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

add copy-write text to preliminaries

© 2013 William Augustine Dunn

DEDICATION

write dedications

(Optional dedication page)
To ...

TABLE OF CONTENTS

Page

LIST OF FIGURES **v**

LIST OF TABLES **vi**

ACKNOWLEDGMENTS **vii**

CURRICULUM VITAE **viii**

ABSTRACT OF THE DISSERTATION **x**

Todo list **1**

1 Introduction **3**
1.1 Dengue fever and malaria are mosquito borne diseases 3
1.2 Hematophagy (bloodfeeding) is an essential part of mosquito-borne disease 4
1.2.1 Transmission cycle 4
1.3 Public health: *multiple targets* 6
1.3.1 Addressing the human side of the cycle 7
1.3.2 Addressing the vector side of the cycle (conventionally) 7
1.3.3 Transgenesis provides new vector control strategy 8
1.4 Elements of a successful transgenic vector control product 9
1.4.1 Introducing heritable transgenes into mosquito genomes . . . 9
1.4.2 Effector genes 10
1.4.3 Controlling when and where the effector genes are expressed 13
1.4.4 Achieving swift, dramatic, and clinically meaningful effects on the native vector population 15
1.4.5 Conservation of bloodmeal-regulated transcription 16
1.5 Evolutionary relationship between mosquitoes and hematophagy . . 17
1.5.1 Implications for design of transgenic vector control systems? . 18

2 Development of Analytical Approach **19**
2.1 Comparative genomics predicts of *cis*-regulatory elements in mosquitoes [70] 19

2.2 RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti* [12] 25
2.2.1 Transcripts found exclusively in bloodfed mosquitoes 26
2.2.2 *Cis*-regulatory element discovery 27
2.2.3 RNA-seq identifies annotation corrections 28
2.3 Strain variation in the transcriptome of the dengue fever vector, *Aedes aegypti* 32
2.4 Complex modulation of the *Aedes aegypti*transcriptome in response to dengue virus infection 34

3 Approach and Custom Software solutions **35**
3.1 Purpose 35
3.2 The phylogenetic transcriptional correspondence index (PTCI) 36
3.3 Blacktie RNA-seq pipeline 39

4 Analysis of Conservation of Midgut Transcriptional Response to Bloodfeeding in Four Vector Mosquitoes **41**
4.1 Background 41

5 Contributions to Literature Not Otherwise Mentioned **48**
5.1 Purpose 48

6 Conclusions **49**
6.1 Purpose 49

Bibliography **50**

A Methods **62**
A.1 Comparative genomics allows the discovery of *cis*-regulatory elements in mosquitoes 62
A.1.1 Sequence Datasets 62
A.1.2 Discovery of Evolutionarily Conserved Putative CREs Among Mosquitoes 63
A.1.3 Clustering of Temporal- and Spatially Regulated *An. gambiae* . 64
A.1.4 Determination of Association of Mosquito Motifs Within Expression Clusters 65
A.2 RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti* 66
A.2.1 Mosquito strains and rearing 66
A.2.2 RNA extraction and Illumina library preparation 67
A.2.3 Processing of Illumina sequencing data 67
A.2.4 Real-time quantitative RT-PCR validation of RNA-seq data . . . 68
A.2.5 Splice-site predictions 69
A.2.6 Motif discovery 69

A.3 Strain Variation in the Transcriptome of the Dengue Fever Vector, *Aedes aegypti* 70
A.3.1 *Aedes aegypti* strains 70
A.3.2 Weight comparisons of sugar and blood-fed mosquitoes 70
A.3.3 RNA-seq library preparation and sequencing 71
A.3.4 Reverse transcription gene amplification 71
A.3.5 Quantitative real-time gene amplification 72
A.3.6 RNA-seq data analyses 72
A.4 Complex modulation of the *Aedes aegypti*transcriptome in response to dengue virus infection 74
A.4.1 Mosquitoes 74
A.4.2 Virus and Mosquito Oral Infection 74
A.4.3 RNA-seq Library Preparation and Sequencing 75
A.4.4 Data Analyses 76

B Code **78**

LIST OF FIGURES

Page

1.1 Cross section of a female mosquito 5
1.2 *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti* 11
1.3 Elements of transcriptional regulation 14

2.1 Phylogenetic relationships between four vector mosquitoes 20
2.2 Associations of mosquito motifs with gene expression profiles in *An. gambiae* 21
2.3 Examples of amendments to the *Ae. aegypti* annotation supported by HMMSplicer results 30
2.4 Motif map of putative CREs from transcripts detected only in bloodfed female *Ae. aegypti* 31
2.5 Fatty acid biosynthesis and elongation in mitochondria 33
2.6 MEME analysis of nine genes with $\mathrm{FPKM}_{\mathrm{DENVI}} \geq 100$ in carcasses and salivary glands at 14 dpi 34

3.1 Exploring the parameter space of the expression and TFBS components of the PTCI 37
3.2 Graph model used to integrate data types 38
3.3 Diagram of Abbreviated Tophat/Cufflinks Inputs and Outputs 40

4.1 Pairwise PTCI results 42
4.2 Mean PTCI results 43
4.3 *An. gambiae* clustered abundance profiles 44
4.4 Orthologs of cluster 6 46
4.5 Orthologs of cluster 19 47

LIST OF TABLES

Page

1.1 Primary geographic vectors of dengue fever and malaria 4

2.1 Predicted novel junctions within various symmetric distance-windows from annotated transcripts 29

ACKNOWLEDGMENTS

write 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

in-line my CV document and match styles

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

write abstract

The abstract of your contribution goes here.

Todo list

add copy-write text to preliminaries ii

write dedications ii

write acknowledgments vii

in-line my CV document and match styles viii

write abstract x

replace CITEME tags in Introduction 3

expand how the evolutionary relationship between hematophagy and mosquitoes may be useful 17

write the implications your work will have on designing transgenic vector control systems 18

replace CITEME tags in Development of Analytical Approach 19

import text for STRAIN VARIATION 32

import text for COMPLEX MODULATION OF AEDES TXOME 34

replace CITEME tags in Approach and Custom Software solutions 35

MEGA: you need to write most of this chapter 35

replace CITEME tags in Analysis of Conservation of Midgut Transcriptional Response to Bloodfeeding in Four Vector Mosquitoes 41

write all the figures 41

MEGA: you need to write most of this chapter 41

Finish Fig 4.4 46

Finish Fig 4.5 47

decide whether this gets included 48

replace CITEME tags in Introduction 49

write all the figures 49

MEGA: you need to write most of this chapter 49

Chapter 1

Introduction

replace CITEME tags in 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).

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

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.

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.

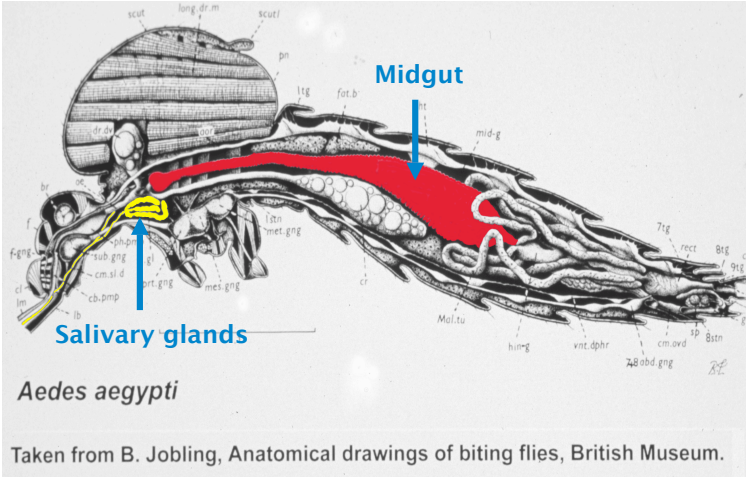


Figure 1.1: Cross section of a female mosquito with tissues important to the transmission cycle highlighted.
CITEME

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

[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).

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. Additionally, 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), spray-

ing 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).

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 [17].

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 con-

ceptualized 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 number of copies at any **single** genomic location was unpredictable [2, 69]. 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* [74, 57, 38], 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.

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 design-

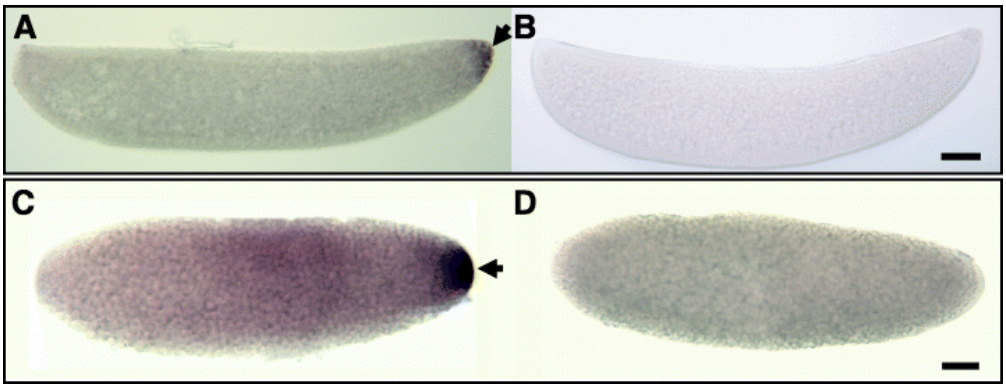


Figure 1.2: *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti*:
Hybridizations *in situ* of *Anga osk* and *Aeae osk* RNA probes to whole-mount *An. gambiae* and *Ae. aegypti* embryos. Embryos are orientated with anterior on the left. *An. gambiae* embryos (A,B) 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 [40]

ing an anti-transmission phenotype: population reduction and population conversion.

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 [82, 7, 28].

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

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 [30, 53]. 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.

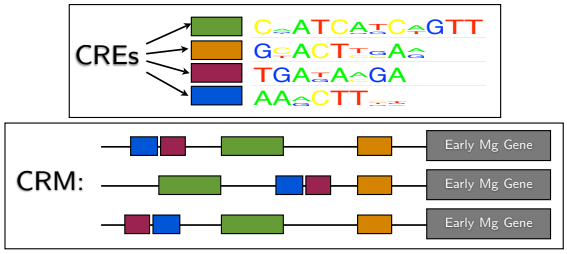
1.4.3 Controlling when and where the effector genes are expressed

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.

