medicine and parasitology 57: 452–60.

[78] Mahony S, Auron PE, Benos PV (2007) DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. PLoS computational biology 3: e61.

[79] Mahony S, Benos PV (2007) STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic acids research 35: W253–8.

[80] Marinotti O, Calvo E, Nguyen QK, Dissanayake S, Ribeiro JMC, et al. (2006) Genome-wide analysis of gene expression in adult Anopheles gambiae. Insect molecular biology 15: 1–12.

[81] Marinotti O, Nguyen QK, Calvo E, James AA, Ribeiro JMC (2005) Microarray analysis of genes showing variable expression following a blood meal in Anopheles gambiae. Insect molecular biology 14: 365–73.

[82] Markland FS, Ann M (2011) Toxins and Hemostasis. Dordrecht: Springer Netherlands.

[83] Marrelli MT, Moreira CK, Kelly D, Alphey L, Jacobs-Lorena M (2006) Mosquito transgenesis: what is the fitness cost? Trends in parasitology 22: 197–202.

[84] Mathur G, Sanchez-Vargas I, Alvarez D, Olson KE, Marinotti O, et al. (2010) Transgene-mediated suppression of dengue viruses in the salivary glands of the yellow fever mosquito, Aedes aegypti. Insect molecular biology 1.

[85] Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic acids research 34: D108–10.

[86] Mercado-Curiel RF, Black WC, Muñoz MdL (2008) A dengue receptor as possible genetic marker of vector competence in Aedes aegypti. BMC microbiology 8: 118.

[87] Meredith JM, Munks RJL, Grail W, Hurd H, Eggleston P, et al. (2006) A novel association between clustered NF-kappaB and C/EBP binding sites is required for immune regulation of mosquito Defensin genes. Insect molecular biology 15: 393–401.

[88] Miller BR, Mitchell CJ (1991) Genetic selection of a flavivirus-refractory strain of the yellow fever mosquito Aedes aegypti. The American journal of tropical medicine and hygiene 45: 399–407.

[89] Miller M (2009) The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators. Current Protein & Peptide Science 10: 244–269.

[90] Molina-Cruz A, DeJong RJ, Charles B, Gupta L, Kumar S, et al. (2008) Reactive oxygen species modulate Anopheles gambiae immunity against bacteria and Plasmodium. The Journal of biological chemistry 283: 3217–23.

[91] Moreira LA, Edwards MJ, Adhami F, Jasinskiene N, James AA, et al. (2000) Robust gut-specific gene expression in transgenic Aedes aegypti mosquitoes. Proceedings of the National Academy of Sciences of the United States of America 97: 10895–8.

[92] Nene V, Wortman JR, Lawson D, Haas B, Kodira C, et al. (2007) Genome sequence of Aedes aegypti, a major arbovirus vector. Science (New York, NY) 316: 1718–23.

[93] Nimmo DD, Alphey L, Meredith JM, Eggleston P (2006) High efficiency site-specific genetic engineering of the mosquito genome. Insect molecular biology 15: 129–36.

[94] Oliveira JHM, Gonçalves RLS, Lara FA, Dias FA, Gandara ACP, et al. (2011) Blood meal-derived heme decreases ROS levels in the midgut of Aedes aegypti and allows proliferation of intestinal microbiota. PLoS pathogens 7: e1001320.

[95] Organization WH (2009) Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition. Geneva: WHO Press.

[96] Pane A, Wehr K, Schüpbach T (2007) zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Developmental cell 12: 851–62.

[97] Pham DQD, Douglass PL, Chavez CA, Shaffer JJ (2005) Regulation of the ferritin heavy-chain homologue gene in the yellow fever mosquito, Aedes aegypti. Insect molecular biology 14: 223–36.

[98] Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ (2011) Transcriptome Profiling of Chemosensory Appendages in the Malaria Vector Anopheles gambiae Reveals Tissue- and Sex-Specific Signatures of Odor Coding. BMC genomics 12: 271.

[99] Ptitsyn AA, Reyes-Solis G, Saavedra-Rodriguez K, Betz J, Suchman EL, et al. (2011) Rhythms and synchronization patterns in gene expression in the Aedes aegypti mosquito. BMC genomics 12: 153.

[100] Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) 26: 841–2.

[101] Rai KS, Black WC (1999) Mosquito genomes: structure, organization, and evolution. Advances in genetics 41: 1–33.

[102] Rajeswari M, Jain A (2002) High Mobility Group chromosomal proteins, HMGA1 as potential tumor markers. Current Sci 82: 838–844.

