Transcriptional regulation of mosquito oogenesis

by

Patrick David Jennings

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Entomology

Program of Study Committee:
Lyric C. Bartholomay, Major Professor
Bryony C. Bonning
Bradley J. Blitvich

Iowa State University

Ames, Iowa

2011

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 4

ABSTRACT

The Eastern treehole mosquito, *Aedes triseriatus* (Say), is the primary vector of La Crosse virus (LACV) in the United States. LACV is maintained in natural populations in part by transovarial transmission from infected females to their progeny. Our understanding of the innate immune response in mosquito ovaries is a fundamental step toward understanding the transmission cycle of LACV and other vertically transmitted arthropod-borne viruses. We investigated signaling mechanisms between the nutrient content of the vertebrate blood meal and activation of mosquito egg development. In particular, the target of rapamycin (TOR) serine/threonine protein kinase is a central regulator of nutritional signaling, and we examined the roles of TOR and its downstream target, S6K, in *Ae. triseriatus* by RNA interference. We also examined the transcriptomes of *Ae. triseriatus* and *Ae. aegypti* (L.) ovaries by Illumina RNA-seq. We assembled 17,743 novel transcript sequences from *Ae. triseriatus* ovary mRNA using a novel and highly accurate assembly method, and we identified genes that are differentially regulated during the processes of follicular atresia, apoptosis, and autophagy. Distinct differences exist between the transcriptional regulation of *Ae. triseriatus* and *Ae. aegypti* oogenesis. TOR and many other proteins with previously unknown function are transcriptionally regulated during apoptosis and autophagy in mosquito ovaries during oogenesis.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

More than two million people die each year from mosquito-borne disease. Numerous arthropod-borne viruses (arboviruses) are vectored by mosquitoes (Leake 1984; Zhang, Zhang, and Zhu 2010), and many of those arboviruses are maintained in nature through transovarial transmission (TOT). In TOT, a virus infects the ovary of a female mosquito, and the viral infection is passed on to the offspring. The virus persists throughout the entire life cycle of the offspring, ultimately infecting the salivary glands of vertically-infected individuals, facilitating transmission of the virus to new vertebrate hosts. TOT is demonstrated in nature by detection of the virus in eggs, larvae, nulliparous adult females, or adult males.

Transovarial transmission has been studied extensively in the transmission cycles of the mosquito-borne yellow fever virus (Aitken et al. 1979; Cordellier 1991; Beaty, Tesh, and Aitken 1980), dengue virus (Arunachalam et al. 2008; Joshi, Singhi, and Chaudhary 1996), and La Crosse virus (Watts et al. 1974; Borucki et al. 2001; Watts et al. 1973). Evidence also exists for transovarial transmission of many other mosquito-borne viruses (Table 1). Some viruses can be transmitted transovarially in one mosquito species but not in a different species, which complicates the study of arbovirus transmission. For example, Hardy *et al.* (1980) showed experimentally that *Aedes albopictus* (Skuse) and *Ae. epactius* Dyar & Knab transmitted St. Louis encephalitis virus to their F1 progeny, but they could not detect TOT of the same virus in seven other mosquito species that were inoculated with St. Louis encephalitis virus. One possible explanation for this phenomenon is that some species have a

more effective antiviral immune response to a particular virus, such that an individual can effectively clear the viral infection from ovary cells. Another hypothesis is that the membranes of ovarian cells may lack the requisite receptors that would allow the virus to enter cells and replicate.

To date, no research has been done on antiviral immunity in the mosquito ovary. Apoptosis and autophagy are two cellular processes that participate in antiviral immunity in both vertebrates and invertebrates (Cashio, Lee, and Bergmann 2005; Siegel 2006; Pritchett, Tanner, and McCall 2009; Arsham and Neufeld 2009; Levine and Deretic 2007). In the genetic model fly *Drosophila melanogaster* Meigen, genes responsible for many aspects of cell death have been shown to be regulated during oogenesis at the level of RNA transcription. Because mosquitoes are also flies, it is likely that transcriptional regulation is behaving similarly during cell death processes in mosquito egg development. The molecular basis of cell death has never before been studied in the mosquito ovary. We posit that deeper understanding of these processes will be informative in terms of the physiology of oogenesis and, in turn, the dynamics of vertical transmission.

The research described here represents a novel analysis of the transcriptional regulation of oogenesis, with an emphasis on programmed cell death pathways. These processes were studied using RNA interference (RNAi) and "next-gen" massively parallel cDNA sequencing in two mosquito species: *Aedes aegypti* (L.), which is the primary vector of dengue virus, yellow fever virus, and chikungunya virus, and *Aedes triseriatus* (Say), the primary vector of La Crosse virus.

Thesis Organization

This thesis is organized into five chapters, each of which addresses a different aspect of the transcriptional regulation of mosquito egg development. The following section of this first chapter is a comprehensive literature review of the current state of research on mosquito oogenesis and molecular signaling pathways related to cell death in mosquitoes.

Chapter two describes my research on two important proteins related to nutritional signaling and cell death in the mosquito ovary, the target of rapamycin and p70 S6 kinase. Expression of these two proteins was decreased in the mosquito *Ae. triseriatus* by RNA interference *in vivo*, and the effects on oogenesis were measured.

Chapters three and four describe my efforts to profile transcriptional regulation of cell death in the mosquito ovary. In order to accurately measure transcript abundance in *Ae. triseriatus*, I first assembled a complete *de novo* transcriptome from Illumina RNA-seq short reads. Chapter three describes the bioinformatics approach that I used to assemble a complete *de novo* transcriptome for the *Ae. triseriatus* ovary using a custom-written Perl program, called *Aedenovo* (a portmanteau of the mosquito genus, *Aedes*, and the type of assembly performed, *de novo*). Chapter four is a detailed analysis of differential gene transcription during three distinct cell death events that occur during mosquito oogenesis. Comparisons are also made between *Ae. aegypti* and *Ae. triseriatus*, and genes are identified that play a role in the induction of follicular atresia by protein deprivation.

Finally, Chapter five contains overall conclusions of my research on transcriptional regulation of mosquito oogenesis. The current state of scientific knowledge related to mosquito egg development is summarized, and the importance of my research is evaluated with respect to arbovirus transovarial transmission and public health.

Literature Review

Blood is an essential source of protein for egg production in adult females of most species of mosquitoes. The recurring need for blood makes mosquitoes efficient vectors of vertebrate disease agents, including protozoan parasites in the genus *Plasmodium*, a wide variety of RNA viruses, and nematode worms in the family *Onchocercidae*. These disease agents cause hundreds of millions of infections each year, which result in more than two million estimated deaths annually. In addition to the public heath burden, economic losses caused by mosquito-borne diseases are also staggering. Because of the medical and economic significance of mosquito-borne disease, mosquito reproductive biology and virus transmission dynamics are incredibly important areas of research. Many arboviruses are transmitted transovarially (Table 1) from an infected female mosquito to her offspring, and in order to persist throughout the entire mosquito life cycle, these viruses must evade or suppress the antiviral innate immune response of mosquito cells, which includes the cell death processes of autophagy and apoptosis. Current research is reviewed here relating to mosquito oogenesis and molecular signaling pathways related to cell death in mosquitoes.

Oogenesis

Hematophagous mosquitoes are broadly categorized into two groups with respect to egg development. Anautogenous mosquitoes require a blood meal before they can lay their first batch of eggs, whereas autogenous mosquitoes can use teneral protein reserves—stored in fat body tissue during larval development—to synthesize yolk proteins and develop one batch of eggs before taking a blood meal. Although the biochemical processes that regulate oogenesis are very similar in autogenous and anautogenous mosquitoes, the differences that

do exist in reproductive biology have profound ecological, and perhaps, epidemiological effects. One important distinction is that anautogenous mosquitoes must feed on vertebrate hosts more frequently than autogenous mosquitoes, thereby increasing their overall vectorial capacity (the potential to transmit viruses to a new vertebrate host). The protein that is acquired from a vertebrate blood meal is used for synthesis of vitellogenin and other yolk protein precursors (YPPs) in the fat body, which are transported to the oocyte and converted to vitellin yolk proteins.

There are at least three independent pathways that converge to regulate production of YPP in the mosquito fat body. First, juvenile hormone prepares the fat body for YPP production and prepares the ovaries for YPP uptake. The second pathway begins with a hormonal signal that releases stored ovarian ecdysteroidogenic hormone (OEH I) from the corpus cardiacum (CC). OEH I stimulates the ovaries to produce ecdysone, which is converted to 20-hydroxyecdysone (20E). This in turn stimulates the fat body to produce YPPs. A third pathway involves the transduction of nutritional signals from the midgut to the fat body and inhibits YPP production when nutrient reserves are low. Successful activation of all three pathways is required for activation of YPP genes.

Juvenile hormone (JH) III is the primary hormone that regulates the previtellogenic phase of reproduction. If larvae have poor nutrition or develop in crowded conditions, JH synthesis in newly emerged adults is too low to initiate the previtellogenic phase of the reproductive cycle (Caroci, Li, and Noriega 2004; Shiao et al. 2008). However, JH synthesis and initiation of the reproductive cycle occurs after the adult female feeds. In newly emerged females with sufficient teneral nutrient reserves, JH is produced in the corpora allata and is released within the first 72 hours following eclosion. In a series of experiments reviewed by

Klowden (1997), allatectomized *Ae*. *aegypti* adults failed to develop primary follicles to the resting stage unless the allatectomy was delayed until 3 days post eclosion. However, Hernández-Martínez *et al*. (2007) found that in the same species, decapitation of mosquitoes arrested primary follicle development only when the decapitation was performed within the first 12 hours following eclosion. This suggests that synthesis or secretion of JH by the adult corpora allata may be under the control of a factor produced in the brain.