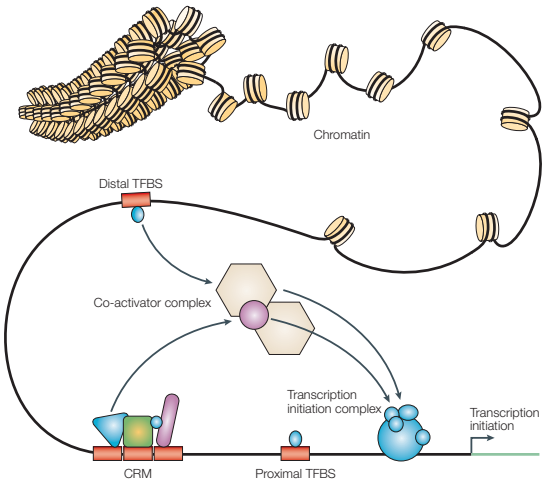
The region of DNA directly before the sequence of the gene is a major determinant of the pattern of expression and is referred to as the promoter. It influences when the gene will be turned on (e.g., directly following the ingestion of a bloodmeal), and in which tissue type (e.g., the midgut). Promoters 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

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

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) CREs make up CRMs



(B) One model of transcriptional regulation CITEME

Figure 1.3: Elements of transcriptional regulation

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 [55], the promoter sequences of a gene called carboxypeptidase (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 [55]:

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:

[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.

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.

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 [51]

expand how the evolutionary relationship between hematophagy and mosquitoes may be useful

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?

write the implications your work will have on designing 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.

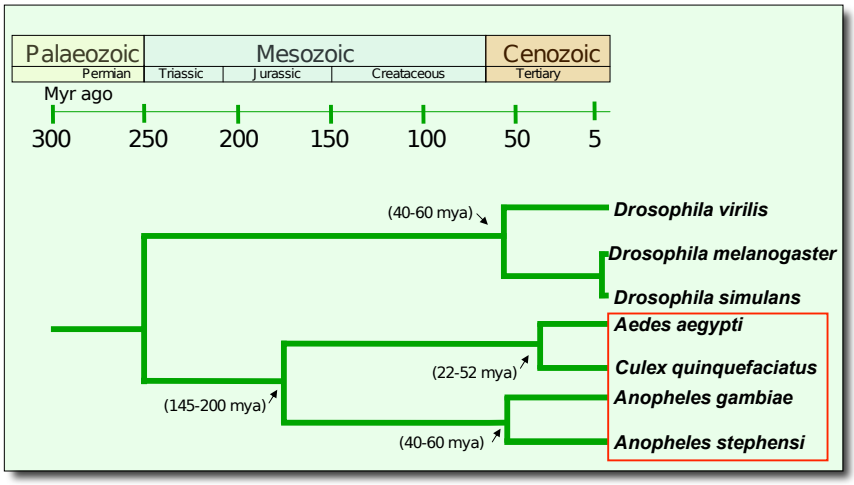
Chapter 2

Development of Analytical Approach

replace CITEME tags in Development of Analytical Approach

2.1 Comparative genomics predicts of *cis*-regulatory elements in mosquitoes [70]

The role CREs play in regulating gene expression during development is well-established [23], 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 [22, 35, 75, 80]. 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



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:
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.
Adapted from [70]

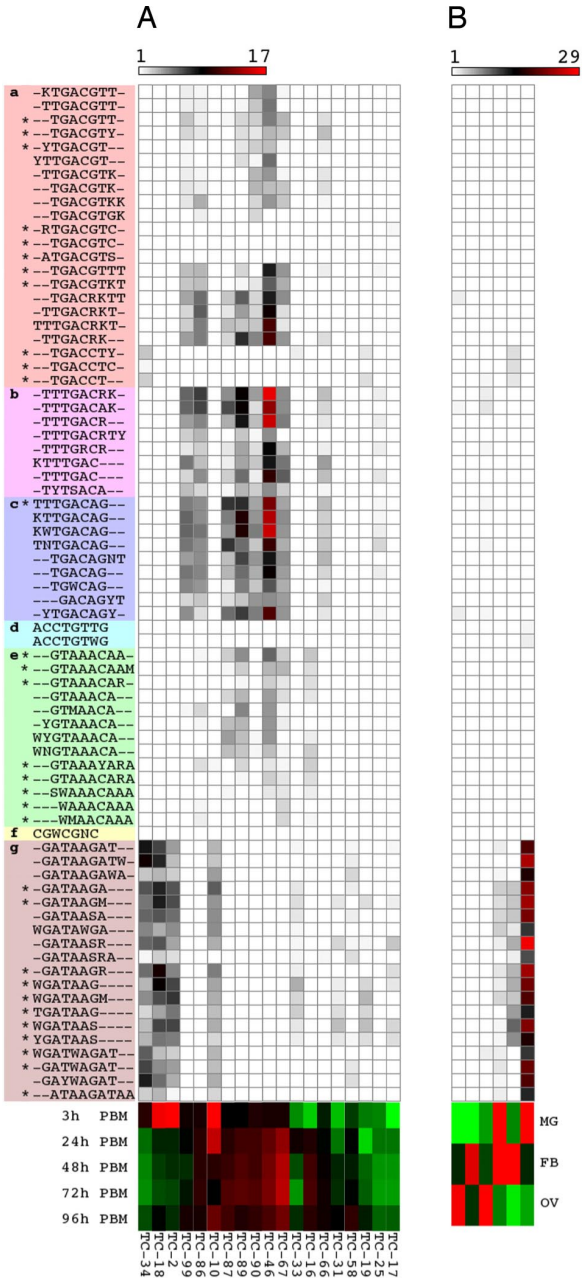


Figure 2.2: Associations of mosquito motifs with gene expression profiles in *An. gambiae*:
Motif enrichment within (A) 5′-end flanking regions of genes in clusters responsive to blood meal ingestion, and in (B) 5′-end flanking regions of genes in clusters enriched in selected tissues. The significance 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 [70]

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 [27, 71, 84] 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 million years ago (Mya) [71] (Figure 2.1)], in which orthologous intergenic sequences are aligned more easily. The work represented here includes mosquito species with phylogenetic relationships spanning 20 to 200 Mya with genomic comparisons complicated by large amounts of dispersed repetitive elements [56].

The Motif Discovery using Orthologous Sequences (MDOS) algorithm [83] 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 putative 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 site (TFBS) that are identical to those whose function was established experimentally [42, 20, 5, 3, 59, 31, 26, 54] 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 [33]. 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 this study require a blood meal 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 [63], but it is hypothesized that the occurrence of this trait in the larger clade is monophyletic [15, 18]. Most of the discovered mosquito motifs are associated with blood meal-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 transcription product accumulation within tissues responsible for blood meal 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.

This study and similar genome-wide approaches to identify putative CREs in mosquitoes [33] 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. Promoters that direct expression of transgenes specifically to the germ cells would be useful for the development of gene-drive mechanisms for spreading a desired (pathogen-resistance) trait in a mosquito population [1]. Gene-specific knockdown or robust expression of exogenous odorant receptors or odorant binding proteins in the antenna of anthropophilic mosquitoes could redirect their preferences toward other animals [85]. Targeted expression of antipathogen effector genes in the midguts, salivary glands, and hemolymph (via fat body-specific control DNA), 3 sites of interaction of most pathogens with their insect vectors, could reduce mean intensities of infection to zero, preventing pathogen transmission and disease [39]. 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 will 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* [12]

Bonizzoni and Dunn [12] provides a detailed examination of the changes in transcripts accumulation occurring at the whole-body level of *Ae. aegypti* females 5 hours post bloodmeal (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 blood fed mosquitoes. This finding supports the hypothesis that there may be a gap in immunity following a bloodmeal. Reduced expression of immune genes in blood fed 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 [67]. 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* [9, 11]. The quantity and quality of data generated by RNA-seq technology makes this 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 sugar-fed (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 [73]. 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 [16]. 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 paralogous 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 [29]. 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 [58].

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 [52, 52]. 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 SCOPE [19] 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 sugar-fed mosquitoes but detected at significant levels in the RNA-seq libraries from blood-fed mosquitoes and identified a redundant list of 22 motifs that are enriched significantly in these sequences (Figure 2.4). 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* paralogues of the G12 gene of *An. gambiae* (AGAP006187) (Figure 2.4). *Ae. aegypti* has 17 G12 genes, many more relative to other insects, which have 4.5 on average (according to OrthoDB; group EOG95TCTG) [43]. The transcripts of nine of the G12 paralogues 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 2.4). 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* [70].

2.2.3 RNA-seq identifies annotation corrections

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 [70] to compare junctions revealed by our data to the annotation provided by Vectorbase and Ensembl [45, 36]. 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 un-annotated alternatively-spliced transcripts. To estimate how many of the remaining splice junctions might be truly novel, we mapped them to increasingly larger DNA fragments 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 [36]. Figure 2.3 highlights three determined amendments to the current annotation, all supported by EST data. Figure 2.3 (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 2.3 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 2.3 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.

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 [12]

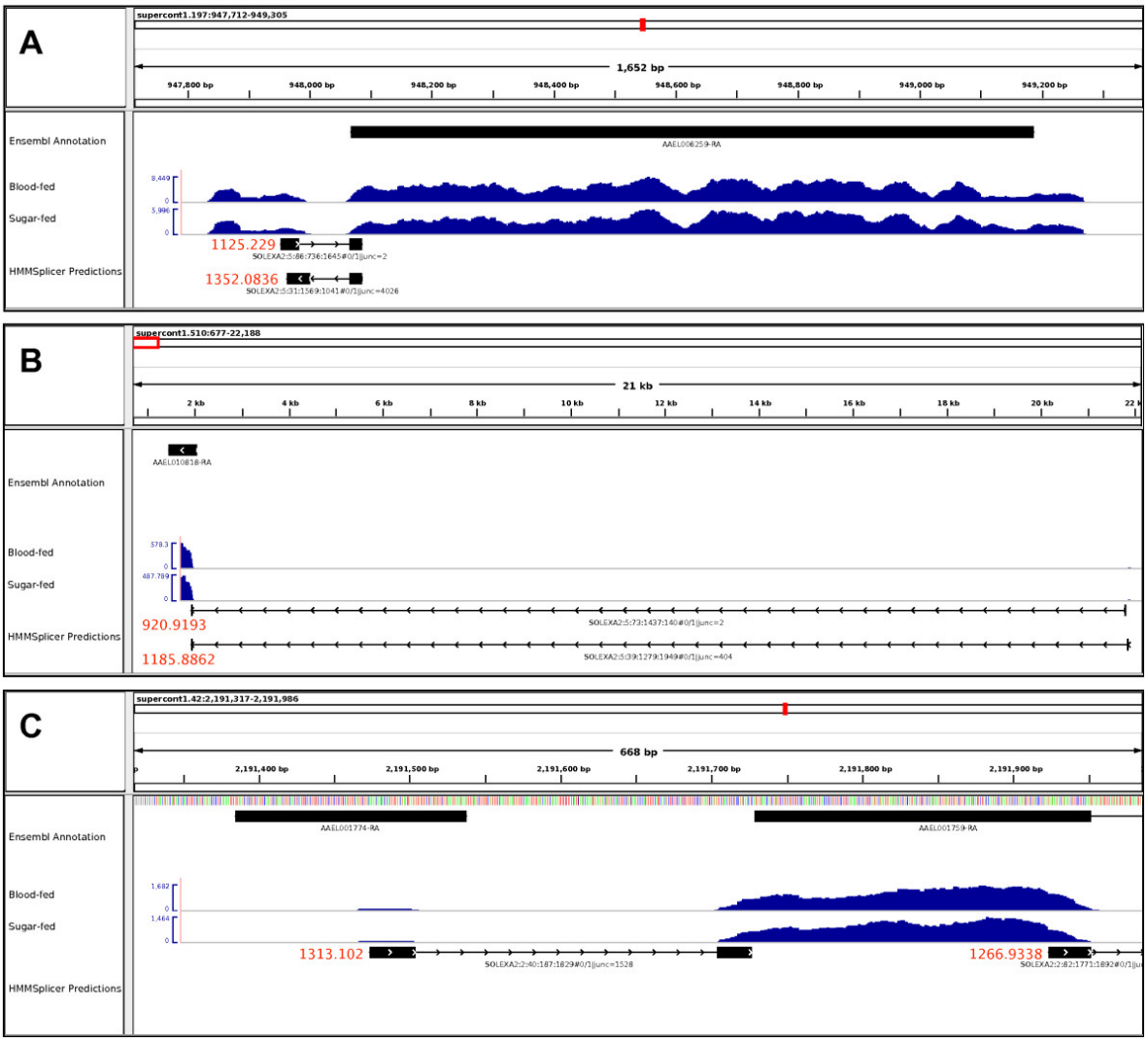


Figure 2.3: Examples of amendments to the *Ae. aegypti* annotation supported by HMMSplicer results: Black bars in the top tracks represent the current gene annotations. Blue histograms in the second track represent the non-normalized coverage of RNA-seq reads at each position. The range of the histogram 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 [12]