[103] Rey FA (2003) Dengue virus envelope glycoprotein structure: new insight into its interactions during viral entry. Proceedings of the National Academy of Sciences of the United States of America 100: 6899–901.

[104] Ribeiro JMC (????) AegyXcel.

[105] Richardson J, Molina-Cruz A, Salazar MI, Black W (2006) Quantitative analysis of dengue-2 virus RNA during the extrinsic incubation period in individual Aedes aegypti. The American journal of tropical medicine and hygiene 74: 132–41.

[106] Richardson JE (2006) Fjoin: Simple and Efficient Computation of Feature Overlaps. Journal of computational biology : a journal of computational molecular cell biology 13: 1457–64.

[107] Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: a free, open-source system for microarray data management and analysis. BioTechniques 34: 374–8.

[108] Salazar MI, Richardson JH, Sánchez-Vargas I, Olson KE, Beaty BJ (2007) Dengue virus type 2: replication and tropisms in orally infected Aedes aegypti mosquitoes. BMC microbiology 7: 9.

[109] Samsa MM, Mondotte JA, Iglesias NG, Assunção Miranda I, Barbosa-Lima G, et al. (2009) Dengue virus capsid protein usurps lipid droplets for viral particle formation. PLoS pathogens 5: e1000632.

[110] Sanchez-Vargas I, Scott JC, Poole-Smith KB, Franz AWE, Barbosa-Solomieu V, et al. (2009) Dengue Virus Type 2 Infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. PLOS Pathogens 5: e1000299.

[111] Sandstrom DJ, Bayer CA, Fristrom JW, Restifo LL (1997) Broad-complex transcription factors regulate thoracic muscle attachment in Drosophila. Developmental biology 181: 168–85.

[112] Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nature protocols 3: 1101–8.

[113] Sethuraman N, Fraser MJ, Eggleston P, O'Brochta DA (2007) Post-integration stability of piggyBac in Aedes aegypti. Insect biochemistry and molecular biology 37: 941–51.

[114] Shore P, Sharrocks AD (1995) The MADS-Box Family of Transcription Factors. European Journal of Biochemistry 229: 1–13.

[115] Sieglaff DH, Dunn WA, Xie XS, Megy K, Marinotti O, et al. (2009) Comparative genomics allows the discovery of cis-regulatory elements in mosquitoes. Proceedings of the National Academy of Sciences of the United States of America 106: 3053–8.

[116] Sim S, Dimopoulos G (2010) Dengue virus inhibits immune responses in Aedes aegypti cells. PloS one 5: e10678.

[117] Sim S, Ramirez JL, Dimopoulos G (2012) Dengue virus infection of the Aedes aegypti salivary gland and chemosensory apparatus induces genes that modulate infection and blood-feeding behavior. PLoS pathogens 8: e1002631.

[118] Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, et al. (2007) Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature 450: 219–32.

[119] Stobbart R (1977) The control of the diuresis following a blood meal in females of the yellow fever mosquito Aedes aegypti (L). J Exp Biol 69: 53–85.

[120] Tchankouo-Nguetcheu S, Khun H, Pincet L, Roux P, Bahut M, et al. (2010) Differential protein modulation in midguts of Aedes aegypti infected with chikungunya and dengue 2 viruses. PloS one 5.

[121] Terenius O, Marinotti O, Sieglaff D, James AA (2008) Molecular genetic manipulation of vector mosquitoes. Cell host & microbe 4: 417–23.

[122] Thorpe HM, Smith MC (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proceedings of the National Academy of Sciences of the United States of America 95: 5505–10.

[123] Tompa M, Li N, Bailey TL, Church GM, De Moor B, et al. (2005) Assessing computational tools for the discovery of transcription factor binding sites. Nature biotechnology 23: 137–44.

[124] Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England) 25: 1105–11.

[125] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7: 562–78.

[126] Trapnell C, Williams Ba, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28: 511–5.

[127] Urdaneta-Marquez L, Failloux AB (2011) Population genetic structure of Aedes aegypti, the principal vector of dengue viruses. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 11: 253–61.

[128] Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, et al. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome research 19: 327–35.

[129] Wang L, Feng Z, Wang X, Wang X, Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics (Oxford, England) 26: 136–8.

[130] Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nature reviews Genetics 5: 276–87.

[131] Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, et al. (2007) Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science (New York, NY) 316: 1738–43.

[132] Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV (2013) OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic acids research 41: D358–65.

[133] Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A (1987) Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. The American journal of tropical medicine and hygiene 36: 143–52.

[134] Weaver SC, Vasilakis N (2009) Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 9: 523–40.