Oogenesis is also affected by signaling cascades that are regulated by JH. The broad gene (*br*) is involved in the 20E regulatory pathway, and the expression of *br* is regulated by JH during the previtellogenic and vitellogenic phases of reproduction. Surprisingly, JH has opposite effects on *br* expression during those two phases. During the previtellogenic phase, JH stimulates *br* expression, but in older, vitellogenic mosquitoes (when 20E levels are higher), JH inhibits *br*. The broad proteins Z1, Z2, and Z4 are found in circulation during the vitellogenic phase, and these proteins are essential for proper activation and termination of the vitellogenin gene (*Vg*) in response to 20E (Chen et al. 2004).

OEH I is a hormone that has been implicated as a key initiation factor of oogenesis. It is synthesized in the medial neurosecretory cells (MNC) of the brain, the ventral nerve cord, and the midgut endocrine cells (Brown and Cao 2001), and OEH I is stored in the corpus cardiacum (Klowden 1987). When released from the CC, OEH I initiates the vitellogenic stage of ovarian growth by stimulating the ovaries to produce ecdysone (Klowden 1997; Brown et al. 1998), which is then converted to 20-hydroxyecdysone (20E) by the enzyme 20-monooxygenase. OEH I is essential to the initiation of vitellogenesis in the fat body, and the mechanism of OEH I release appears to be a major factor in determining whether a mosquito species can develop eggs autogenously (Klowden 1997).

Anautogenous mosquitoes have been studied more extensively, due to their increased vectorial capacity and public health significance. In anautogenous mosquitoes, a blood meal causes the midgut epithelium to stretch and the concentration of amino acids in the hemolymph to increase. The combination of midgut stretching and the increase in amino acids in the hemolymph triggers the CC to release its stored OEH I.

Autogenous mosquitoes have conserved the hormonal mechanisms that initiate oogenesis, but the triggers for the release of those hormones are different. In the autogenous species *Wyeomyia smithii* Coquillett, 1901, the release of OEH I from the CC occurs before adult ecdysis (Smith and Mitchell 1986). In many autogenous mosquitoes, the OEH releasing signal comes from the ovaries or male accessory gland (MAG) fluid, providing a stimulus for oogenesis only when sperm are present to fertilize the eggs. Pondeville *et al*. (2008) found that high levels of 20E were produced in the MAG. This 20E was delivered to the female during copulation, and it acted as an allohormone to influence the female reproductive cycle and increase fertility.

JH and OEH I are essential for mosquito oogenesis to occur. The expression of these two hormones is dependent on factors that increase the mosquito's overall fitness, such as larval nutrition (Telang et al. 2006; Telang, Frame, and Brown 2007) and presence of MAG hormones (Pondeville et al. 2008). However, if oogenesis begins while the mosquito's nutritional reserves are low, eggs fail to develop correctly, and there is clearly an amino acid signaling pathway involved.

Nutritional signaling adds an important third checkpoint to the initiation of vitellogenesis. A model for a nutrient signaling pathway related to oogenesis is presented in Figure 1. After an anautogenous mosquito has taken an adequate blood meal, the resulting

increase in amino acids in the hemolymph removes a block to 20E-stimulated *Vg* transcription via the target of rapamycin (TOR) signaling cascade. TOR is involved in two protein complexes, TOR complex 1 and 2 (TORC1 and TORC2). TORC1 includes regulatory associated protein of TOR (raptor) and LST8 (Hara et al. 2002; Chen and Kaiser 2003), and this protein complex is more directly involved with vitellogenesis and oogenesis. During the previtellogenic arrest stage, the tuberous sclerosis protein complex (TSC2 and TSC1) inactivates the GTPase Rheb, causing inhibition of TOR kinase activity. This inhibition of TOR effectively arrests follicular development by inhibiting transcription of *Vg* and other YPP genes. Hansen *et al*. (2004) found that an increase in amino acid titers in *Ae*. *aegypti* hemolymph activates TOR, which in turn phosphorylates S6 kinase (Hansen et al. 2005) and allows 20E to stimulate *Vg* gene expression. The mechanism of amino acid signaling is not fully understood, but the family of Rag guanosine triphosphatases are involved in the activation of TORC1 by affecting its intracellular localization to a compartment that also contains its activator Rheb (Sancak et al. 2008).

There are many hormones and other proteins that regulate mosquito oogenesis. The production and release of JH, OEH I, and nutrient signaling proteins are essential to the control of oogenesis and the success of mosquitoes as disease vectors. Our understanding of the complex interactions of the proteins regulating mosquito oogenesis is fundamental to the goal of eradicating deadly mosquito-borne diseases through disruption of the reproductive capacity of the mosquito vector.

Regulation of Programmed Cell Death

Programmed cell death (PCD) is inherent to the processes of growth and development, and three distinct PCD events occur during mosquito oogenesis that involve both apoptosis and autophagy. A mosquito follicle contains an oocyte and seven nurse cells; those eight cells are surrounded by many small follicular epithelial cells. At the end of each gonotropic cycle, the oocyte is the only cell that survives from each follicle, and the coordinated regulation of cell death in the nurse cells and follicular epithelium are prerequisite to development of the oocyte, and thereby, a viable embryo.

The genetic regulation of cell death has been studied extensively in *Drosophila*, and the same types of cell death have been observed to a lesser extent in mosquitoes. Cell death is genetically regulated by two opposing forces: signals that repress cell death, and signals that activate cell death. Inhibitors of apoptosis (IAPs) are proteins involved in repressing cell death, and IAP antagonists and caspase activators both stimulate the action of caspases to activate cell death.

Caspases are cysteine proteases that are active during apoptosis. Initiator caspases (*e.g.* CASP2, CASP8, CASP9, and CASP10) cleave and activate the effector caspases (*e.g.* CASP3, CASP6, CASP7), which are the main effectors of apoptosis. The functions of caspases have been reviewed elsewhere (Yuan and Horvitz 2004; Salvesen and Riedl 2007; Bryant et al. 2008). Briefly, initiator caspases lead to the formation of a protein complex called the apoptosome, which contains the protein Apaf-1, cytochrome c, and an initiator caspase (Schafer and Kornbluth 2006). The role of cytochrome c has been well studied in mammals, but its role in dipterans has not been determined. The apoptosome activates downstream effector caspases, which cleave key cellular proteins and lead to apoptosis.

Caspase action is blocked by the IAP family of proteins, which contains at least four members in *Drosophila*: Diap1, Diap2, Bruce, and Deterin (Yoo et al. 2002; Hay, Wassarman, and Rubin 1995; Jones et al. 2000; Qiu et al. 2004). In the mosquito *Ae. aegypti*, 1:1 orthologs of *Drosophila* Diap1, Bruce, and Deterin were found in the genome sequence, and other unique IAPs have been identified in mosquitoes (Pridgeon et al. 2008). IAPs inhibit apoptosis by poly-ubiquitination of caspases and other proapoptotic proteins, which causes deactivation or degradation of those poly-ubiquitinated proteins (Herman-Bachinsky et al. 2007; Ditzel et al. 2008). Apoptosis is strongly inhibited by IAPs, and expression of IAP antagonists is necessary for induction of apoptosis.

IAP antagonists in *Drosophila* are represented by the genes *grim*, *reaper*, and *hid*, as well as the more recently characterized *sickle* (Christich et al. 2002) and *jafrac2* (Tenev et al. 2002). These IAP antagonists interact as homodimers of Rpr or heterodimers of Rpr and other IAP antagonist proteins to trigger apoptosis (Sandu, Ryoo, and Steller 2010). In mosquitoes, IAP antagonists were only recently identified. To date, only two mosquito IAP antagonists have been characterized—Michelob_x and IMP. These proteins induce apoptosis in a manner similar to Rpr/Grim in *Drosophila* (Zhou et al. 2005; Wang and Clem 2011). IAP antagonists are difficult to identify due to the low level of sequence similarity, and additional IAP antagonists may exist in mosquitoes (Wang and Clem 2011).

The main caspase activator in *Drosophila* is Apaf-1 related killer (Ark). Ark binds to the initiator caspase Dronc, and the complex of Ark and Dronc is involved in forming the apoptosome. Ark has recently been found in mosquitoes (Bryant et al. 2008), and shares many similarities to *Drosophila* Ark in both structure and function.

PCD is a natural process in oogenesis, but PCD is also very likely involved in controlling arbovirus infection in the mosquito host. Transovarial transmission of an arbovirus requires a persistent, low-level infection throughout the mosquito life cycle. In order for transovarial transmission to be effective, the virus must strike a delicate balance to not be eliminated by the mosquito immune system at low levels and to not replicate to levels that would prove lethal to the mosquito host; both situations would terminate transmission of the virus to the next host generation. We posit that mosquitoes may use autophagy as an antiviral response to low-level, persistent infection with an arbovirus.

Autophagy is a normal cellular process by which cytoplasmic contents such as damaged organelles are engulfed in an organelle called the autophagosome, which merges with lysosomes to digest the contents of the autophagosome. Autophagy can also be used to clear viral components from a cell. In *Drosophila* infected with vesicular stomatitis virus (VSV), autophagy is upregulated and helps to clear the virus from infected cells. This upregulation of autophagy in *Drosophila* is mediated through the TOR pathway (Sabin, Hanna, and Cherry 2010; DiAngelo et al. 2009). VSV most likely activates an extracellular pattern recognition receptor (Shelly et al. 2009), and the resulting signal inhibits PI3K activation of Akt in the TOR pathway. This attenuation of TOR activity leads to a decrease in cell growth and fat storage as well as an in increase in autophagy and reallocation of resources to other immune functions (DiAngelo et al. 2009). However, mosquito cells with persistent viral infections of a co-evolved virus do not demonstrate an increase in the transcription of autophagy-related genes (Bartholomay et al. 2010). Autophagy is involved in normal PCD during oogenesis in the higher flies (Nezis et al. 2006, 2006; Velentzas et al. 2007), and we are just beginning to understood the role of autophagy in the regulation of

oogenesis in mosquitoes. The research presented in this thesis will identify genes that are transcriptionally involved in such cell death processes during oogenesis, and our research will help to unravel some of the complex molecular mechanisms that regulate autophagy and apoptosis during mosquito oogenesis.