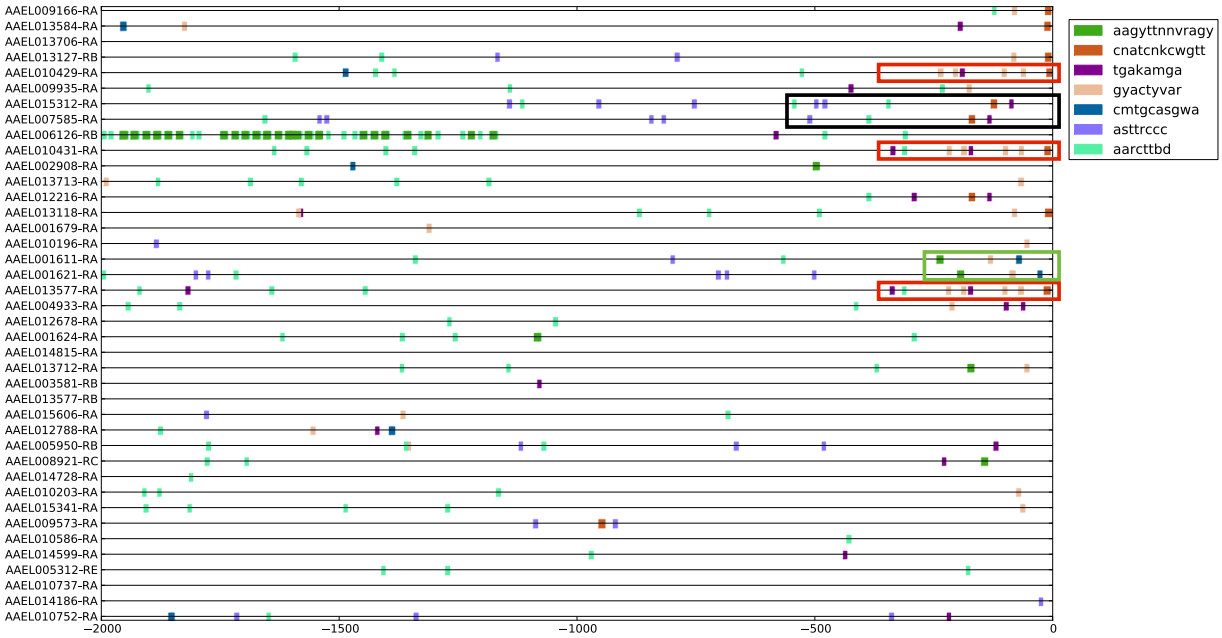


Figure 2.4: Motif map of putative CREs discovered by SCOPE using transcripts detected significantly only in bloodfed female *Ae. aegypti*: Locations of representative SCOPE-derived CRE motifs in the 2000 bp upstream of the annotated translational start site in the 40 transcripts detected significantly 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 [12]

2.3 Strain variation in the transcriptome of the dengue fever vector, *Aedes aegypti*

import text for STRAIN VARIATION

Figure 2.5

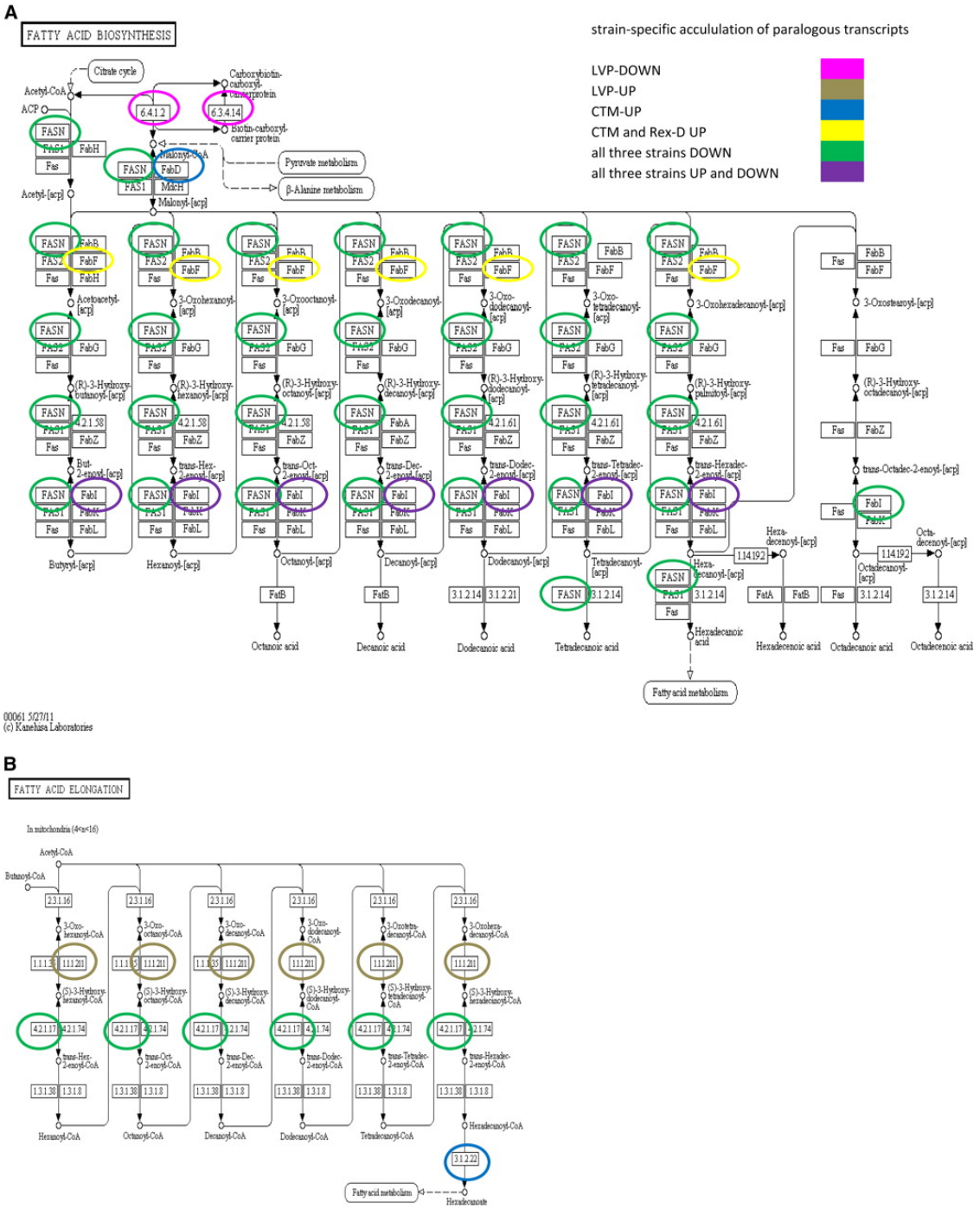


Figure 2.5: Fatty acid biosynthesis and elongation in mitochondria: Proteins corresponding to transcripts accumulated in a strain-specific manner are circled in a color corresponding to the strain and condition defined in the legend (panels A and B). Excerpted from [14]

2.4 Complex modulation of the *Aedes aegypti*transcriptome in response to dengue virus infection

import text for COMPLEX MODULATION OF AEDES TXOME

Figure 2.6

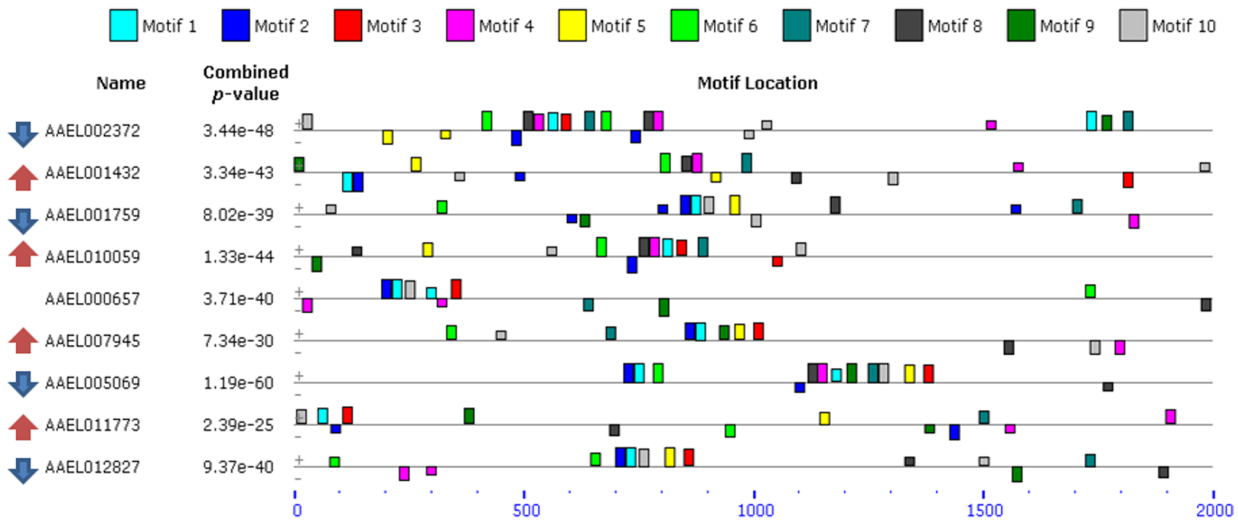


Figure 2.6: MEME analysis of nine genes with $FPKM_{DENVI} \geq 100$ in carcasses and salivary glands at 14 dpi These genes also were identified with transcripts exhibiting significant differential accumulation in analyses of salivary gland samples of the Liverpool strain infected with DEV2 Thailand 16881 [26]. 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 [26]. Distances in base-pairs are provided below the schematic of each gene. doi:10.1371/journal.pone.0050512.g004
Excerpted from [13]

Chapter 3

Approach and Custom Software solutions

replace CITEME tags in Approach and Custom Software solutions

3.1 Purpose

Illustrate how lessons learned in previous chapter (as well as new lessons introduced here) are addressed through custom software solutions.