[135] Wise de Valdez MR, Nimmo D, Betz J, Gong HF, James Aa, et al. (2011) Genetic elimination of dengue vector mosquitoes. Proceedings of the National Academy of Sciences of the United States of America 108.

[136] Wu J, Sieglaff DH, Gervin J, Xie XS (2008) Discovering regulatory motifs in the Plasmodium genome using comparative genomics. Bioinformatics (Oxford, England) 24: 1843–9.

[137] Xi Z, Ramirez JL, Dimopoulos G (2008) The Aedes aegypti toll pathway controls dengue virus infection. PLoS pathogens 4: e1000098.

[138] Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434: 338–45.

Glossary

gene drive molecular genetic tactics that cause a linked trait or gene to spread through a population at faster rates than expected based on fitness alone; generally operating independently of natural selection and genetic drift. 13

germline the cells responsible for eventually forming the sperm or egg cells in a mature sexual organism. 8

hematophagy the characteristic of feeding on blood. 20

hemolymph A fluid analogous to blood that constitutes the circulatory system of most invertebrates. 3

incidence a measure of the frequency with which new cases of illness, injury, or other health condition occurs among a population during a specified period. 4

monophyletic having developed from a single ancestral source. 20

ϕC31 a site-specific bacteriophage recombinase that performs precise, unidirectional recombination between two sites called attB and attP. 8

population conversion a novel vector control strategy that aims to drive a transmission-refractory trait into a local vector population with the goal of preventing individuals from every becoming infectious. 6, 9

population reduction a vector control strategy that focuses on reducing the local population of a disease vector with the goal of reducing the probability that an infectious individual will have an opportunity to participate in a transmission event. 6, 9

prevalence the number or proportion of cases or events or attributes among a given population. 4

transgene a gene that has been transfered from one organism to another. 7

transposable element relatively short regions of DNA in an organism's genome that have the ability to cut (or copy) themselves from the genome and insert themselves into a new location in the genome. 7

transposon . 7, *see* transposable element

Acronyms

NFκB Nuclear Factor κ-light-chain-enhancer of activated B cells. 20, 31

20E 20-hydroxyecdysone. 55

AMP attacin. 47

CRE *cis*-regulatory element. 11, 17–21, 24, 26, 30, 35, 36

CRM *cis*-regulatory module. 11, 17, 18, 21, 24, 25

CTM Chetumal. 28, 29, 34, 39–43, 45–48

DENV dengue virus. 27–29, 31, 34–38

DENV2 dengue virus serotype 2. 39, 40, 48

DENVI dengue virus infected. 29–34

dpi days post infection. 27–33, 35

EIP extrinsic incubation period. 27, 28

FPKM fragments per kilobase of transcript per million mapped reads. 88

gFunc graph-based functional genomics framework. 58

LVP Liverpool. 33–36, 39, 40, 42, 43, 45, 46

mAP mRNA accumulation profile. 57

MDOS Motif Discovery using Orthologous Sequences. 17, 19

Mya million years ago. 18, 19

NBF non-bloodfed. 57, 67

PBM post bloodmeal. 17, 18, 22, 40, 42, 57, 67

PTCI Phylogenetic Transcriptional Correlation Index. 57, 59

PyPI Python Package Index. 58

Rex-D Rexville D Puerto Rico. 39–43, 45–48

RIDL release of insects carrying a dominant lethal phenotype. 9

RNAi RNA interference. 23

ROS reactive oxygen species. 47

SCOPE Suite for Computational identification Of Promoter Elements. 24

SIT sterile insect technique. 5, 6, 9

TFBS transcription factor binding site. 19, 55

TSS transcriptional start sites. 26

Appendix A

Methods

A.1 Comparative genomics allows the discovery of *cis*-regulatory elements in mosquitoes

A.1.1 Sequence Datasets

Orthologous gene pairs among *Culex quinquefasciatus* (genebuild CpiJ1.2), *Aedes aegypti* (genebuild AaegL1.1), *Anopheles gambiae* (genebuild AgamP3.4), and *Drosophila melanogaster* (genebuild BDGP4.3) were determined using the Ensemble Compara pipeline [128]. This pipeline is based on maximum likelihood phylogenetic gene trees built from the gene transcripts and representing the evolutionary history of gene families. Duplication or speciation events are differentiated by comparing the gene trees to the species tree. This method is analogous to the reciprocal best-hit approach in the simple case of unique orthologous genes (one-to-one orthologues). The resulting lists of genes are available from Vectorbase at http://www.vectorbase.org/Other/ComparativeAnalyses. All mosquito gene coordinates were obtained from VectorBase