Figures/Tables

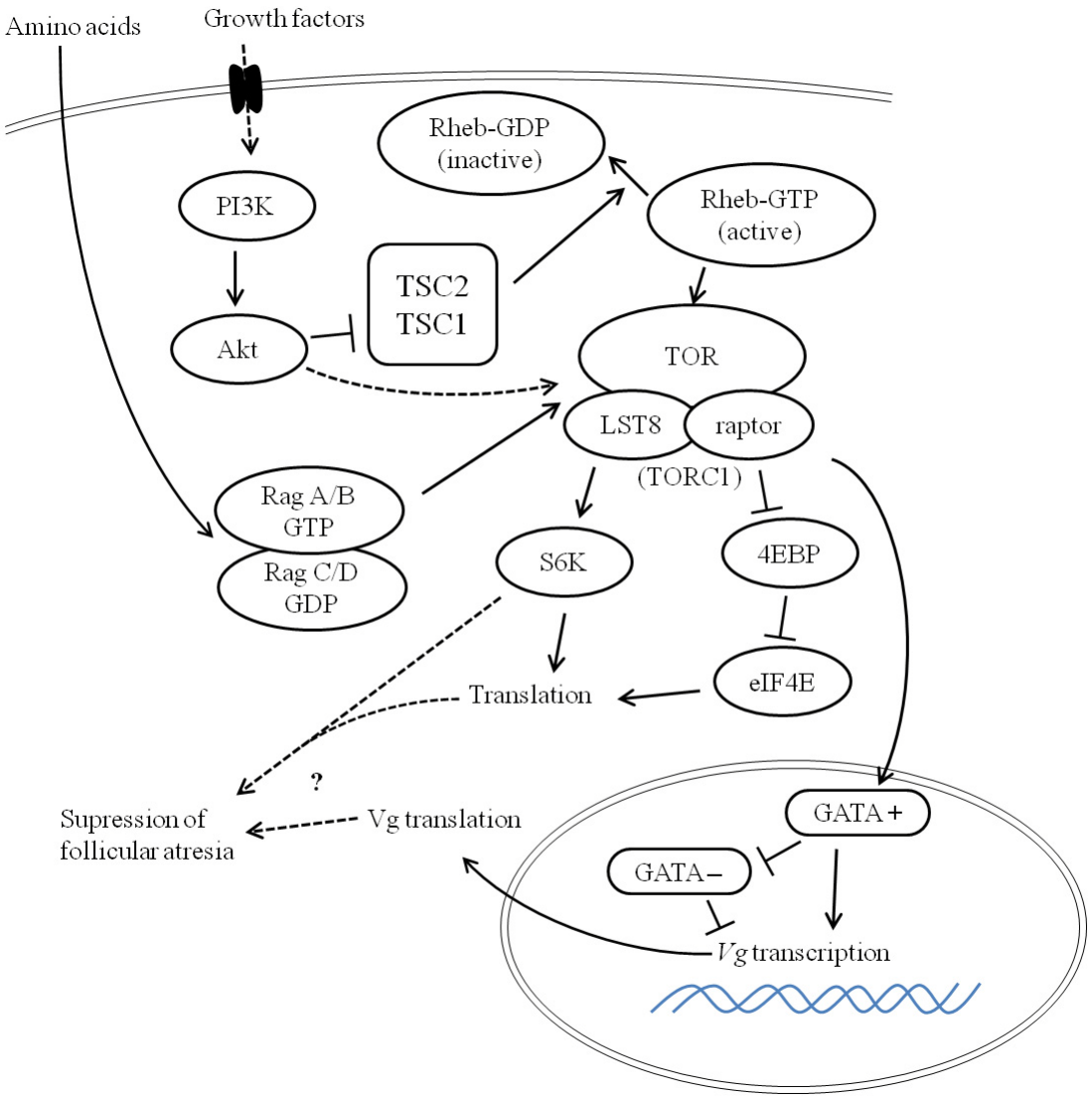


Figure 1. The target of rapamycin (TOR) pathway in the context of mosquito oogenesis. Insulin-like peptides and growth factors repress tuberous sclerosis protein complex activity (TSC1 and TSC2), which activates Rheb and TOR complex 1 (TORC1). High levels of amino acids also stimulate TORC1 via the Rag GTPase proteins. Downstream of TORC1, a repressor GATA factor is displaced by an enhancer GATA factor, resulting in expression of vitellogenin (Vg), which is transported to the ovary. Within the ovary, S6K may play a major role in supression of follicular atresia.

Table 1. Mosquito-borne viruses that show evidence of transovarial transmission (TOT).

Virus Family(:Genus) and Species	Experimental Evidence of TOT
Bunyaviridae	
La Crosse virus	(Watts et al. 1973)
	(Watts et al. 1974)
	(Pantuwatana et al. 1974)
	(Watts et al. 1975)
California encephalitis virus	(Crane and Elbel 1977)
	(Turell and Hardy 1980)
Jamestown Canyon virus	(Berry et al. 1977)
Keystone virus	(LeDuc et al. 1975)
San Angelo virus	(Tesh and Shroyer 1980)
	(Tesh 1980)
Snow shoe hare virus	(McLean et al. 1975)
Tahyna virus	(Bardos 1976)
Trivittatus virus	(Andrews et al. 1977)
	(Christensen et al. 1978)
Flaviviridae: *Flavivirus*	
Koutango virus	(Coz et al. 1976)
Yellow fever virus	(Aitken et al. 1979)
	(Tesh et al. 1979)
	(Beaty, Tesh, and Aitken 1980)
Japanese encephalitis virus	(Rosen et al. 1978)
	(Rosen, Shroyer, and Lien 1980)
	(Tesh et al. 1979)
Dengue 2 virus	(Tesh 1980)
	(Jousset 1981)
Dengue 1, *3*, and *4 viruses*	(Tesh 1980)
Murray Valley encephalitis virus	(Kay and Carley 1980)
West Nile virus	(Mishra and Mourya 2001)
Kunjin virus (now classified as a subtype of *West Nile virus*)	(Tesh 1980)
	(Tesh 1980)
St. Louis encephalitis virus	(Hardy et al. 1980)
	(Francy et al. 1981)
Togaviridae: *Alphavirus*	
Semliki Forest virus	(Stones 1960)
Eastern equine encephalitis virus	(Chamberlain and Sudia 1961)
Western equine encephalitis virus	(Thomas 1963)

References

Aitken, Thomas H. G., Robert B. Tesh, Barry J. Beaty, and Leon Rosen. 1979. Transovarial transmission of yellow fever virus by mosquitoes (*Aedes aegypti*). *Am J Trop Med Hyg* 28 (1):119-121.

Andrews, W. N., W. A. Rowley, Y. W. Wong, D. C. Dorsey, and W. J. Hausler, Jr. 1977. Isolation of trivittatus virus from larvae and adults reared from field-collected larvae of *Aedes trivittatus* (Diptera: Culicidae). *J Med Entomol* 13 (6):699-701.

Arsham, Andrew M., and Thomas P. Neufeld. 2009. A genetic screen in *Drosophila* reveals novel cytoprotective functions of the autophagy-lysosome pathway. *PLoS ONE* 4 (6):e6068.

Arunachalam, N., S. C. Tewari, V. Thenmozhi, R. Rajendran, R. Paramasivan, R. Manavalan, K. Ayanar, and B. K. Tyagi. 2008. Natural vertical transmission of dengue viruses by *Aedes aegypti* in Chennai, Tamil Nadu, India. *Indian J Med Res.* 127 (4):395-7.

Bardos, V. 1976. The ecology and medical importance of the Tahyna Virus. *MMW Munch Med Wochenschr* 118 (49):1617-20.

Bartholomay, Lyric C., Robert M. Waterhouse, George F. Mayhew, Corey L. Campbell, Kristin Michel, Zhen Zou, Jose L. Ramirez, Suchismita Das, Kanwal Alvarez, Peter Arensburger, Bart Bryant, Sinead B. Chapman, Yuemei Dong, Sara M. Erickson, S. H. P. Parakrama Karunaratne, Vladimir Kokoza, Chinnappa D. Kodira, Patricia Pignatelli, Sang Woon Shin, Dana L. Vanlandingham, Peter W. Atkinson, Bruce Birren, George K. Christophides, Rollie J. Clem, Janet Hemingway, Stephen Higgs, Karine Megy, Hilary Ranson, Evgeny M. Zdobnov, Alexander S. Raikhel, Bruce M. Christensen, George Dimopoulos, and Marc A. T. Muskavitch. 2010. Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* 330 (6000):88-90.

Beaty, Barry J., Robert B. Tesh, and Thomas H. G. Aitken. 1980. Transovarial transmission of yellow fever virus in *Stegomyia* mosquitoes. *Am J Trop Med Hyg* 29 (1):125-132.

Berry, R. L., B. J. L. Weigert, C. H. Calisher, M. A. Parsons, and G. T. Bear. 1977. Evidence for transovarial transmission ofJamestown Canyon virus in Ohio. *Mosquito News* 3:494-496.

Borucki, Monica K., Brian J. Kempf, Carol D. Blair, and Barry J. Beaty. 2001. The effect of mosquito passage on the La Crosse virus genotype. *J Gen Virol* 82 (12):2919-2926.