1. RNA-seq analysis, reproducibility, updating
2. Integration of multiple disparate Omics scale data types
3. Documentation of exactly how analyses were carried out (IPython notebooks)

MEGA: you need to write most of this chapter

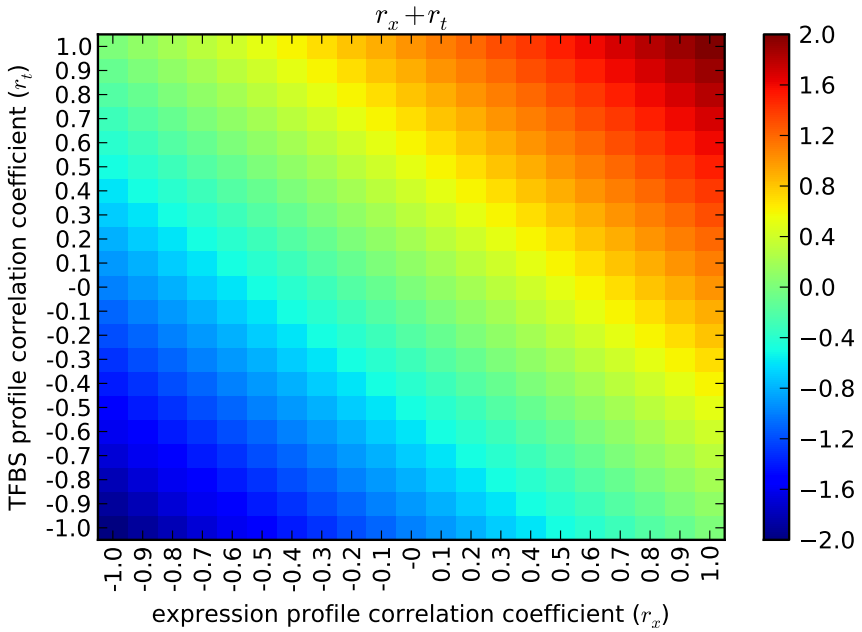
3.2 The phylogenetic transcriptional correspondence index (PTCI)

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

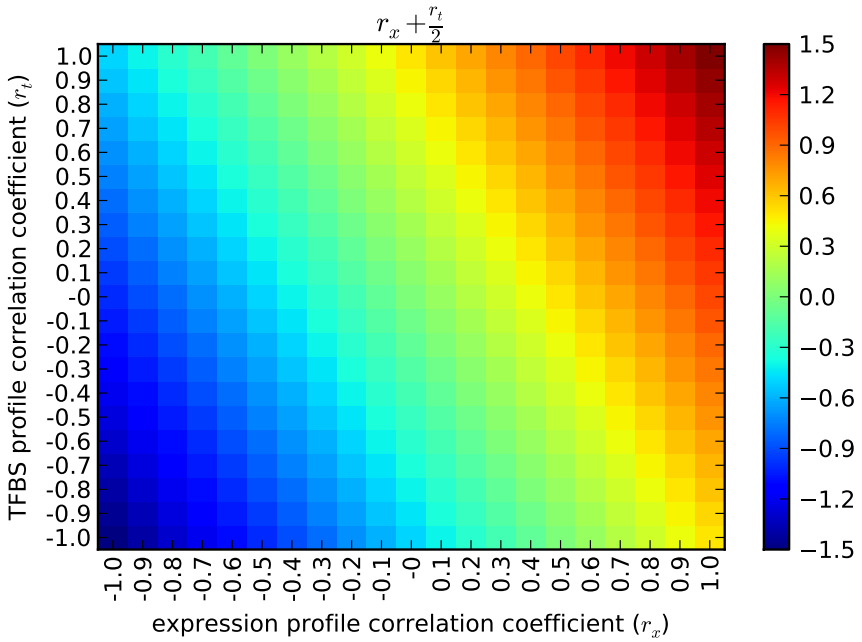
Here is the PTCI calculation:

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

The graph model is useful (Figure 3.2)

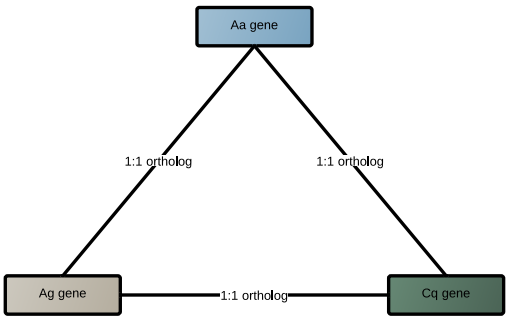


(A)

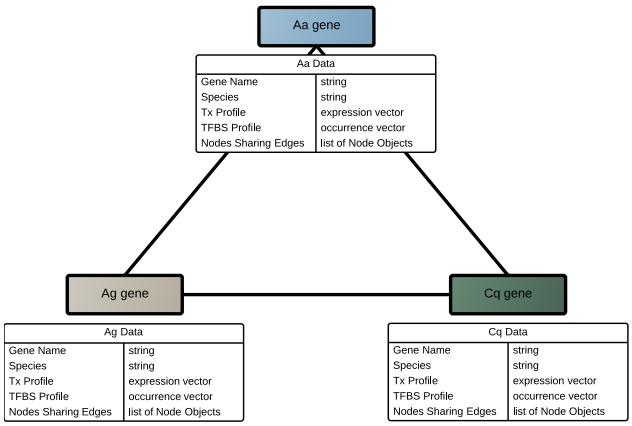


(B)

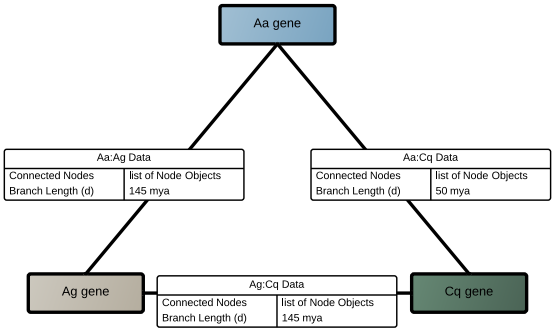
Figure 3.1: Exploring the parameter space of the expression and TFBS components of the PTCI:
(A) The expression and TFBS profile correlations each carry the same weight. TFBS data is not strictly empirical, and position weight matrix models inherently are prone to produce false positive predictions. 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.



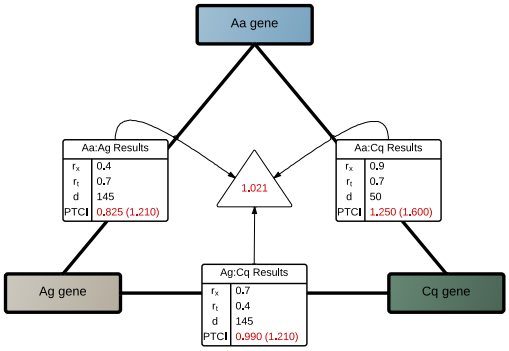
(A) N-way 1:1 ortholog graph structure



(B) Node data model



(C) Edge data model



(D) Example PTCI data

Figure 3.2: Graph model used to integrate data types

3.3 Blacktie RNA-seq pipeline

See Figure 3.3

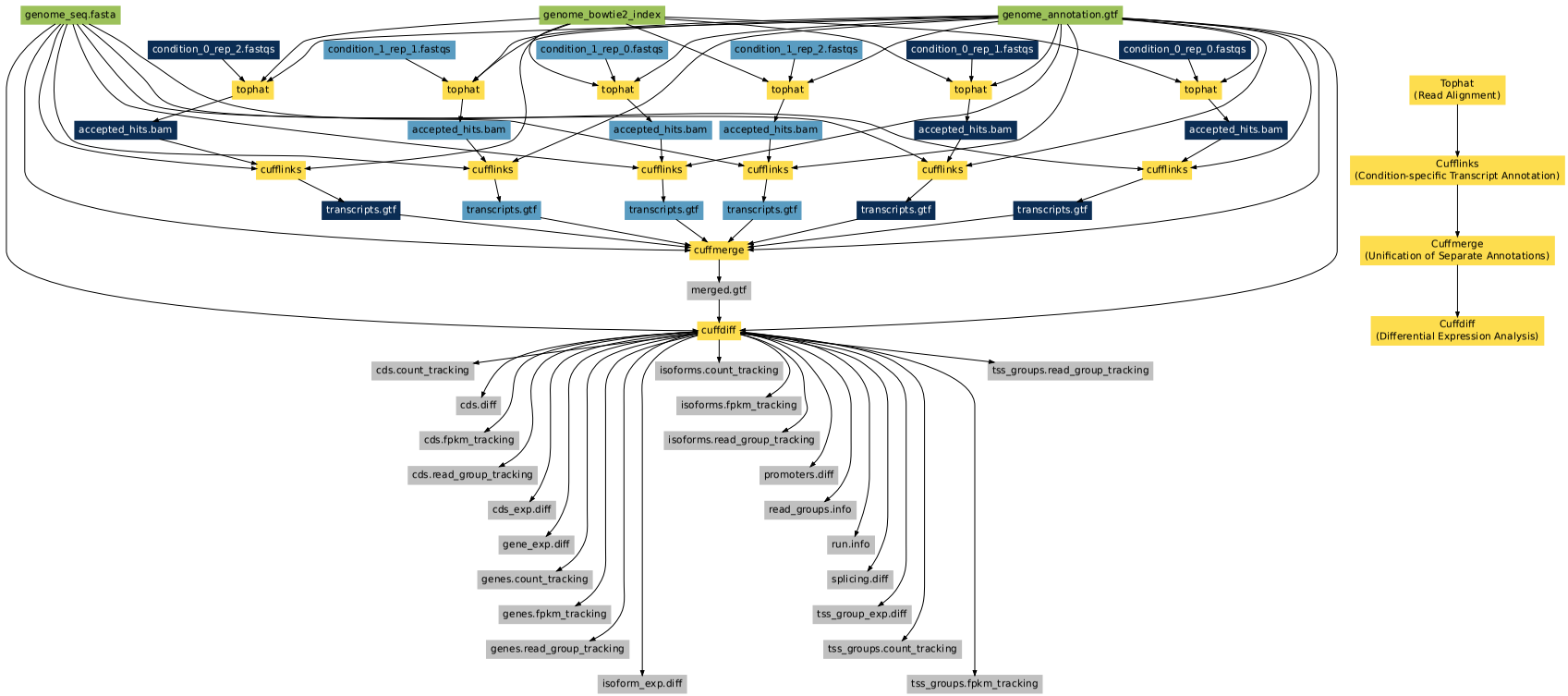


Figure 3.3: Diagram of Abbreviated Tophat/Cufflinks Inputs and Outputs:
This figure demonstrates the complexities of a typical RNA-seq experiment as analyzed with the "Tuxedo Protocol". It models a fairly **simple** two condition experiment with each condition having three replications. It is "abbreviated" 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