and *D. melanogaster* data were from Ensembl API (*Ae. aegypti*: Ensembl 49 genebuild AaeL1.1; *An. gambiae*: Ensembl 49, genebuild Agam3.4; *Cx. quinquefasciatus*: VectorBase, genebuild CpiJ1.2; *D. melanogaster*: Ensemble 49, genebuild BDGP4.3). Repeat-masked *Cx. quinquefasciatus* sequences were obtained from VectorBase and all other genome sequences were retrieved premasked using the Ensembl perl API. The one-to-one mosquito orthologous datasets were evaluated further before using in the MDOS analyses. The pronounced intron elongation in 5′- and 3′-end UTRs resulting from the insertion of repetitive elements within these regions [92] and the presence of coding sequence incorrectly included in annotated UTRs were mitigated by only using sequences found within fragments 2 kb in length at the 5′-end of the annotated gene boundaries. Overlaps of these DNA sequences with adjacent genes were determined through use of fjoin [106] and the sequences truncated accordingly. Only sequences with a final size $\geq$ 10 base (bp) were analyzed. Pairwise comparisons were conducted with MDOS limits set for the discovery of 7-, 8-, and 9-mers.

A.1.2 Discovery of Evolutionarily Conserved Putative CREs Among Mosquitoes

Motifs receiving a conservation z-score $\geq$3 in all 3 mosquito pairwise comparisons were combined into a nonredundant list. To discover motifs with greater exclusivity within the 3 mosquitoes, the conservation z-scores for each motif in 2-kb 5′-end flanking regions of shared *D. melanogaster* orthologues also were determined. A reciprocal analysis was conducted in which 8-mers conserved in 5′-end flanking regions of one-to-one orthologs of *D. melanogaster* and each mosquito species also were determined (conservation z-score $\geq$3), followed by conservation z-scores determination of these motifs in the other 2 mosquito species. This analysis addresses the effect of the order of motif discovery, and whether the discovery process was biased by first discovering conserved

motifs in mosquitoes followed by assessment in *D. melanogaster* or vice versa. The discovered motifs were grouped by a "Familial Binding Profile" construction through use of the STAMP program [79, 78], using default settings (Metric = PCC, Alignment = SWU, Gap-open = 1,000, Gap-extend = 1,000, nonoverlap-align Multiple Alignment = IR, Tree = UPGMA). Putative identifications of the discovered motifs were determined using STAMP through comparisons to mosquito TFBS reported in the literature, with acceptable matches defined as those with E-values $< 1\times10^{-5}$ and no more than 1 mismatched nucleotide.

A.1.3 Clustering of Temporal- and Spatially Regulated *An. gambiae*

Preexisting microarray data [81, 80] were used to identify groups of genes with specific temporal- or spatial-mRNA accumulation profiles. Alignments of probe sequences to the *An. gambiae*genome (Ensembl 49) were provided by Nathan Johnson (Ensembl group, EBI). Probe-sets aligning to multiple genes or with $\geq$2 probes with more than 1 mismatch were not included. One-way ANOVA was performed to identify probe-sets (genes) with significant changes in expression with a conservative false discovery rate of 0.001 [15], followed by k-means clustering with Euclidean distance separation using open-source software (MeV MultiExperiment Viewer v4.1.01, TM4 [107]). The probe sets showing significant dynamic expression patterns following a blood meal were clustered into distinct TC groups. To further refine the expression gene/cluster assignments, probe-sets that align to the same gene were required to have a Pearson's Correlation Coefficient $\geq$0.9; otherwise, the respective gene was removed from further analysis. Expression values from 4 samples (whole-body females, midguts, fat body, and ovaries, all processed at 24 hPBM) were analyzed by one-way ANOVA. Probe-sets (genes) from each sample displaying $\geq$ 3-fold enrichment over the remaining samples

as well as having a p-value ≤ 0.05 (Tukey honest significant difference) were considered to be enriched within the respective sample.