Brown, MR, and C Cao. 2001. Distribution of ovary ecdysteroidogenic hormone I in the nervous system and gut of mosquitoes. *J Insect Sci* 1:3.

Brown, MR, R Graf, KM Swiderek, D Fendley, TH Stracker, DE Champagne, and AO Lea. 1998. Identification of a steroidogenic neurohormone in female mosquitoes. *J Biol Chem* 273 (7):3967-71.

Bryant, Bart, Carol D. Blair, Ken E. Olson, and Rollie J. Clem. 2008. Annotation and expression profiling of apoptosis-related genes in the yellow fever mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* 38 (3):331-345.

Caroci, Abrahim S., Yiping Li, and Fernando G. Noriega. 2004. Reduced juvenile hormone synthesis in mosquitoes with low teneral reserves reduces ovarian previtellogenic development in *Aedes aegypti*. *J Exp Biol* 207 (15):2685-2690.

Cashio, Peter, Tom V. Lee, and Andreas Bergmann. 2005. Genetic control of programmed cell death in *Drosophila melanogaster*. *Seminars in Cell & Developmental Biology* 16 (2):225-235.

Chamberlain, R W, and W D Sudia. 1961. Mechanism of transmission of viruses by mosquitoes. *Annual Review of Entomology* 6 (1):371-390.

Chen, Esther J., and Chris A. Kaiser. 2003. LST8 negatively regulates amino acid biosynthesis as a component of the TOR pathway. *The Journal of Cell Biology* 161 (2):333-347.

Chen, L, J Zhu, G Sun, and AS Raikhel. 2004. The early gene Broad is involved in the ecdysteroid hierarchy governing vitellogenesis of the mosquito *Aedes aegypti*. *J Mol Endocrinol* 33 (3):743-61.

Christensen, Bruce M., Wayne A. Rowley, Yau W. Wong, Donald C. Dorsey, and William J. Hausler, Jr. 1978. Laboratory studies of transovarial transmission of trivittatus virus by *Aedes trivittatus*. *Am J Trop Med Hyg* 27 (1):184-186.

Christich, Anna, Saila Kauppila, Po Chen, Naoko Sogame, Su-Inn Ho, and John M. Abrams. 2002. The damage-responsive *Drosophila* gene *sickle* encodes a novel IAP binding protein similar to but distinct from *reaper*, *grim*, and *hid*. *Current Biology* 12 (2):137-140.

Cordellier, R. 1991. L'épidémiologie de la fièvre jaune en Afrique de l'Ouest. *Bull World Health Organ.* 69 (1):73-84.

Coz, J., M. Valade, M. Cornet, and Y. Robin. 1976. Transovarian transmission of a Flavivirus, the Koutango virus, in *Aedes aegypti* L. *C R Acad Sci Hebd Seances Acad Sci D* 283 (1):109-10.

Crane, G. T., and R. E. Elbel. 1977. Transovarial transmission of California encephalitis virus in the mosquito *Aedes dorsalis* at Blue Lake, Utah. *Mosquito News* 37:479-482.

DiAngelo, Justin R., Michelle L. Bland, Shelly Bambina, Sara Cherry, and Morris J. Birnbaum. 2009. The immune response attenuates growth and nutrient storage in *Drosophila* by reducing insulin signaling. *Proceedings of the National Academy of Sciences* 106 (49):20853-20858.

Ditzel, Mark, Meike Broemer, Tencho Tenev, Clare Bolduc, Tom V. Lee, Kristoffer T. G. Rigbolt, Richard Elliott, Marketa Zvelebil, Blagoy Blagoev, Andreas Bergmann, and Pascal Meier. 2008. Inactivation of effector caspases through nondegradative polyubiquitylation. *Molecular Cell* 32 (4):540-553.

Francy, D. B., W. A. Rush, M. Montoya, D. S. Inglish, and R. A. Bolin. 1981. Transovarial transmission of St. Louis encephalitis virus by *Culex pipiens* complex mosquitoes. *Am J Trop Med Hyg* 30 (3):699-705.

Hansen, IA, GM Attardo, JH Park, Q Peng, and AS Raikhel. 2004. Target of rapamycin-mediated amino acid signaling in mosquito anautogeny. *Proc Natl Acad Sci U S A* 101 (29):10626-31.

Hansen, Immo A., Geoffrey M. Attardo, Saurabh G. Roy, and Alexander S. Raikhel. 2005. Target of rapamycin-dependent activation of S6 kinase is a central step in the transduction of nutritional signals during egg development in a mosquito. *Journal of Biological Chemistry* 280 (21):20565-20572.

Hara, Kenta, Yoshiko Maruki, Xiaomeng Long, Ken-ichi Yoshino, Noriko Oshiro, Sujuti Hidayat, Chiharu Tokunaga, Joseph Avruch, and Kazuyoshi Yonezawa. 2002. Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell* 110 (2):177-189.

Hardy, James L., Leon Rosen, Laura D. Kramer, Sally B. Presser, Donald A. Shroyer, and Michael J. Turell. 1980. Effect of rearing temperature on transovarial transmission of St. Louis encephalitis virus in mosquitoes. *Am J Trop Med Hyg* 29 (5):963-968.

Hay, Bruce A., David A. Wassarman, and Gerald M. Rubin. 1995. Drosophila homologs of baculovirus inhibitor of apoptosis proteins function to block cell death. *Cell* 83 (7):1253-1262.

Herman-Bachinsky, Y., H. D. Ryoo, A. Ciechanover, and H. Gonen. 2007. Regulation of the *Drosophila* ubiquitin ligase DIAP1 is mediated via several distinct ubiquitin system pathways. *Cell Death Differ* 14 (4):861-871.

Hernández-Martínez, Salvador, Jaime G. Mayoral, Yiping Li, and Fernando G. Noriega. 2007. Role of juvenile hormone and allatotropin on nutrient allocation, ovarian development and survivorship in mosquitoes. *Journal of Insect Physiology* 53 (3):230-234.

Jones, Grace, Davy Jones, Lei Zhou, Hermann Steller, and Yanxia Chu. 2000. Deterin, a new inhibitor of apoptosis from *Drosophila melanogaster*. *Journal of Biological Chemistry* 275 (29):22157-22165.

Joshi, Vinod, Manju Singhi, and R. C. Chaudhary. 1996. Transovarial transmission of dengue 3 virus by *Aedes aegypti*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 90 (6):643-644.

Jousset, Françoise-Xavière. 1981. Geographic *Aedes aegypti* strains and dengue-2 virus: Susceptibility, ability to transmit to vertebrate and transovarial transmission. *Annales de l'Institut Pasteur. Virologie* 132 (3):357-370.

Kay, B. H., and J. G. Carley. 1980. Transovarial transmission of Murray Valley encephalitis virus by *Aedes aegypti* (L). *Aust J Exp Biol Med Sci* 58 (5):501-4.

Klowden, Marc J. 1987. Distention-mediated egg maturation in the mosquito, *Aedes aegypti*. *Journal of Insect Physiology* 33 (2):83-87.

Klowden, Marc J. 1997. Endocrine aspects of mosquito reproduction. *Archives of Insect Biochemistry and Physiology* 35 (4):491-512.

Leake, C. J. 1984. Transovarial transmission of arboviruses by mosquitoes. *Vectors in Virus Biology*:63-91.

LeDuc, J. W., J. F. Burger, B. F. Eldridge, and P. K. Russell. 1975. Ecology of Keystone virus, a transovarially maintained arbovirus. *Ann N Y Acad Sci* 266:144-51.

Levine, Beth, and Vojo Deretic. 2007. Unveiling the roles of autophagy in innate and adaptive immunity. *Nat Rev Immunol* 7 (10):767-777.

McLean, D. M., P. N. Grass, M. A. Miller, and K. S. Wong. 1975. Arbovirus growth in *Aedes aegypti* mosquitoes throughout their viable temperature range. *Arch Virol* 49 (1):49-57.

Mishra, AC, and DT Mourya. 2001. Transovarial transmission of West Nile virus in *Culex vishnui* mosquito. *Indian J Med Res.* 114:212-214.

Nezis, I. P., D. J. Stravopodis, L. H. Margaritis, and I. S. Papassideri. 2006. Autophagy is required for the degeneration of the ovarian follicular epithelium in higher Diptera. *Autophagy* 2 (4):297-8.

Nezis, I. P., D. J. Stravopodis, L. H. Margaritis, and I. S. Papassideri. 2006. Programmed cell death of follicular epithelium during the late developmental stages of oogenesis in the fruit flies *Bactrocera oleae* and *Ceratitis capitata* (Diptera, Tephritidae) is mediated by autophagy. *Dev Growth Differ* 48 (3):189-98.

Pantuwatana, Somsak, Wayne H. Thompson, Douglas M. Watts, Thomas M. Yuill, and Robert P. Hanson. 1974. Isolation of La Crosse virus from field collected *Aedes triseriatus* larvae. *Am J Trop Med Hyg* 23 (2):246-250.

Pondeville, Emilie, Annick Maria, Jean-Claude Jacques, Catherine Bourgouin, and Chantal Dauphin-Villemant. 2008. *Anopheles gambiae* males produce and transfer the vitellogenic steroid hormone 20-hydroxyecdysone to females during mating. *Proceedings of the National Academy of Sciences* 105 (50):19631-19636.

Pridgeon, Julia, Liming Zhao, James Becnel, Gary Clark, and Kenneth Linthicum. 2008. Developmental and environmental regulation of AaeIAP1 transcript in *Aedes aegypti*. *Journal of medical entomology* 45 (6):1071-1079.