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

write all the figures

MEGA: you need to write most of this chapter

4.1 Background

1. Biological importance of bloodmeal to mosquito

2. Biological importance of bloodmeal to transmission cyc

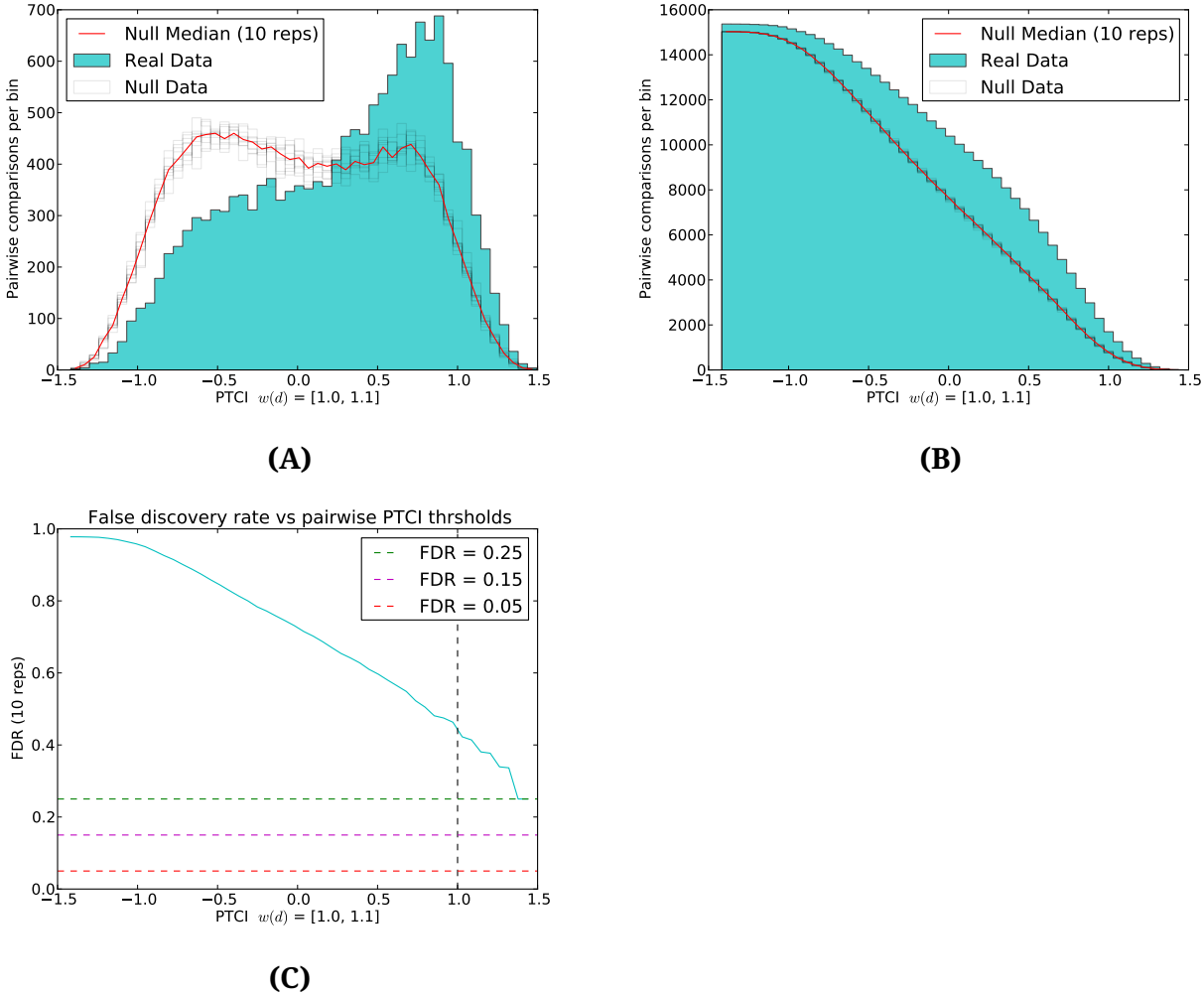


Figure 4.1: Pairwise PTCI results:
(A) Histogram of pairwise PTCI results vs null distributions. **(B)** Reverse Cumulative histogram of pairwise PTCI results vs null distributions. **(C)** False discover rate vs PTCI threshold.

Figure 4.1

Figure 4.2

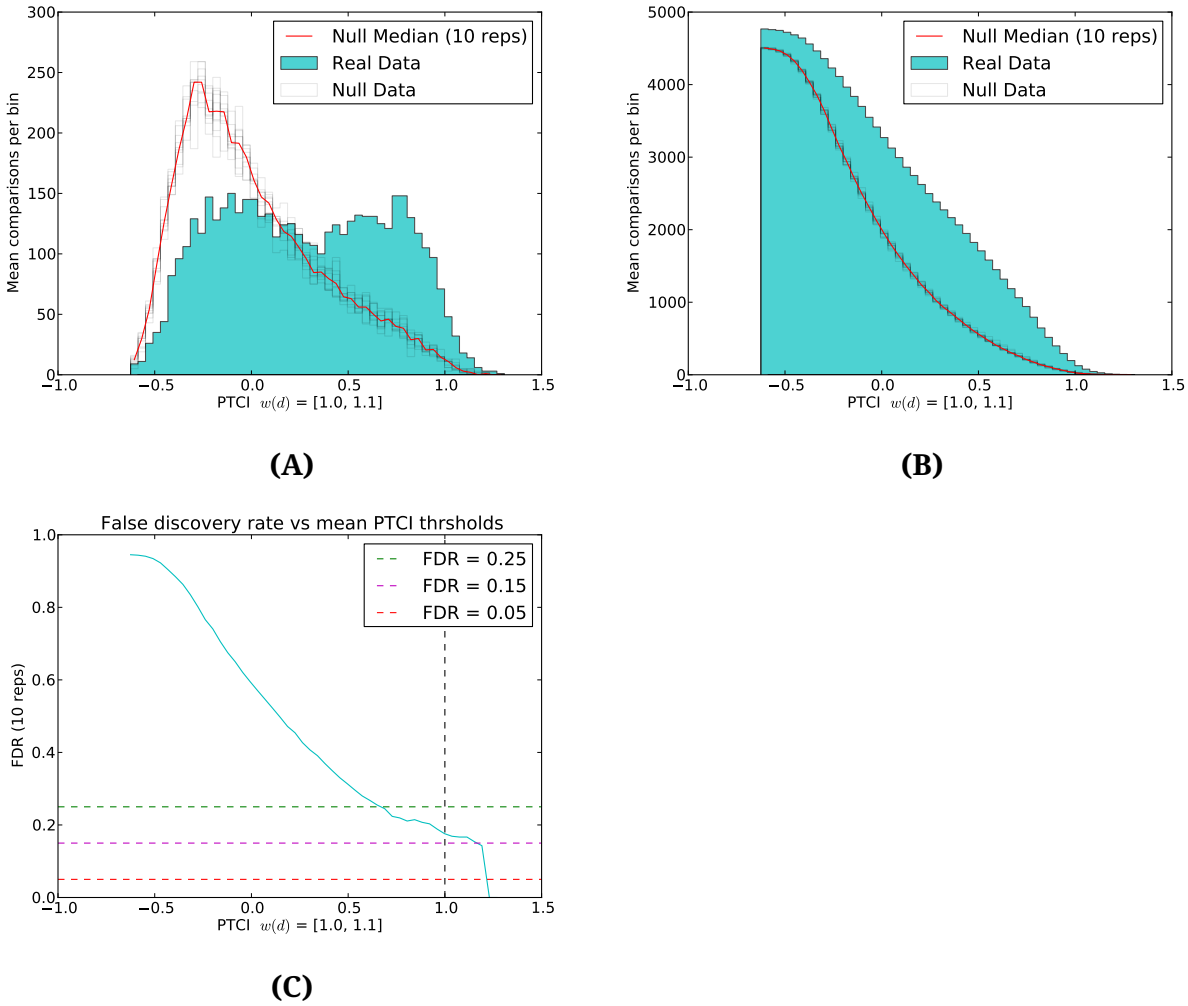


Figure 4.2: Mean PTCI results:
(A) Histogram of mean PTCI results vs null distributions. **(B)** Reverse Cumulative histogram of mean PTCI results vs null distributions. **(C)** False discover rate vs PTCI threshold.