A.1.4 Determination of Association of Mosquito Motifs Within Expression Clusters

The 5′-end flanking sequences of genes within each *An. gambiae* expression cluster were scanned for the occurrence of the mosquito motifs, and their enrichment scored using the hypergeometric distribution. The number of genes containing a particular motif in their 5′-end flanking sequences is designated K, and those occurring within a specific expression cluster, k. If the total number of 5′-end sequences analyzed is N, and the number of genes in that particular cluster is represented as n, all sequences without the motif ("negative set") would be $N - K$ and those within the sample $n - k$. The probability of observing by chance at least k matches within the cluster n can be calculated through the equation:

$$P(k) = \sum_{k'=k}^{\min n,K} \frac{\binom{K}{k'}\binom{N-K}{n-k'}}{\binom{N}{n}} \tag{A.1}$$

Distributions of p-values obtained from mosquito-motif associations with expression clusters were compared with those derived from alternative sequences. To generate alternative sequences, mosquito motifs were shuffled following 2 different procedures: the first used a translation key (A = G; G = T; T = C; C = A) to substitute the nucleotides at each position; the second produced random permutations by shuffling the order of motif constituents, maintaining the nucleotide composition.

A.2 RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti*

A.2.1 Mosquito strains and rearing

The *Ae. aegypti* Liverpool strain (LTV) used in this study originated from West Africa where it was selected for susceptibility to the filarial worm parasite, Brugia malayi [77], and has been maintained at the Liverpool School of Tropical Medicine since 1936. DNA from mosquitoes of this strain, derived after twelve consecutive generation of single-pair inbreeding, was used to generate the currently available *Ae. aegypti* genome sequence [92]. Mosquitoes were maintained at 28°C, 70-80% relative humidity, with 12-12 h light-dark photoperiod at Colorado State University (Fort Collins, Colorado). Larvae were fed on a finely-ground fish food (Tetramin, Tetra Werke, Germany). Males and females were kept together in a cage with unlimited access to water and sugar (raisin) until blood feeding. Mosquitoes aged 3-5 days after eclosion were allowed to feed on immobilized mice. The study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Female mosquitoes were flash-frozen in dry ice and promptly stored (-80°C) five hours after blood feeding and shipped to the University of California, Irvine for RNA extraction.

A.2.2 RNA extraction and Illumina library preparation

Total RNA was extracted with TRIZOL (Invitrogen) from pools of three females (3-5 days old) either exclusively kept on a sugar diet (S) or five hours after blood feeding (B).

After checking for the quality of RNA with an Agilent 2100 bioanalyzer, two samples of S and B were pooled to reach the 20 micrograms necessary for the preparation of two single-read Illumina libraries. Illumina libraries were prepared and run for 40 cycles by the Expression Analysis Core at the UC Davis Genome Center. Libraries were run at a concentration of 4-5 pM.

A.2.3 Processing of Illumina sequencing data

Sequencing data were retrieved from the UC Davis Genome Center through r-sync. Sequencing data have been deposited at the Short Read Archive (NCBI) under accession number GSE24872. Data from the two technical replicates were combined to gain sequencing depth after having verified the technical reproducibility of the two libraries generated for each condition (B and S). Bowtie [71] was used to align the Illumina reads against the *Ae. aegypti* genome (version AaegL1) [72], allowing a maximum of two mismatches and with the -m option, which returns only reads with a single best match in the genome. Reads mapping to ribosomal RNA genes were filtered out from the Bowtie output using a custom Python script. The percentage of covered transcriptome was determined using BEDTools [100]. Differential expression between conditions was assessed by the likelihood ratio test as implemented in the program DEGseq [129], after accounting for the different total gene counts of each library, at a p value of 0.001 and with a false discovery rate (FDR) of 0.1% [15]. Transcript description was based on the *Ae. aegypti* protein database AegyXcel [104].

A.2.4 Real-time quantitative RT-PCR validation of RNA-seq data

A total of 13 genes identified by RNA-seq to be expressed differentially between S and B mosquitoes were chosen for real-time quantitative PCR analysis. Total RNA was

extracted by TRIZOL (Invitrogen) from a pool of eight females kept exclusively on a sugar diet or a similar pool collected five hours after blood feeding. Following DNAse I (Invitrogen) treatment, a total of 10 µg of RNA were used for cDNA synthesis with superscript III (Invitrogen) and random primers. Real-time quantitative PCR reactions of 20 µl were performed in triplicate with SYBR Green Supermix (Biorad) and 0.3 µM of each primer on three sequential five-fold dilutions each of the original cDNA. Real-time quantitative PCR reactions were run on an iQ3 system (Biorad). No primer dimer was detected when inspecting the melting curves and primer pairs were chosen that displayed greater than 90% amplification efficiency, in all cases except AAEL002565, where efficiency was 89.313 ± 5.384. Fold-changes in gene expression between S and B mosquitoes were derived by the comparative CT method [112], using the constitutive gene rp49 (GenBank Acc. No.:AY539746; AAEL003396) as the reference and four samples each for S and B mosquitoes. Correlation between the expression values detected by RNA-seq and qRT-PCR for the 13 genes tested was estimated by calculating Spearman's Rho correlation in the JMP501 statistical software (SAS Institute INC., Cary, NC). The paired t-test in Excel was used to compare the expression values for each transcript in the two methods. The significance of the qRT-PCR-based difference in expression values between B and S mosquitoes based on four samples each for B and S were calculated using a standard t-test.