Pritchett, Tracy L., Elizabeth A. Tanner, and Kimberly McCall. 2009. Cracking open cell death in the *Drosophila* ovary. *Apoptosis* 14 (8):969-979.

Qiu, Xiao-Bo, Shirley L. Markant, Junying Yuan, and Alfred L. Goldberg. 2004. Nrdp1-mediated degradation of the gigantic IAP, BRUCE, is a novel pathway for triggering apoptosis. *EMBO J* 23 (4):800-810.

Rosen, L., R. B. Tesh, J. C. Lien, and J. H. Cross. 1978. Transovarial transmission of Japanese encephalitis virus by mosquitoes. *Science* 199 (4331):909-11.

Rosen, Leon, Donald A. Shroyer, and Jih Ching Lien. 1980. Transovarial transmission of Japanese encephalitis virus by *Culex tritaeniorhynchus* mosquitoes. *Am J Trop Med Hyg* 29 (4):711-712.

Sabin, Leah R., Sheri L. Hanna, and Sara Cherry. 2010. Innate antiviral immunity in *Drosophila*. *Current Opinion in Immunology* 22 (1):4-9.

Salvesen, Guy S., and Stefan J. Riedl. 2007. Caspase mechanisms. In *Programmed Cell Death in Cancer Progression and Therapy*, edited by R. Khosravi-Far and E. White: Springer Netherlands.

Sancak, Yasemin, Timothy R. Peterson, Yoav D. Shaul, Robert A. Lindquist, Carson C. Thoreen, Liron Bar-Peled, and David M. Sabatini. 2008. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* 320 (5882):1496-1501.

Sandu, Cristinel, Hyung Don Ryoo, and Hermann Steller. 2010. Drosophila IAP antagonists form multimeric complexes to promote cell death. *The Journal of Cell Biology* 190 (6):1039-1052.

Schafer, Zachary T., and Sally Kornbluth. 2006. The apoptosome: physiological, developmental, and pathological modes of regulation. *Developmental Cell* 10 (5):549-561.

Shelly, Spencer, Nina Lukinova, Shelly Bambina, Allison Berman, and Sara Cherry. 2009. Autophagy is an essential component of *Drosophila* immunity against vesicular stomatitis virus. *Immunity* 30 (4):588-598.

Shiao, Shin-Hong, Immo A. Hansen, Jinsong Zhu, Douglas H. Sieglaff, and Alexander S. Raikhel. 2008. Juvenile hormone connects larval nutrition with target of rapamycin signaling in the mosquito *Aedes aegypti*. *Journal of Insect Physiology* 54 (1):231-239.

Siegel, Richard M. 2006. Caspases at the crossroads of immune-cell life and death. *Nat Rev Immunol* 6 (4):308-317.

Smith, Stan L., and Martin J. Mitchell. 1986. Ecdysone 20-monooxygenase systems in a larval and an adult dipteran: An overview of their biochemistry, physiology and pharmacology. *Insect Biochemistry* 16 (1):49-55.

Stones, P. B. 1960. Discussion. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 54 (2):130-134.

Telang, Aparna, Laura Frame, and Mark R. Brown. 2007. Larval feeding duration affects ecdysteroid levels and nutritional reserves regulating pupal commitment in the yellow fever mosquito *Aedes aegypti* (Diptera: Culicidae). *J Exp Biol* 210 (5):854-864.

Telang, Aparna, Yiping Li, Fernando G. Noriega, and Mark R. Brown. 2006. Effects of larval nutrition on the endocrinology of mosquito egg development. *J Exp Biol* 209 (4):645-655.

Tenev, Tencho, Anna Zachariou, Rebecca Wilson, Angela Paul, and Pascal Meier. 2002. Jafrac2 is an IAP antagonist that promotes cell death by liberating Dronc from DIAP1. *EMBO J* 21 (19):5118-5129.

Tesh, Robert B. 1980. Establishment of two cell lines from the mosquito *Toxorhynchites amboinensis* (Diptera: Culicidae) and their susceptibility to infection with arboviruses. *Journal of Medical Entomology* 17 (4):338-343.

Tesh, Robert B. 1980. Experimental studies on the transovarial transmission of Kunjin and San Angelo viruses in mosquitoes. *Am J Trop Med Hyg* 29 (4):657-666.

Tesh, Robert B., Leon. Rosen, Barry J. Beaty, and Thomas H. G. Aitken. 1979. Studies of transovarial transmission of yellow fever and Japanese encephalitis viruses in *Aedes* mosquitoes and their implications for the epidemiology of dengue. *Anon. Dengue in the Caribbean: Proc Pan American Health Organization*:179-182.

Tesh, Robert B., and Donald A. Shroyer. 1980. The mechanism of arbovirus transovarial transmission in mosquitoes: San Angelo virus in *Aedes albopictus*. *Am J Trop Med Hyg* 29 (6):1394-1404.

Thomas, Leo A. 1963. Distribution of the virus of Western equine encephalomyelitis in the mosquito vector, *Culex tarsalis*. *American Journal of Epidemiology* 78 (2):150-165.

Turell, M. J., and J. L. Hardy. 1980. Carbon dioxide sensitivity of mosquitoes infected with California encephalitis virus. *Science* 209 (4460):1029-30.

Velentzas, A. D., I. P. Nezis, D. J. Stravopodis, I. S. Papassideri, and L. H. Margaritis. 2007. Mechanisms of programmed cell death during oogenesis in *Drosophila virilis*. *Cell Tissue Res* 327 (2):399-414.

Wang, Hua, and Rollie Clem. 2011. The role of IAP antagonist proteins in the core apoptosis pathway of the mosquito disease vector *Aedes aegypti*. *Apoptosis* 16 (3):235-248.

Watts, D. M., P. R. Grimstad, G. R. DeFoliart, and T. M. Yuill. 1975. *Aedes hendersoni*: failure of laboratory-infected mosquitoes to transmit LaCrosse virus (California encephalitis group). *J Med Entomol* 12 (4):451-3.

Watts, D. M., S. Pantuwatana, G. R. DeFoliart, T. M. Yuill, and W. H. Thompson. 1973. Transovarial transmission of La Crosse virus (California encephalitis group) in the mosquito, *Aedes triseriatus*. *Science* 182 (4117):1140-1141.

Watts, Douglas M., Wayne H. Thompson, Thomas M. Yuill, Gene R. DeFoliart, and Robert P. Hanson. 1974. Overwintering of La Crosse virus in *Aedes triseriatus*. *Am J Trop Med Hyg* 23 (4):694-700.

Yoo, Soon Ji, Jun R. Huh, Israel Muro, Hong Yu, Lijuan Wang, Susan L. Wang, R. M. Renny Feldman, Rollie J. Clem, H. Arno J. Muller, and Bruce A. Hay. 2002. Hid, Rpr and Grim negatively regulate DIAP1 levels through distinct mechanisms. *Nat Cell Biol* 4 (6):416-424.

Yuan, Junying, and H. Robert Horvitz. 2004. A first insight into the molecular mechanisms of apoptosis. *Cell* 116 (Supplement 2):S53-S56.

Zhang, X., J. Zhang, and K. Y. Zhu. 2010. Chitosan/double-stranded RNA nanoparticle-mediated RNA interference to silence chitin synthase genes through larval feeding in the African malaria mosquito (*Anopheles gambiae*). *Insect Molecular Biology* 19 (5):683-693.

Zhou, Lei, Guohua Jiang, Gina Chan, Carl P Santos, David W Severson, and Lei Xiao. 2005. Michelob_x is the missing inhibitor of apoptosis protein antagonist in mosquito genomes. *EMBO Reports* 6 (8):769-774.

CHAPTER 2. MOLECULAR MECHANISMS OF NUTRITIONAL SIGNALING DURING OOGENESIS IN *AEDES TRISERIATUS*

Patrick Jennings and Lyric Bartholomay

Abstract

The target of rapamycin (TOR) pathway is a highly conserved mechanism of nutritional signaling in eukaryotes, and it plays a central role in mosquito egg development. The roles of TOR and its downstream target S6K were assessed by *in vivo* RNA interference (RNAi) in the Eastern treehole mosquito, *Aedes* (*Ochlerotatus*) *triseriatus* (Say). TOR transcript knockdown by RNAi slows ovarian development and increases follicular atresia, resulting in fewer eggs. Knockdown of S6K transcripts by RNAi confers a similar effect to egg development *in vivo*. These results support the current model of TOR signaling in mosquitoes and help to clarify the downstream effects of TOR that relate to oogenesis.

Introduction

Encephalitis of arbovirus origin is one of the most common and harmful disease manifestations of mosquito-borne viruses in the United States. With changing climates and an ever-increasing human population, the incidence of arboviral encephalitis is likely to increase (McMichael et al. 2003; Wilson, Lush, and Baker 2010). In some species, the virus infects the ovaries of the mosquito and passes vertically to her offspring. This process, called transovarial transmission, facilitates the maintenance of the virus in mosquito populations, especially in aedine mosquitoes that diapause in the egg stage. However, little is known about why transovarial transmission of viruses can occur in certain mosquito species but not

in others (Erickson et al. 2006). Surprisingly little is known about the molecular nuances of oogenesis within the mosquito ovariole. We reasoned that cell death and autophagy—both of which are inherently antiviral and physiologically essential processes—play pivotal roles during oogenesis in determining the efficacy of transovarially transmitted mosquito-borne virus transmission.