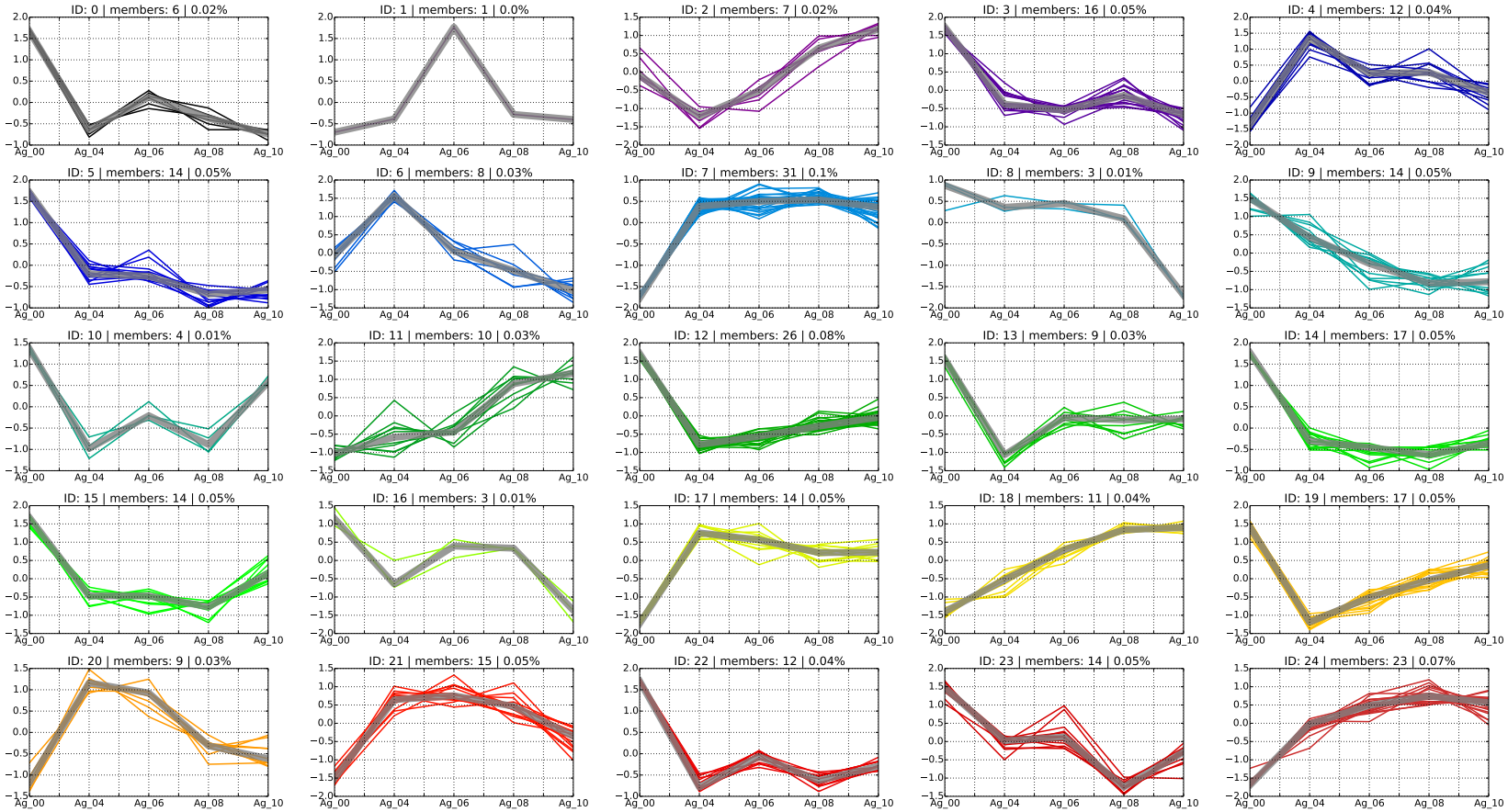
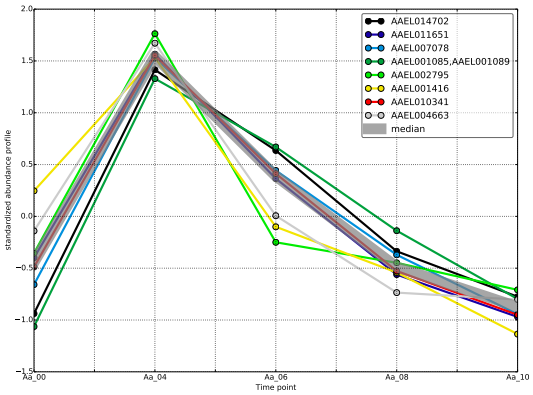


Figure 4.3: *An. gambiae* clustered abundance profiles.
Clusters were generated using k-means clustering as implemented in Biopython version 1.62 [21]. Abundance profile data was log transformed after one FPKM was added to all data to remove zeros ($log_{10}(\mathrm{FPKM}+1)$). K-means was then applied 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.

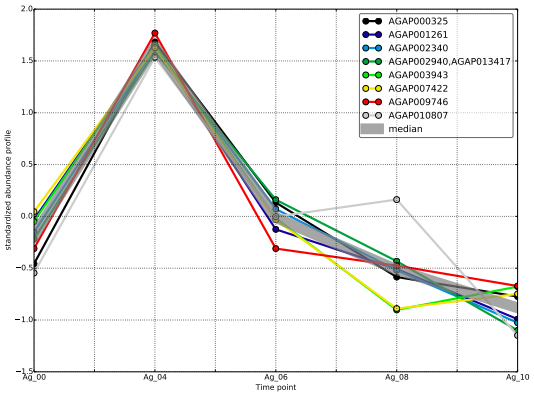
Figure 4.3

Figure 4.4

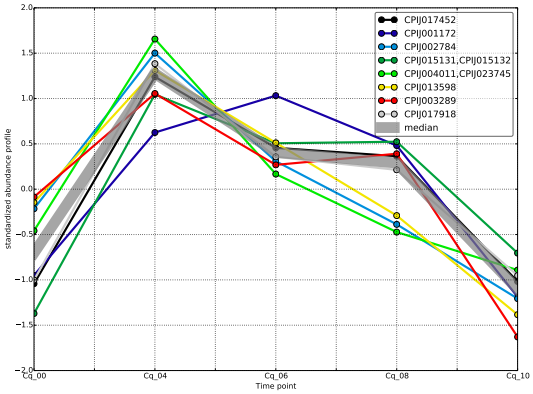
Figure 4.5



(A)



(B)

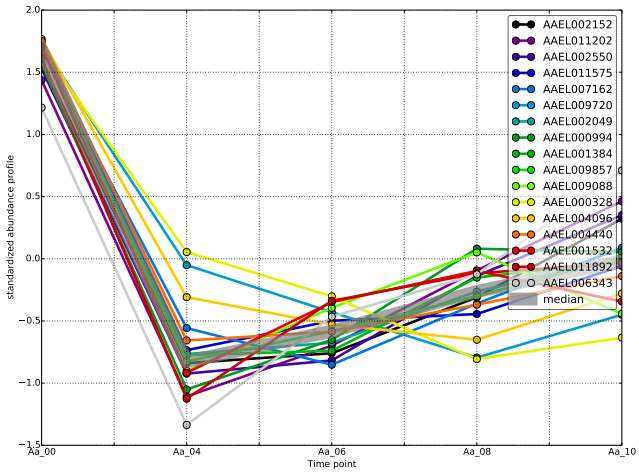


(C)

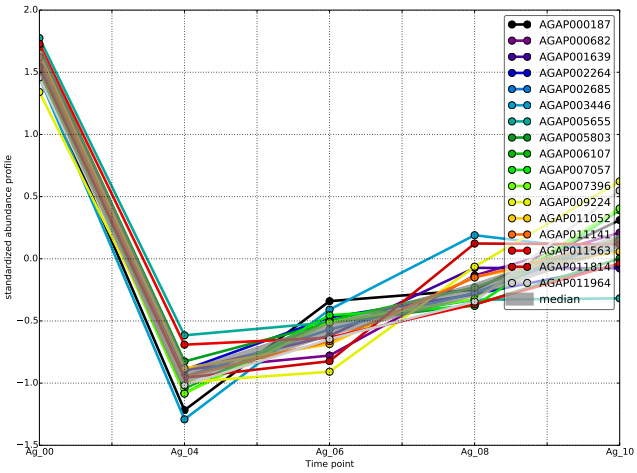
Figure 4.4: Orthologs of cluster 6:

(A) *Ae. aegypti*. **(B)** *An. gambiae*. **(C)** *Cx. quinquefasciatus*.

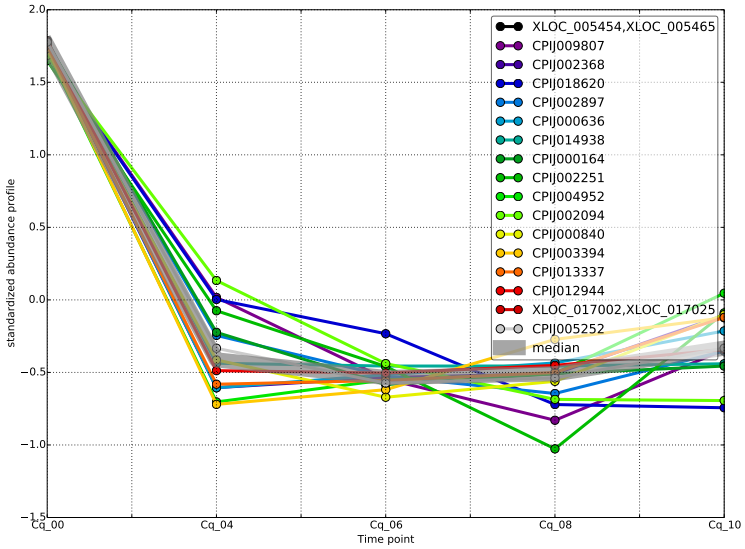
- the text
- double gene names
- fix fig widths



(A)



(B)



(C)

Figure 4.5: Orthologs of cluster 19:

(A) *Ae. aegypti*. **(B)** *An. gambiae*. **(C)** *Cx. quinquefasciatus*.

Finish this fig:

- the text
- double gene names
- fix fig widths

Chapter 5

Contributions to Literature Not Otherwise Mentioned

decide whether this gets included

5.1 Purpose

xxxxxxxxxxxxxxxxxxxxxxxxxxx

1. xxx

Chapter 6

Conclusions

replace CITEME tags in Introduction

write all the figures

MEGA: you need to write most of

6.1 Purpose

xxxxxxxxxxxxxxxxxxxxxxxxxx

1. xxx

Bibliography

[1] Z. N. Adelman, N. Jasinskiene, S. Onal, J. Juhn, A. Ashikyan, M. Salampessy, T. MacCauley, and A. A. James. nanos gene control DNA mediates developmentally regulated transposition in the yellow fever mosquito Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, 104(24):9970–5, June 2007.

[2] Z. N. Adelman, N. Jasinskiene, K. J. M. Vally, C. Peek, E. A. Travanty, K. E. Olson, S. E. Brown, J. L. Stephens, D. L. Knudson, C. J. Coates, and A. A. James. Formation and loss of large, unstable tandem arrays of the piggyBac transposable element in the yellow fever mosquito, Aedes aegypti. *Transgenic research*, 13(5):411–25, Oct. 2004.

[3] a. Ahmed, D. Martín, a. G. Manetti, S. J. Han, W. J. Lee, K. D. Mathiopoulos, H. M. Müller, F. C. Kafatos, A. Raikhel, and P. T. Brey. 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(26):14795–800, Dec. 1999.

[4] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, Jan. 2010.

[5] G. M. Attardo, S. Higgs, K. a. Klingler, D. L. Vanlandingham, and A. S. Raikhel. 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(23):13374–9, Nov. 2003.

[6] T. L. Bailey and C. Elkan. 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(1553-0833):28–36, Jan. 1994.

[7] I. Bargielowski, C. Kaufmann, L. Alphey, P. Reiter, and J. Koella. 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, N.Y.)*, 12(12):1053–8, Dec. 2012.

[8] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):pp. 289–300, 1995.

[9] K. E. Bennett, K. E. Olson, M. d. L. Muñoz, I. Fernandez-Salas, J. A. Farfan-Ale, S. Higgs, W. C. Black, and B. J. Beaty. 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(1):85–92, July 2002.

[10] S. A. Bernhardt, M. P. Simmons, K. E. Olson, B. J. Beaty, C. D. Blair, and W. C. Black. Rapid intraspecific evolution of miRNA and siRNA genes in the mosquito Aedes aegypti. *PloS one*, 7(9):e44198, Jan. 2012.

[11] W. C. Black, K. E. Bennett, N. Gorrochótegui-Escalante, C. V. Barillas-Mury, I. Fernández-Salas, M. de Lourdes Muñoz, J. A. Farfán-Alé, K. E. Olson, and B. J. Beaty. Flavivirus susceptibility in Aedes aegypti. *Archives of medical research*, 33(4):379–88, July 2002.

[12] M. Bonizzoni, W. A. Dunn, C. L. Campbell, K. E. Olson, M. T. Dimon, O. Marinotti, and A. A. James. RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, Aedes aegypti. *BMC Genomics*, 12(1):82, Jan. 2011.