A.2.5 Splice-site predictions

The program HMMsplicer [35] followed by custom Python scripts was used to assess transcriptome plasticity. Initial HMMsplicer runs were performed separately for sugar-fed and blood-fed samples using all RNA-seq reads that passed Illumina's quality filtering, regardless of whether they aligned to the genome. Junctions were predicted initially for single reads and then combined with perfectly matching junctions

and junctions within 3 bp of each other. The combined junction inherits the location of the highest scoring junction and the combined score is adjusted appropriately. Only junctions predicting canonical splice sites after this combination were retained. Predictions for sugar-fed and blood-fed samples were combined and scores adjusted similar to above to improve the predictive power, but perfectly matching junctions were required for junctions to be combined. Finally, only junctions with more than one supporting RNA-seq read and an HMMsplicer score of 600 or greater were considered here.

A.2.6 Motif discovery

SCOPE [27] uses an ensemble method to combine the results of three specialized motif finders that separately concentrate on non-degenerate motifs, degenerate motifs and motifs that contain two separate "half-sites". It generates significance scores by combining overrepresentation, positional bias and the proportion of the co-regulated promoters to contain at least one instance of the motif. It is resistant to the common problem of extraneous or "non-informative" promoter regions included in the co-regulated set. SCOPE was run using the 2000 bp upstream of the start codon for each transcript with SCOPE's OccurrenceKSScorer to generate the significance values.

A.3 Strain Variation in the Transcriptome of the Dengue Fever Vector, *Aedes aegypti*

A.3.1 *Aedes aegypti* strains

LVP, CTM, and Rex-D mosquitoes were reared under identical laboratory conditions to prevent the effects of environmental factors on transcription. Female and male mosquitoes were kept together in cages with unlimited access to sugar (raisins) and water until blood feeding. Three- to five day-old female mosquitoes were allowed to feed on anesthetized mice. Blood-fed females were transferred to another cage, kept with access to water and sugar, and whole-animal samples harvested at 5, 8, 12, 24, or 72 hr post-bloodmeal (hPBM) and stored at $-80°$. Blood-feedings occurred between 8 and 10 AM to avoid differences in expression profiles attributable to circadian rhythms [99].

A.3.2 Weight comparisons of sugar and blood-fed mosquitoes

Three-day old females were separated into pools of 20 mosquitoes each, denied access to water for 4 hr, immobilized by CO_2, and weighed. Mosquitoes then were given access to water and sucrose until 4 hr before blood-feeding. At 24 hr after the first weight measurement, six pools of each strain were allowed to feed on anesthetized mice for 15 min. Immediately after blood-feeding, each pool was immobilized by CO_2 and reweighed. Two additional weight measurements were recorded at 30-min intervals to account for diuresis [119]. The differences in weight before and after blood-feeding and among the three strains were analyzed by a t-test, assuming equal variance.

A.3.3 RNA-seq library preparation and sequencing

Total RNA was extracted with TRIZOL (Invitrogen) from pools of three female mosquitoes either kept on a sugar diet (S) or 5 hr after blood-feeding (B), The quality of the RNA was checked on an Agilent 2100 Bioanalyzer, and two samples each of S and B mosquitoes were pooled for RNA-seq library preparation. Libraries were prepared following the standard Illumina protocol (http://www.illumina.com/products/mrna_seq_8_sample_prep_kit.ilmn) and sequenced with 40-bp single reads at the Expression Analysis Core at the UC Davis Genome Center (http://genomecenter.ucdavis.edu/expression_analysis/). Libraries were run at a nucleic acid concentration of 4 to 5 pM.