Mosquitoes have polytrophic meroistic ovaries, each containing as many as 250 ovarioles. After eclosion, ovarioles develop in size until they reach Christophers' stage IIa or IIb (Christophers 1911). In anautogenous mosquitoes, which are more epidemiologically significant vectors of human disease, the development of ovarioles is arrested at this stage until a blood meal is taken. In the adult stage, mosquitoes feed on a low-amino acid diet of flower nectar, and only take a blood meal when additional protein is required for egg development. This blood meal provides the necessary protein for synthesis of vitellogenin (the major yolk protein precursor) in the mosquito fat body. When the protein content of a blood meal is insufficient for development of every primary follicle in the ovaries, some follicles will degrade by the process of follicular atresia, and the protein content of those atretic follicles is reallocated to healthy, developing follicles. This was demonstrated by feeding mosquitoes to repletion on a blood meal that was diluted with various concentrations of saline, causing a decrease in the number of follicles that developed to maturity with a decrease in blood content of the meal (Klowden 1987; Uchida et al. 2004).

Regardless of the number of ovarioles present at the beginning of the gonotrophic cycle, the volume of blood ingested and the source of the blood meal have a strong effect on the number of follicles that develop to maturity. More follicles are degraded by the process of follicular atresia if teneral protein reserves are low, if a small blood meal is taken, or if an

unsuitable vertebrate host is the blood source (Pritchett, Tanner, and McCall 2009; Clements and Boocock 1984). Several proteins and neuropeptides have been identified in the process of sensing a blood meal and initiating the vitellogenic phase of follicle development. Among the most important of the protein cascades triggered is the target of rapamycin (TOR) pathway, which activates vitellogenesis (Hansen et al. 2004) and stimulates cell growth and protein synthesis during oogenesis (Brandon et al. 2008; Hansen et al. 2005).

The TOR kinase is a serine/threonine protein kinase that is involved in nutritional signaling. Amino acids acquired in the blood meal activate TOR complex 1 (TORC1), a protein complex that includes TOR, LST8, and raptor (regulatory associated protein of TOR). TORC1 affects translation, metabolism, and many other cellular processes (Hay and Sonenberg 2004). Interestingly, autophagy is intimately linked with the TOR pathway because TOR shuts off autophagy when nutrients are abundant and regulates nutrient availability to developing follicles in mosquitoes (Hansen et al. 2004). Only recently has autophagy been linked to cell death functions during oogenesis in flies; based on those results, autophagy is occurring at times when cell death (follicular atresia) is directly related to the nutritional status of the female fly (Nezis et al. 2006).

Reverse genetics studies have shown that knockdown of *Aedes aegypti* TOR by RNAi results in a decrease in *Ae*. *aegypti* vitellogenin production, inhibits egg development, and decreases both egg production and viability (Hansen et al. 2004). TOR phosphorylates the Gln-3 GATA transcription factor and affects its nuclear translocation. The phosphorylated Gln-3 then binds upstream of the vitellogenin gene (*Vg*) and displaces a GATA repressor protein, causing a decrease in vitellogenin transcription and a corresponding decrease in egg development (Park et al. 2006; Attardo et al. 2003).

TOR is clearly necessary for nutritional signaling and initiation of vitellogenin synthesis (Raikhel et al. 2002; Hansen et al. 2004), but TOR also plays an important role in the growth and development of oocytes. TORC1 activates p70 ribosomal S6 kinase (S6K) and deactivates 4E binding protein (4EBP), a repressor of eukaryotic translation initiation factor 4E (eIF4E), leading to an increase in cell growth and proliferation. Hansen *et al*. (2004) found that TOR knockdown by RNAi interferes with *Vg* transcription in fat body cell culture and inhibits follicle development *in vivo* in *Ae*. *aegypti*. After two injections with double-stranded TOR RNA and subsequent blood feeding, *Ae. aegypti* mosquitoes developed fewer eggs, and the egg viability dropped from 75% in control dsRNA-injected mosquitoes to 27% viability in the TOR dsRNA-injected mosquitoes (Hansen et al. 2004). The phosphorylation of S6K by TORC1 appears to be a key factor in blood meal-induced activation of mosquito vitellogenesis and egg development (Hansen et al. 2005). and others have elucidated some of the complex molecular mechanisms of TOR signaling in mosquito egg development (Park et al. 2006; Roy, Hansen, and Raikhel 2007; Attardo et al. 2006; Avruch et al. 2009; Liao et al. 2008; Shiao et al. 2008). However, little work has been done to understand the effects of TOR and its downstream targets on specific aspects of cell death in the mosquito ovary during oogenesis.

Because of the multiple ways that TOR affects vitellogenesis and egg development, the TOR kinase pathway could be an important focus for development of novel modes of action for insecticides, but in mosquitoes, the pathway is not well characterized *in vivo*. Knockdown of TOR leads to delayed ovarian development and greatly reduced fecundity (Hansen et al. 2004), but it is not clear whether this decrease in fecundity is a result of decreased vitellogenin transcription, decreased S6K phosphorylation, or decreased 4EBP

phosphorylation. Because the regulation of TOR and S6K are likely to be complex *in vivo* (Hansen et al. 2005), the approach used in this study was to inject mosquitoes with dsRNA, blood feed, and observe the effects on ovarian development by dissection at specified intervals after the blood meal. Here, we used *in vivo* RNAi knockdown experiments to reveal the effects of TOR and its downstream target, S6K, during mosquito egg development.

Materials and Methods

Mosquito Culture.

The *Ae*. *triseriatus* colony used in this study was established in 2003 from adults collected in Scott County, IA (Erickson et al. 2006). The *Ae*. *triseriatus* colony and a colony of *Ae*. *aegypti* Liverpool strain were maintained in a controlled environmenal chamber (27 ± 1°C and 80 ± 5% RH with a 16:8-h photoperiod) and fed a 10% sucrose solution.

Sequencing.

Total RNA was extracted from 5 adult female mosquitoes using the one-step phenol/guanidinium thiocyanate method (TRIzol, Invitrogen). Whole-body RNA was used to create cDNA using a commercially available reverse transcription polymerase chain reaction (RT-PCR) kit (SuperScript™ III One-step RT-PCR System with Platinum® Taq DNA Polymerase, Invitrogen). A forward primer was designed from the highly conserved FAT domain of *Ae*. *aegypti* TOR sequence AY438003.1, and a reverse primer was designed from the PI3K catalytic domain of the same sequence (Forward: 5′-GGCGGAGACAAAGA GAAGTG; Reverse: 5′-ATCACCTGGAGGTTCGACTG). The PCR products were excised from an agarose gel and purified using a gel extraction kit (QIAquick GEL Extraction Kit,

Qiagen). An 884 bp PCR product was ligated into the pCR4-TOPO vector (Invitrogen), and the vector was transformed into One Shot® Mach1™-T1^{R} competent *E. coli*. Plasmid DNA was purified from an overnight culture of a single colony and sequenced with a standard T7 primer. *Ae*. *triseriatus* specific primers (AtrTOR-F and AtrTOR-R) were chosen from the sequence, and a primer walking technique was used to sequence the remainder of the gene. The 3′ and 5′ ends were sequenced using a rapid amplification of cDNA ends (RACE) PCR protocol (Beck et al. 2007). The full-length transcript sequence was later confirmed and completed using Illumina mRNA-seq, and Bowtie (Langmead et al. 2009) was used to map 104-nt short reads against sequences generated by traditional Sanger sequencing.

Illumina mRNA-seq was used to sequence *Aedes triseriatus* ribosomal S6K (AtrS6K). Briefly, poly-A mRNA was isolated using Illumina's mRNA-seq prep kit, mRNA was fragmented by addition of divalent cations under elevated temperature, double-stranded cDNA (dscDNA) was synthesized (SuperScript II/DNA polymerase I), blunt ends were created with T4 DNA polymerase and Klenow DNA polymerase, 3′ ends were adenylated, and Illumina PE PCR adapters were ligated to the fragmented dscDNA. The cDNA templates were size selected by agarose gel electrophoresis and amplified with 15 cycles of PCR. Sequencing was performed on an Illumina GA_{II} System. Assembly of AtrS6K was accomplished by matching the 104 nt reads against *Aedes aegypti* S6K transcript sequence using the Perl pipeline program, *Aedenovo* (see Chapter 3).

Rapamycin Injections.

Forty female *Ae*. *triseriatus* in three biological replicates were cold-anesthetized 5 days post-eclosion and injected intrathoracically with 0.20 μL rapamycin, which was

dissolved in dimethyl sulfoxide (DMSO) and diluted in phosphate-buffered saline (PBS) buffer to 20 μM. Forty mosquitoes were cold-anesthetized and injected with DMSO/PBS solution, and forty were cold-anesthetized but not injected. All mosquitoes were maintained on 10% sucrose for 24 h and then fed 34% PCV defibrinated sheep blood through a Parafilm membrane on a water-jacketed membrane feeder (Rutledge, Ward, and Gould 1964). Mosquitoes that did not take a full blood meal were removed. Ovaries were dissected into 0.9% saline, stained for 10 seconds in neutral red (Clements and Boocock 1984), and destained in phosphate-buffered saline.

RNA Interference.

Mosquitoes were injected with several different types of double-stranded RNA (dsRNA). Hereafter, dsRNA will be referred to by the abbreviation of the mRNA that was used as a template for creation of the dsRNA, preceded by the letters "ds" (*e.g.* "dsTOR" refers to double-stranded RNA synthesized from a 422 nt region of target of rapamycin mRNA).

All dsRNA was synthesized *in vitro* using the MEGAscript® RNAi Kit (Ambion) according to the manufacturer's protocol with a 10-hour transcription step, and the purified dsRNA was concentrated to 1-5 μg/μL under vacuum centrifugation at 42°C. A 422 nt segment from the 3′ end of the *Ae. triseriatus* TOR gene was amplified by PCR and used for synthesis of dsTOR. Plasmid DNA from enhanced green fluorescent protein (EGFP)-expressing *E. coli* was extracted, and a 720 bp region was amplified by PCR with primers designed by Dong *et al.* (2006) and used for synthesis of dsEGFP. *Ae. triseriatus* S6K mRNA was sequenced and assembled with Velvet (Zerbino and Birney 2008) from Illumina

RNA-seq short reads, and primers were designed from the *de novo* transcript assembly (see methods in Chapter 3). A 384 nt region was amplified by PCR and used for synthesis of dsS6K. The primer sequences for each dsRNA synthesis and real-time PCR are presented in Table 1.