[13] M. Bonizzoni, W. A. Dunn, C. L. Campbell, K. E. Olson, O. Marinotti, and A. A. James. Complex modulation of the Aedes aegypti transcriptome in response to dengue virus infection. *PloS one*, 7(11):e50512, Jan. 2012.

[14] M. Bonizzoni, W. A. Dunn, C. L. Campbell, K. E. Olson, O. Marinotti, A. A. James, and R. Kulathinal. Strain Variation in the Transcriptome of the Dengue Fever Vector, Aedes aegypti. *G3: Genes|Genomes|Genetics*, 2(1):103–114, Jan. 2012.

[15] A. Borkent and D. A. Grimaldi. The Earliest Fossil Mosquito (Diptera: Culicidae), in Mid-Cretaceous Burmese Amber. *Annals of the Entomological Society of America*, 97(5):882–888, Sept. 2004.

[16] D. E. Brackney, J. Isoe, B. W C, J. Zamora, B. D. Foy, R. L. Miesfeld, and K. E. Olson. Expression profiling and comparative analyses of seven midgut serine proteases from the yellow fever mosquito, Aedes aegypti. *Journal of insect physiology*, 56(7):736–44, July 2010.

[17] R. C. Bushland, A. W. Lindquist, and E. F. Knipling. Eradication of Screw-Worms through Release of Sterilized Males. *Science*, 122(3163):287–8, Aug. 1955.

[18] E. Calvo, B. J. Mans, J. F. Andersen, and J. M. C. Ribeiro. Function and evolution of a mosquito salivary protein family. *The Journal of biological chemistry*, 281(4):1935–42, Jan. 2006.

[19] J. M. Carlson, A. Chakravarty, C. E. DeZiel, and R. H. Gross. SCOPE: a web server for practical de novo motif discovery. *Nucleic acids research*, 35(Web Server issue):W259–64, July 2007.

[20] K.-H. Cho, H.-M. Cheon, V. Kokoza, and A. S. Raikhel. 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(4):273–81, Apr. 2006.

[21] P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. L. de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*, 25(11):1422–3, June 2009.

[22] M. K. Das and H.-K. Dai. A survey of DNA motif finding algorithms. *BMC bioinformatics*, 8 Suppl 7:S21, Jan. 2007.

[23] E. H. Davidson. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution (Google eBook)*. Academic Press, 2010.

[24] V. Deubel, R. M. Kinney, and D. W. Trent. Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. *Virology*, 155(2):365–77, Dec. 1986.

[25] M. T. Dimon, K. Sorber, and J. L. DeRisi. HMMSplicer: A Tool for Efficient and Sensitive Discovery of Known and Novel Splice Junctions in RNA-Seq Data. *PLoS ONE*, 5(11):16, 2010.

[26] N. T. Dittmer, G. Sun, S.-f. Wang, and A. S. Raikhel. CREB isoform represses yolk protein gene expression in the mosquito fat body. *Molecular and cellular endocrinology*, 210(1-2):39–49, Nov. 2003.

[27] O. Elemento and S. Tavazoie. Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome biology*, 6(2):R18, Jan. 2005.

[28] L. Facchinelli, L. Valerio, J. M. Ramsey, F. Gould, R. K. Walsh, G. Bond, M. A. Robert, A. L. Lloyd, A. A. James, L. Alphey, and T. W. Scott. Field cage studies and progressive evaluation of genetically-engineered mosquitoes. *PLoS neglected tropical diseases*, 7(1):e2001, Jan. 2013.

[29] H. M. Fischer, C. W. Wheat, D. G. Heckel, and H. Vogel. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular biology and evolution*, 25(5):809–20, May 2008.

[30] A. W. E. Franz, I. Sanchez-Vargas, Z. N. Adelman, C. D. Blair, B. J. Beaty, A. A. James, and K. E. Olson. 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(11):4198–203, Mar. 2006.

[31] F. Giannoni, H. M. Müller, J. Vizioli, F. Catteruccia, F. C. Kafatos, and A. Crisanti. Nuclear factors bind to a conserved DNA element that modulates transcription of Anopheles gambiae trypsin genes. *The Journal of biological chemistry*, 276(1):700–7, Jan. 2001.

[32] X. Guo, Y. Xu, G. Bian, A. D. Pike, Y. Xie, and Z. Xi. Response of the mosquito protein interaction network to dengue infection. *BMC genomics*, 11:380, Jan. 2010.

[33] J. Hernandez-Romano, F. J. Carlos-Rivera, H. Salgado, H. Lamadrid-Figueroa, V. Valverde-Garduño, M. H. Rodriguez, and J. Martinez-Barnetche. 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, Jan. 2008.

[34] A. M. Hess, A. N. Prasad, A. Ptitsyn, G. D. Ebel, K. E. Olson, C. Barbacioru, C. Monighetti, and C. L. Campbell. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. *BMC microbiology*, 11:45, Jan. 2011.

[35] J. Hu, B. Li, and D. Kihara. Limitations and potentials of current motif discovery algorithms. *Nucleic acids research*, 33(15):4899–913, Jan. 2005.

[36] T. Hubbard, D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, T. Down, R. Durbin, E. Eyras, J. Gilbert, M. Hammond, L. Huminiecki, A. Kasprzyk, H. Lehvaslaiho, P. Lijnzaad, C. Melsopp, E. Mongin, R. Pettett, M. Pocock, S. Potter, A. Rust, E. Schmidt, S. Searle, G. Slater, J. Smith, W. Spooner, A. Stabenau, J. Stalker, E. Stupka, A. Ureta-Vidal, I. Vastrik, and M. Clamp. The Ensembl genome database project. *Nucleic acids research*, 30(1):38–41, Jan. 2002.

[37] S. Ingsriswang, S. Yokwai, and D. Wichadakul. LinkinPath: from sequence to interconnected pathway. *Bioinformatics (Oxford, England)*, 27(14):2015–7, July 2011.

[38] A. T. Isaacs, N. Jasinskiene, M. Tretiakov, I. Thiery, A. Zettor, C. Bourgouin, and A. A. James. 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(28):E1922–30, July 2012.

[39] N. Jasinskiene, J. Coleman, A. Ashikyan, M. Salampessy, O. Marinotti, and A. A. James. Genetic control of malaria parasite transmission: threshold levels for infection in an avian model system. *The American journal of tropical medicine and hygiene*, 76(6):1072–8, June 2007.

[40] J. Juhn and A. A. James. oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti. *Insect molecular biology*, 15(3):363–72, June 2006.

[41] P. J. Kersey, D. M. Staines, D. Lawson, E. Kulesha, P. Derwent, J. C. Humphrey, D. S. T. Hughes, S. Keenan, A. Kerhornou, G. Koscielny, N. Langridge, M. D. McDowall, K. Megy, U. Maheswari, M. Nuhn, M. Paulini, H. Pedro, I. Toneva, D. Wilson, A. Yates, and E. Birney. Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic acids research*, 40(Database issue):D91–7, Jan. 2012.

[42] V. a. Kokoza, D. Martin, M. J. Mienaltowski, A. Ahmed, C. M. Morton, and a. S. Raikhel. Transcriptional regulation of the mosquito vitellogenin gene via a blood meal-triggered cascade. *Gene*, 274(1-2):47–65, Aug. 2001.

[43] E. V. Kriventseva, N. Rahman, O. Espinosa, and E. M. Zdobnov. OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic acids research*, 36(Database issue):D271–5, Jan. 2008.

[44] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25, Jan. 2009.

[45] D. Lawson, P. Arensburger, P. Atkinson, N. J. Besansky, R. V. Bruggner, R. Butler, K. S. Campbell, G. K. Christophides, S. Christley, E. Dialynas, M. Hammond, C. a. Hill, N. Konopinski, N. F. Lobo, R. M. MacCallum, G. Madey, K. Megy, J. Meyer, S. Redmond, D. W. Severson, E. O. Stinson, P. Topalis, E. Birney, W. M. Gelbart, F. C. Kafatos, C. Louis, and F. H. Collins. VectorBase: a data resource for invertebrate vector genomics. *Nucleic acids research*, 37(Database issue):D583–7, Jan. 2009.

[46] W. W. MACDONALD. 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, Dec. 1963.

[47] S. Mahony, P. E. Auron, and P. V. Benos. DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. *PLoS computational biology*, 3(3):e61, Mar. 2007.

[48] S. Mahony and P. V. Benos. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic acids research*, 35(Web Server issue):W253–8, July 2007.

[49] O. Marinotti, E. Calvo, Q. K. Nguyen, S. Dissanayake, J. M. C. Ribeiro, and A. A. James. Genome-wide analysis of gene expression in adult Anopheles gambiae. *Insect molecular biology*, 15(1):1–12, Feb. 2006.

[50] O. Marinotti, Q. K. Nguyen, E. Calvo, A. A. James, and J. M. C. Ribeiro. Microarray analysis of genes showing variable expression following a blood meal in Anopheles gambiae. *Insect molecular biology*, 14(4):365–73, Aug. 2005.

[51] F. S. Markland and M. Ann. *Toxins and Hemostasis*. Springer Netherlands, Dordrecht, 2011.

[52] M. T. Marrelli, C. K. Moreira, D. Kelly, L. Alphey, and M. Jacobs-Lorena. Mosquito transgenesis: what is the fitness cost? *Trends in parasitology*, 22(5):197–202, May 2006.

[53] G. Mathur, I. Sanchez-Vargas, D. Alvarez, K. E. Olson, O. Marinotti, and a. a. James. Transgene-mediated suppression of dengue viruses in the salivary glands of the yellow fever mosquito, Aedes aegypti. *Insect molecular biology*, 1, Aug. 2010.

[54] J. M. Meredith, R. J. L. Munks, W. Grail, H. Hurd, P. Eggleston, and M. J. Lehane. 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(4):393–401, Aug. 2006.

[55] L. A. Moreira, M. J. Edwards, F. Adhami, N. Jasinskiene, A. A. James, and M. Jacobs-Lorena. Robust gut-specific gene expression in transgenic Aedes aegypti mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20):10895–8, Sept. 2000.