A.3.4 Reverse transcription gene amplification

Total RNA was extracted using TRIZOL (Invitrogen) from pools of 300 embryos collected 12 hr after egg deposition, 30 3rd instar larvae, 30 pupae, or five adults. Adult samples included 3- to 5-day-old male or female mosquitoes kept on a sugar diet and 3- to 5-day-old females collected 5, 24 and 72 hPBM. Total RNA from six adult samples of each of the time points was pooled in equal amounts to have a representation of 30 individuals as in the case of larvae and pupae. After DNAse I treatment (Ambion), 800 ng of RNA were used for cDNA synthesis with Superscript III (Invitrogen) and random oligonucleotide primers. Gene amplification reactions were performed with the use of 2 µl of cDNA, 0.4 µM of each transcript-specific primer, 0.2 nM each dNTP, 2 mM $MgSO_4$, 1X buffer, and 1 Unit of Platinum Taq DNA Polymerase (Invitrogen). Amplification conditions were 94° for 2 min followed by 25 to 40 cycles of 94° for 15 sec, 45 to 65° for 30 to 45 sec, 68° for 30 sec, and a final step at 68° for 6 min. Ampli-

fication products were resolved in a 2% agarose gel and stained with GelRed (Phenix Research).

A.3.5 Quantitative real-time gene amplification

Total RNA was extracted by TRIZOL (Invitrogen) from pools of eight females either kept exclusively on a sugar diet or eight females collected at 5, 8, 12, and 24 hPBM. After DNAse I (Invitrogen) treatment, 10 μg of RNA were used for cDNA synthesis with Superscript III (Invitrogen) and random primers. Real-time quantitative gene amplification reactions for 13 selected transcripts were run and analyzed as described previously [20].

A.3.6 RNA-seq data analyses

RNA-seq data for the CTM and Rex-D strains are deposited at the NCBI Gene Expression Omnibus under accession number GSE32074. RNA-seq data for the LVP strain are available under the accession number GSE24872 [20]. Sequence reads were mapped to the LVP reference genome with Bowtie [71], allowing a maximum of two mismatches and with the -m option, which returns only reads with a single best match in the genome [20].

Differential transcript accumulation levels among conditions were assessed by the likelihood ratio test implemented in DEGseq [129], after accounting for the different total gene counts of each library, and at a P value of 0.001 with a false discovery rate of 0.1% [15]. Use of the modifier “significant” in the following text implies that accumulation values met the statistical criteria.

The reference file containing transcript annotation information and used in DEGseq was generated by converting a GTF annotation file obtained through http://metazoa.ensembl.org/ to the format (refflat) accepted by DEGseq and represents gene-build AaegL1.2. Function parent attribution of the sequenced transcripts is based on the *Ae. aegypti* database AegyXcel (http://exon.niaid.nih.gov/transcriptome.html#aegyxcel). Transcripts accumulated differentially among samples were visualized in the *Ae. aegypti* protein network [51] with the use of Cytoscape_v2.7.0 (http://www.cytoscape.org/). Metabolic pathways were analyzed by LinkinPath [59], which accepts a maximum of 5000 sequences per session and the color of each group of sequences is assigned automatically.

A.4 Complex modulation of the *Aedes aegypti* transcriptome in response to dengue virus infection

A.4.1 Mosquitoes

The *Ae. aegypti* Chetumal (CTM) strain derives from mosquitoes collected in Chetumal (Yucatan Peninsula, Mexico) in 2005 [17]. Mosquitoes were maintained at 28°C, 70-80% relative humidity, with 12-12 h light-dark photoperiod at Colorado State University (CSU; Fort Collins, Colorado). Larvae were fed on finely-ground fish food (Tetramin, Tetra Werke, Germany). Adult males and females were kept together in cages with unlimited access to water and a sugar source (raisins) until blood feeding. Defibrinated sheep blood (Colorado Serum Company, Denver, CO) was used for artificial blood meals.

A.4.2 Virus and Mosquito Oral Infection

A DENV-2 strain, Jam1409, of the American-Asian genotype isolated in Jamaica in 1983, was used in this study [34]. DENV-2 Jam1409 was cultured in *Ae. albopictus* C6/36 cells and high passage (>25 passages) viruses were used in oral challenges [108]. Specifically, 350 and 330 adult females were fed either a DENV2-infected meal diluted 1:1 in sheep's blood or uninfected C6/36 cell culture medium diluted 1:1 in sheep's blood, respectively. Blood meals had a viral titer of 7.9 log10 pfu/ml. 'DENVI' or 'B' designate mosquitoes fed either the infected or uninfected blood meal, respectively. After blood feeding, 20 DENVI mosquitoes were sacrificed and viral titers determined for each individual using a standard method [56]. Specifically, mosquito bodies were homogenized in 270 μl of Dulbecco's Modified Eagle Medium (DMEM) and then centrifuged to eliminate large debris particles. The supernatant was further filtered and used in serial dilutions to infect monolayers of Vero cells. The lowest concentration infecting Vero cells was used to calculate the viral titer of DENVI mosquitoes. Prevalence of infection was 65% and the median viral titer in infected mosquitoes was 2.2 log10 pfu per mosquito. Samples of 20-30 mosquitoes were removed at 1, 4 and 14 days post infection (dpi), and midguts were dissected from the carcasses. Midguts, salivary glands and corresponding carcasses were collected at 14 dpi. The time points of 1, 4 and 14 dpi were chosen as representing an early time point post-infection when viral particles are confined within the midguts, a time point during the dissemination out of the midgut and a time point within the peak of viral titers in the salivary glands, respectively [108]. Midguts and carcasses were collected individually for DENVI and in pools of 10 for B mosquitoes. Salivary glands were collected in pools of 20. Oral challenges and tissue dissections were conducted in a BSL-3 facility at CSU. All dissected samples were placed in 50 μl of 1:1 PBS:Trizol, frozen in dry ice and shipped to the University of California, Irvine, for RNA extraction.