Adult female mosquitoes at one day post-eclosion were injected once with 0.2 – 1 µg dsTOR or dsEGPF in 0.2 µL ultrapure water and maintained for 3 days on sucrose. In addition to the dsEGPF control, a third group of mosquitoes was injected with 0.20 µL 0.9% saline, and a fourth group was not injected but was cold-anesthetized and handled in an identical manner. Sucrose was replaced with water 14 hours before blood feeding, and mosquitoes were blood fed as described above. After 26 hours, the ovaries were dissected in phosphate-buffered saline and stained for 10 seconds in neutral red (Clements and Boocock 1984). The carcasses were stored in a ribonuclease inhibitor (RNA*later*, Ambion) at 4°C, and RNA was extracted (TRIzol, Invitrogen) for real-time PCR analysis of TOR knockdown.

Verification of transcript knockdown by RNAi.

Knockdown of TOR and S6K by dsRNA injection was confirmed by real-time PCR. Total RNA was extracted from the mosquito carcasses after ovary dissections, and mRNA was reverse transcribed in triplicate with oligo$(dT)_{20}$ primers (ThermoScript two-step RT-PCR kit, Invitrogen). Real-time PCR was performed with actin primers, S6K primers, and TOR primers (Table 1) on a Thermo light cycler real-time PCR machine using iQ SYBR Green Supermix (Bio-Rad) according to the manufacturer's protocol. Ct values were adjusted to calculate relative target transcript abundance based on standard curves, and transcript abundance was normalized using *Aedes triseriatus* β-actin as a reference gene.

Results

Sequence analyses.

The full-length open reading frame of *Aedes triseriatus* TOR cDNA sequence (8099 nt; GenBank Accession no. HQ875574) predicts a protein of 2449 amino acids. We named this protein *Aedes triseriatus* target of rapamycin (AtrTOR). The predicted protein sequence includes a kinase domain, a FK506-BP 12-binding domain, and a FATC domain, which are highly conserved across all eukaryotes (Fig. 1), as well as a FAT domain and N-terminal HEAT repeats that are not as highly conserved. The full-length *Ae. triseriatus* TOR mRNA sequence contains a 458 nt 5′ untranslated region (UTR) and a 291 nt 3′ UTR. *Ae. triseriatus* TOR shares 95% amino acid identity (97% similarity) with *Ae. aegypti* TOR (Fig. 1).

The cDNA sequence of *Aedes triseriatus* S6K (3008 nt; GenBank Accession no. JF746878) was determined by assembly of Illumina mRNA-seq short reads. The full-length open reading frame predicts a protien of 552 amino acids and includes the highly conserved STKc_p70S6K domain, and we named the predicted protein *Aedes triseriatus* ribosomal S6 kinase (AtrS6K). The full-length predicted AtrS6K protein sequence shares 89% amino acid identity (91% similarity) with *Aedes aegypti* S6 kinase.

Rapamycin and AtrTOR knockdown increase follicular atresia *in vivo*.

Ae. triseriatus injected with 0.2 µL rapamycin (20 µM in PBS/DMSO) showed increased follicular atresia and reduced follicle development compared to PBS/DMSO injected mosquitoes (Fig. 2). Mosquitoes injected with 0.2-1 µg dsTOR had fewer normally developing ovarian follicles and more atretic follicles at both 24 hours post blood meal (hpbm)(data not shown; $p < 0.05$) and 48 hpbm (Fig. 3a; $F_{2,26} = 3.37$, $p < 0.01$) as indicated

by neutral red staining (Fig. 3b). Real-time PCR indicated a 40% decrease in TOR transcript abundance in the total RNA extracted at 48 hpbm.

AtrTOR knockdown delays the ovarian developmental cycle.

The follicles of mosquitoes injected with dsTOR developed more slowly than dsEGFP (control)-injected mosquitoes. Follicles from dsTOR-injected mosquitoes were all classified at Christophers' stage V (follicles 75-90% occupied by vitellin, no signs of nurse cell apoptosis) when dissected 48 hpbm, whereas dsEGFP-injected mosquitoes at 48 hpbm were classified at stage VI (follicles >90% occupied by vitellin and nurse cells showing signs of apoptosis). The strain of *Ae. triseriatus* used in these experiments normally oviposit 4 days after a blood meal, but *Ae. triseriatus* injected with 0.2 μg dsTOR oviposited 7-9 days pbm.

AtrS6K knockdown increases follicular atresia and delays follicle growth.

Mosquito ovaries injected with 1 μg dsS6K were phenotypically similar to those of the dsTOR-injected mosquitoes. The number of atretic follicles in each ovary 24 hpbm increased by 59% in dsS6K-injected *Ae. triseriatus* ($F_{2,151} = 9.774$, $p < 0.01$, Fig. 5a), and a corresponding decrease was observed in the number of healthy, developing follicles ($F_{2,151} = 7.426$, $p < 0.01$, Fig. 5b). Preliminary experiments to knock down by RNAi another downstream target of TOR, eukaryotic initiation factor 4E-binding protein (4EBP) in *Ae. aegypti* did not affect follicular atresia; the control (dsEGFP) injected groups and ds4EBP injected treatment groups showed no difference in the amount of follicular atresia or follicle growth at 36 hpbm (data not shown; $p > 0.05$).

Discussion

TOR and its downstream target, S6K, are clearly important regulators of amino acid signaling. Here we showed that TOR and S6K play vital roles in the transduction of amino acid nutritional signals to the oocytes, because knockdown of TOR and S6K transcripts by RNAi increase follicular atresia *in vivo*.

The effects of TOR RNAi on follicular atresia were not as pronounced as rapamycin injections. This could be explained by the incomplete knockdown of TOR and S6K transcripts by *in vivo* RNAi. Mosquitoes exhibit a dose-dependent and tissue-dependent response to injection of dsRNA (Boisson et al. 2006), and many TOR transcripts may evade RNAi degradation and generate functional TOR protein. Inefficient transcript knockdown has also been observed in mosquito RNAi experiments when shRNA was delivered in a sindbis virus vector to knock down transcription of a GATA transcription factor, *AaGATAr* (Attardo et al. 2003). However, in our current study, comparisons can be made between RNAi knockdown of TOR and S6K, because identical experimental conditions were used for dsTOR and dsS6K injection. Injection of dsS6K resulted in a similar phenotype to dsTOR-injected mosquito ovaries, but the induction of follicular atresia was less pronounced than in dsTOR-injected mosquitoes. S6K could be more refractory to RNAi than TOR, but it is also likely that TOR regulates follicular atresia through other targets, not only through phosphorylation of S6K. 4EBP is known to be inhibited by TORC1, but our preliminary studies on 4EBP knockdown in *Ae. aegypti* showed that ds4EBP injection did not inhibit follicular atresia. This result is supported in a study by LaFever *et al.* (2010), who found that the effects of TOR on proliferation, growth, and survival in germline *Drosophila* cells were independent of 4EBP. Further studies on 4EBP and its target, eIF4E, are necessary to

confirm that 4EBP is not a major effector of follicular atresia in mosquitoes, but other downstream targets should also be investigated.

For example, *myc* is one conserved gene that has been shown to be regulated by TOR in *Drosophila* and other animals (LaFever et al. 2010; Teleman et al. 2008; Johnston et al. 1999), but its function has not been studied in mosquitoes. *Drosophila* Myc (dMyc) and its binding partner, Max, together function as a sequence-specific bHLH/LZ transcription factor that regulates expression of genes involved in protein synthesis, apoptosis, and metabolism (Boon et al. 2001; Guo et al. 2000; Orian et al. 2003; Watson et al. 2002). dMyc, which corresponds to the *diminutive* locus (*dm*) in *Drosophila*, plays an essential role in *Drosophila* oogenesis by regulating growth and DNA endoreplication in the follicular epithelium and nurse cells, and *dm* mutants develop smaller ovarioles (Maines et al. 2004). Since dsS6K injections in *Ae. triseriatus* only partially replicated the phenotype observed when dsTOR was injected, the effects of Myc and other targets of the TOR pathway should be investigated in mosquito oogenesis.

In summary, TOR and S6K were identified as key regulators of follicular atresia in the Eastern treehole mosquito, *Aedes triseriatus* (Say). The regulation of autophagy and apoptosis by TOR and S6K are likely to be more complex during a persistent viral infection, and future research will undoubtedly reveal links between the TOR pathway and the mechanisms of transovarial transmission of arboviruses.

Acknowledgments

We thank Caleb Robb for assistance with dsS6K RNAi and his help with mosquito dissections, Grishma Parikh for assistance with dsTOR injections, and Brendan Dunphy for mosquito colony maintenance. This work was funded by a grant from the Roy J. Carver Charitable Trust.