[56] V. Nene, J. R. Wortman, D. Lawson, B. Haas, C. Kodira, Z. J. Tu, B. Loftus, Z. Xi, K. Megy, M. Grabherr, Q. Ren, E. M. Zdobnov, N. F. Lobo, K. S. Campbell, S. E. Brown, M. F. Bonaldo, J. Zhu, S. P. Sinkins, D. G. Hogenkamp, P. Amedeo, P. Arensburger, P. W. Atkinson, S. Bidwell, J. Biedler, E. Birney, R. V. Bruggner, J. Costas, M. R. Coy, J. Crabtree, M. Crawford, B. Debruyn, D. Decaprio, K. Eiglmeier, E. Eisenstadt, H. El-Dorry, W. M. Gelbart, S. L. Gomes, M. Hammond, L. I. Hannick, J. R. Hogan, M. H. Holmes, D. Jaffe, J. S. Johnston, R. C. Kennedy, H. Koo, S. Kravitz, E. V. Kriventseva, D. Kulp, K. Labutti, E. Lee, S. Li, D. D. Lovin, C. Mao, E. Mauceli, C. F. M. Menck, J. R. Miller, P. Montgomery, A. Mori, A. L. Nascimento, H. F. Naveira, C. Nusbaum, S. O'leary, J. Orvis, M. Pertea, H. Quesneville, K. R. Reidenbach, Y.-H. Rogers, C. W. Roth, J. R. Schneider, M. Schatz, M. Shumway, M. Stanke, E. O. Stinson, J. M. C. Tubio, J. P. Vanzee, S. Verjovski-Almeida, D. Werner, O. White, S. Wyder, Q. Zeng, Q. Zhao, Y. Zhao, C. a. Hill, A. S. Raikhel, M. B. Soares, D. L. Knudson, N. H. Lee, J. Galagan, S. L. Salzberg, I. T. Paulsen, G. Dimopoulos, F. H. Collins, B. Birren, C. M. Fraser-Liggett, and D. W. Severson. Genome sequence of Aedes aegypti, a major arbovirus vector. *Science (New York, N.Y.)*, 316(5832):1718–23, June 2007.

[57] D. D. Nimmo, L. Alphey, J. M. Meredith, and P. Eggleston. High efficiency site-specific genetic engineering of the mosquito genome. *Insect molecular biology*, 15(2):129–36, Apr. 2006.

[58] A. Pane, K. Wehr, and T. Schüpbach. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Developmental cell*, 12(6):851–62, June 2007.

[59] D. Q.-D. Pham, P. L. Douglass, C. A. Chavez, and J. J. Shaffer. Regulation of the ferritin heavy-chain homologue gene in the yellow fever mosquito, Aedes aegypti. *Insect molecular biology*, 14(3):223–36, June 2005.

[60] R. J. Pitts, D. C. Rinker, P. L. Jones, A. Rokas, and L. J. Zwiebel. Transcriptome Profiling of Chemosensory Appendages in the Malaria Vector Anopheles gambiae Reveals Tissue- and Sex-Specific Signatures of Odor Coding. *BMC genomics*, 12(1):271, May 2011.

[61] A. A. Ptitsyn, G. Reyes-Solis, K. Saavedra-Rodriguez, J. Betz, E. L. Suchman, J. O. Carlson, and W. C. Black. Rhythms and synchronization patterns in gene expression in the Aedes aegypti mosquito. *BMC genomics*, 12:153, Jan. 2011.

[62] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6):841–2, Mar. 2010.

[63] K. S. Rai and W. C. Black. Mosquito genomes: structure, organization, and evolution. *Advances in genetics*, 41:1–33, Jan. 1999.

[64] J. M. C. Ribeiro. AegyXcel.

[65] J. E. Richardson. Fjoin: Simple and Efficient Computation of Feature Overlaps. *Journal of computational biology : a journal of computational molecular cell biology*, 13(8):1457–64, Oct. 2006.

[66] A. I. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush. TM4: a free, open-source system for microarray data management and analysis. *BioTechniques*, 34(2):374–8, Feb. 2003.

[67] M. I. Salazar, J. H. Richardson, I. Sánchez-Vargas, K. E. Olson, and B. J. Beaty. Dengue virus type 2: replication and tropisms in orally infected Aedes aegypti mosquitoes. *BMC microbiology*, 7:9, Jan. 2007.

[68] T. D. Schmittgen and K. J. Livak. Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols*, 3(6):1101–8, Jan. 2008.

[69] N. Sethuraman, M. J. Fraser, P. Eggleston, and D. A. O'Brochta. Post-integration stability of piggyBac in Aedes aegypti. *Insect biochemistry and molecular biology*, 37(9):941–51, Sept. 2007.

[70] D. H. Sieglaff, W. A. Dunn, X. S. Xie, K. Megy, O. Marinotti, and A. A. James. 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(9):3053–8, Mar. 2009.

[71] A. Stark, M. F. Lin, P. Kheradpour, J. S. Pedersen, L. Parts, J. W. Carlson, M. a. Crosby, M. D. Rasmussen, S. Roy, A. N. Deoras, J. G. Ruby, J. Brennecke, E. Hodges, A. S. Hinrichs, A. Caspi, B. Paten, S.-W. Park, M. V. Han, M. L. Maeder, B. J. Polansky, B. E. Robson, S. Aerts, J. van Helden, B. Hassan, D. G. Gilbert, D. a. Eastman, M. Rice, M. Weir, M. W. Hahn, Y. Park, C. N. Dewey, L. Pachter, W. J. Kent, D. Haussler, E. C. Lai, D. P. Bartel, G. J. Hannon, T. C. Kaufman, M. B. Eisen, A. G. Clark, D. Smith, S. E. Celniker, W. M. Gelbart, and M. Kellis. Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature*, 450(7167):219–32, Nov. 2007.

[72] R. Stobbart. The control of the diuresis following a blood meal in females of the yellow fever mosquito Aedes aegypti (L). *J. Exp. Biol.*, 69(1):53–85, Aug. 1977.

[73] O. Terenius, O. Marinotti, D. Sieglaff, and A. A. James. Molecular genetic manipulation of vector mosquitoes. *Cell host & microbe*, 4(5):417–23, Nov. 2008.

[74] H. M. Thorpe and M. C. Smith. 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(10):5505–10, May 1998.

[75] M. Tompa, N. Li, T. L. Bailey, G. M. Church, B. De Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. Fu, W. J. Kent, V. J. Makeev, A. A. Mironov, W. S. Noble, G. Pavesi, G. Pesole, M. Régnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, Z. Zhu, B. D. Moor, and J. V. Helden. Assessing computational tools for the discovery of transcription factor binding sites. *Nature biotechnology*, 23(1):137–44, Jan. 2005.

[76] C. Trapnell, L. Pachter, and S. L. Salzberg. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, 25(9):1105–11, May 2009.

[77] C. Trapnell, B. a. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5):511–5, May 2010.

[78] A. J. Vilella, J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, and E. Birney. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome research*, 19(2):327–35, Feb. 2009.

[79] L. Wang, Z. Feng, X. Wang, X. Wang, and X. Zhang. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics (Oxford, England)*, 26(1):136–8, Jan. 2010.

[80] W. W. Wasserman and A. Sandelin. Applied bioinformatics for the identification of regulatory elements. *Nature reviews. Genetics*, 5(4):276–87, Apr. 2004.

[81] R. M. Waterhouse, E. V. Kriventseva, S. Meister, Z. Xi, K. S. Alvarez, L. C. Bartholomay, C. Barillas-Mury, G. Bian, S. Blandin, B. M. Christensen, Y. Dong, H. Jiang, M. R. Kanost, A. C. Koutsos, E. a. Levashina, J. Li, P. Ligoxygakis, R. M. Maccallum, G. F. Mayhew, A. Mendes, K. Michel, M. a. Osta, S. Paskewitz, S. W. Shin, D. Vlachou, L. Wang, W. Wei, L. Zheng, Z. Zou, D. W. Severson, A. S. Raikhel, F. C. Kafatos, G. Dimopoulos, E. M. Zdobnov, and G. K. Christophides. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science (New York, N.Y.)*, 316(5832):1738–43, June 2007.

[82] M. R. Wise de Valdez, D. Nimmo, J. Betz, H.-F. Gong, A. a. James, L. Alphey, and W. C. Black. Genetic elimination of dengue vector mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(12), Mar. 2011.

[83] J. Wu, D. H. Sieglaff, J. Gervin, and X. S. Xie. Discovering regulatory motifs in the Plasmodium genome using comparative genomics. *Bioinformatics (Oxford, England)*, 24(17):1843–9, Sept. 2008.

[84] X. Xie, J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander, and M. Kellis. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 434(7031):338–45, Mar. 2005.

[85] P. Xu, R. Atkinson, D. N. M. Jones, and D. P. Smith. Drosophila OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron*, 45(2):193–200, Jan. 2005.

Glossary

anthropophilic the characteristic of preferring to feed on or live amidst humans. 22

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. 21

hemolymph A fluid analogous to blood that constitutes the circulatory system of most invertebrates. 4, 22

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

midgut the location in the mosquito's digestive system where the bloodmeal is processed. 22

monophyletic having developed from a single ancestral source. 21

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, 8

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, 8

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

CRE *cis*-regulatory element. 11, 17, 20, 21, 25, 27

CRM *cis*-regulatory module. 11, 21, 25, 26

FPKM fragments per kilobase of transcript per million mapped reads. 72

MDOS Motif Discovery using Orthologous Sequences. 20

Mya million years ago. 20

NBF non-bloodfed. 42

PBM post bloodmeal. 23, 42

RIDL release of insects carrying a dominant lethal phenotype. 9

RNAi RNA interference. 25

SIT sterile insect technique. 5, 6, 10

TFBS transcription factor binding site. 20

TSS transcriptional start sites. 27

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 [78]. 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 co-

ordinates 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 [56] 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 [65] 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 de-

termination 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 [48, 47], 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 [50, 49] 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 [8], followed by k-means clustering with Euclidean distance separation using open-source software (MeV MultiExperiment Viewer v4.1.01, TM4 [66]). 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 $\leq$ 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 differ-

ent 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 [46], 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 [56]. 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 af-

ter 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 [44] was used to align the Illumina reads against the *Ae. aegypti* genome (version AaegL1) [45], 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 [62]. Differential expression between conditions was assessed by the likelihood ratio test as

implemented in the program DEGseq [79], 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% [8]. Transcript description was based on the *Ae. aegypti* protein database AegyXcel [64].

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 [68], 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 [25] 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 [19] 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 [61].

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 af-

ter 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 [72]. 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. Amplification 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 [12].

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 [12]. Sequence reads were mapped to the LVP reference genome with Bowtie [44], allowing a maximum

of two mismatches and with the -m option, which returns only reads with a single best match in the genome [12].

Differential transcript accumulation levels among conditions were assessed by the likelihood ratio test implemented in DEGseq [79], 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% [8]. 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 [32] with the use of Cytoscape_v2.7.0 (http://www.cytoscape.org/). Metabolic pathways were analyzed by LinkinPath [37], 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 [10]. 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 [24]. DENV-2 Jam1409 was cultured in *Ae. albopictus* C6/36 cells and high passage (>25 passages) viruses were used in oral challenges [67]. 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 [34]. 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 [67]. 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 [60]. 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 [14].

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 [76]. 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 [4], and in terms of fragments per kilobase of transcript per million mapped reads (FPKM) using Cufflinks [77]. 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 [4, 77]. 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 [14, 4]. 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 [14]. Gene description used the biomart function of the EnsemblMetazoa (http://metazoa.ensembl.org/index.html) and Immunodb (http://cegg.unige.ch/Insecta/immunodb) [41, 81].

The motif-based sequence analyses suite MEME [6] 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 [48].

Appendix B

Code