A.4.3 RNA-seq Library Preparation and Sequencing

Total RNA was extracted with TRIZOL (Invitrogen) from pools of 5-10 midguts, 20 salivary glands and corresponding carcasses. The quality of the RNA was checked on an Agilent 2100 Bioanalyser and RNA from a total of 20-40 mosquitoes was pooled in equal amounts for RNA-seq library preparation. Illumina single-end RNA-seq libraries were prepared and sequenced for 40 cycles by the Expression Analysis Core at the UC Davis Genome Center. The need for biological replicates was mitigated as done previously by pooling samples of a large number of mosquitoes for each library [98]. Libraries were run at a concentration of 4-5 pM. Evidence for a correlation in transcript accumulation levels as estimated by RNA-seq data and quantitative real-time polymerase chain reaction (qRT-PCR) was provided previously [22].

A.4.4 Data Analyses

RNA-seq data are deposited at NCBI's Sequence Read Archive (SRA) under accession number SRA058076. Reads were mapped to the Liverpool reference transcriptome (gene build AaegL1. 2) with the splice aligner TopHat, imposing a fragment length of 20 [124]. The accumulation levels of poly-adenylated RNAs were assumed to reflect changes in transcriptional activity of the corresponding genes and quantified in terms of mean number of reads per gene by DESeq [6], and in terms of fragments per kilobase of transcript per million mapped reads (FPKM) using Cufflinks [126]. FPKM accounts for gene/transcript length and allows more accurate comparisons of accumulation levels across genes/transcripts within a sample.

Statistical significance in differential accumulation of poly-adenylated RNAs between DENVI and B mosquitoes at each time point/tissue was assessed conservatively by

two different programs: DESeq and Cuffdiff within Cufflinks [6, 126]. DESeq requires the raw number of reads as input. As a consequence, DESeq was run at the gene level to avoid counting multiple times the reads mapping to exons shared by different transcript isoforms. Cuffdiff output comprises significance in accumulation levels of poly-adenylated RNAs both at the gene and transcript isoform level. Significance in differential gene product representation may not correspond unequivocally to significance at the transcript level because many genes encode a number of transcript isoforms. As a consequence, comparison of results between DESeq and Cuffdiff were done at the gene level. DESeq was run with the "blind" method, the "fit-only" sharing mode and the "local" fit type. Cuffdiff was run requiring a minimum alignment count of 10 and the upper-quartile-norm option, which can improve robustness of differential accumulation calls for less abundant transcripts. Both DESeq and Cuffdiff assess significance in differential accumulation fitting data to a negative binomial distribution, a method shown to maintain control of type-I error and that represents a statistical improvement with respect to the likelihood ratio test we applied previously [22, 6]. As a consequence, differences are expected in transcript quantification levels and significance in differential accumulation between CTM-B and -DENVI mosquitoes of this study and CTM-sugar fed and CTM-B mosquitoes at 5 hours post blood-meal [22]. Gene description used the biomart function of the EnsemblMetazoa (http://metazoa.ensembl.org/index.html) and Immunodb (http://cegg.unige.ch/Insecta/immunodb) [64, 131].

The motif-based sequence analyses suite MEME [11] was used to investigate the 2000 base-pairs (bp) adjacent to the 5′-end of ATG start-codon (hereafter designated "promoters") of selected genes. MEME was run imposing 10 motifs, ranging from 6-20 bp. Each identified motif was analyzed for potential transcription-factor binding sites based on the hypothesis that sequence similarity reflects functional similarity. STAMP,

a web tool for analyzing DNA-binding motif similarity, was used to search the closest match for each motif in the STAMP-supported transcription-factor databases [79].

Appendix B

Code