Figures/Tables

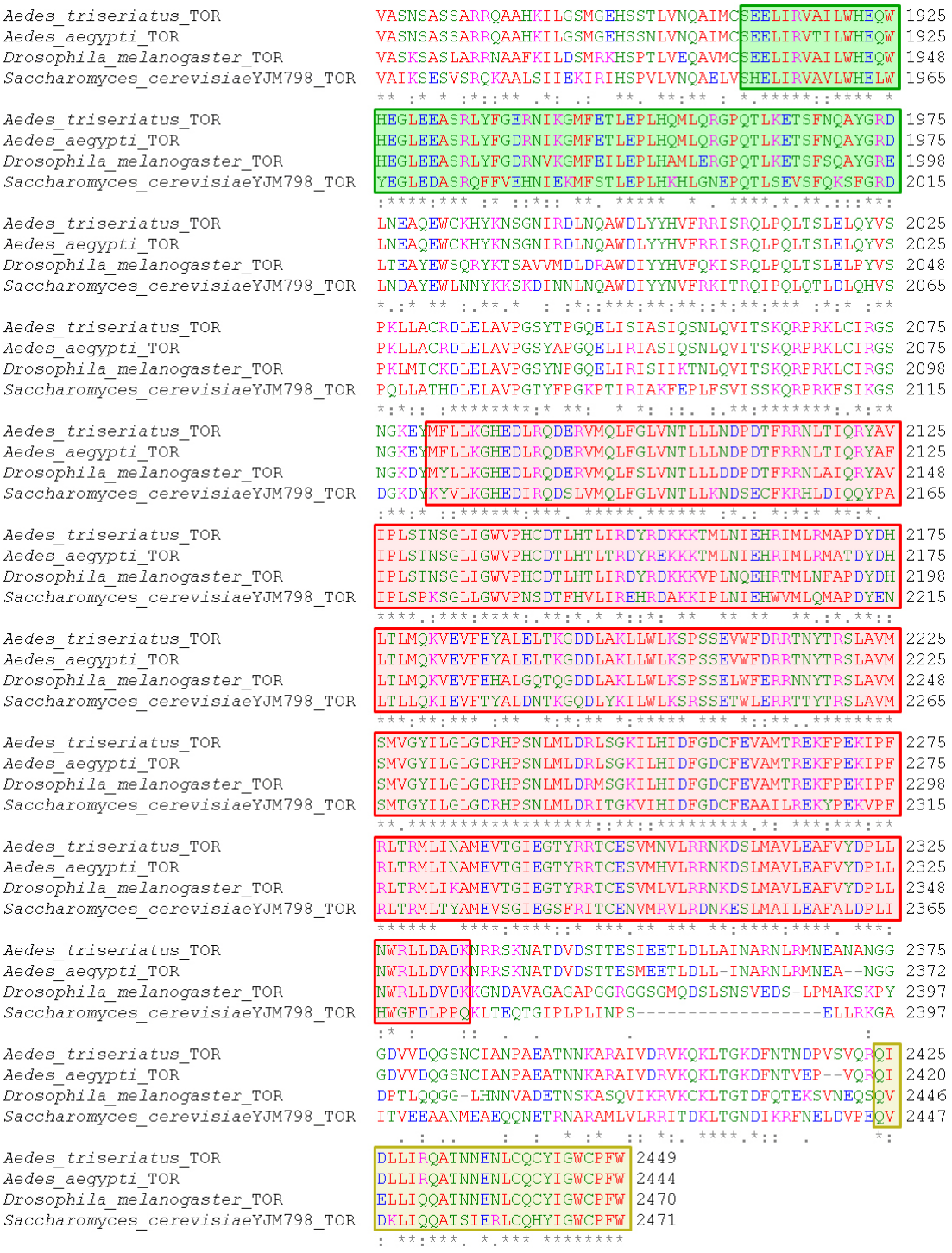


Figure 1. ClustalW2 sequence alignment of the C-terminal ends of *Ae. triseriatus* target of rapamycin (AtrTOR, GenBank accession no. HQ875574) with TOR proteins in three other species. The highlighted regions indicate the conserved FBP (green), kinase (pink), and FATC (yellow) protein domains.

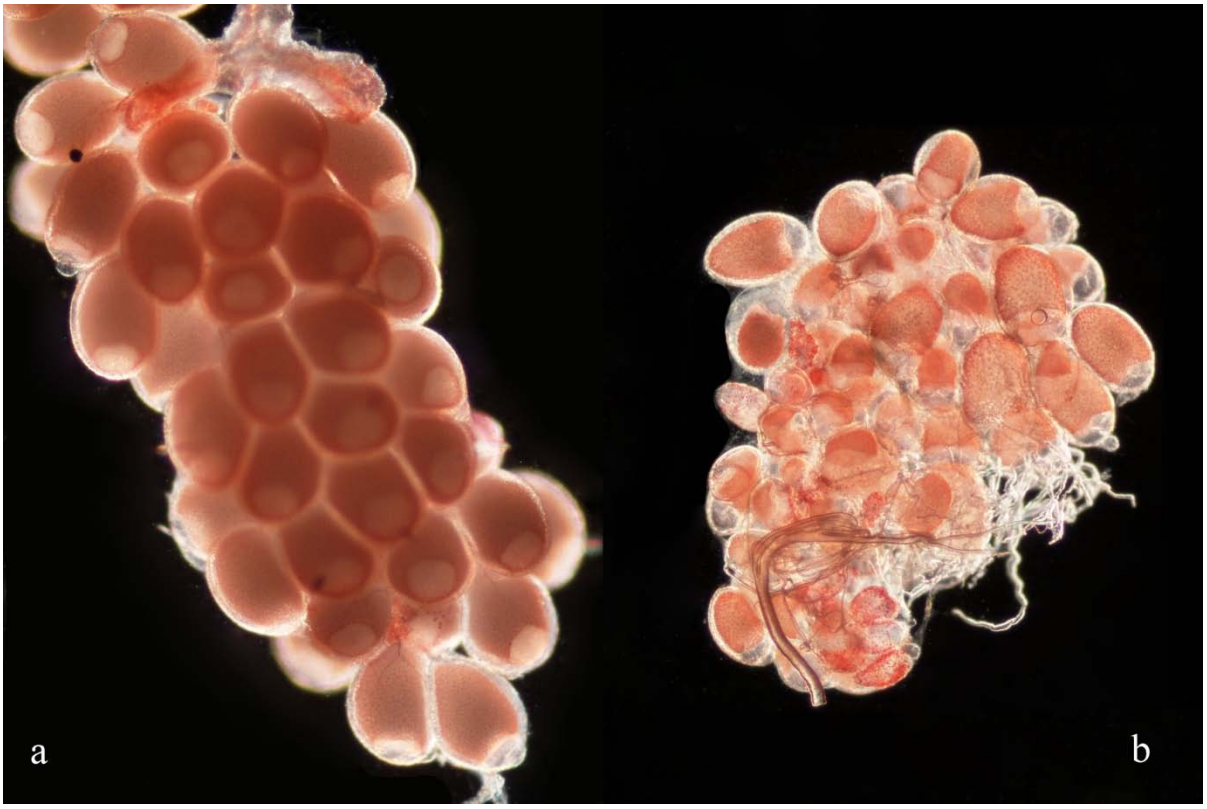


Figure 2. Typical ovary development of *Aedes triseriatus* 48 h after injection with PBS/DMSO (a) and 20 µM rapamycin in PBS/DMSO (b).

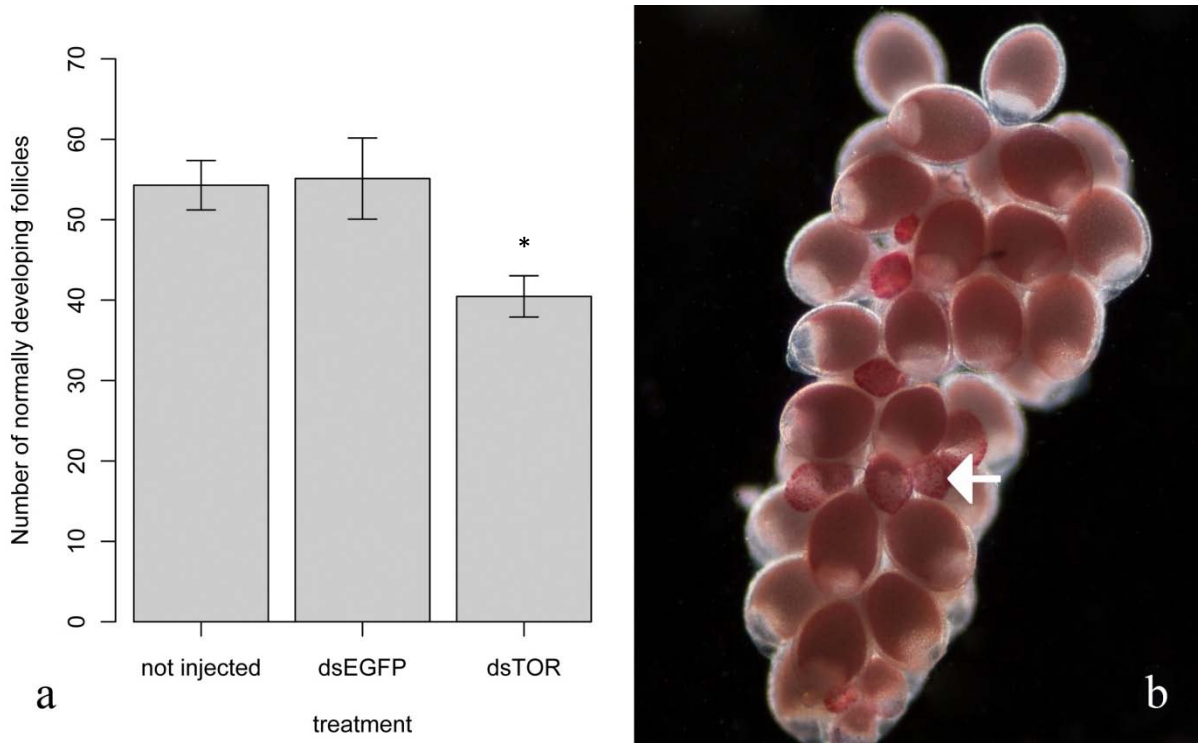


Figure 3. dsTOR injected mosquitoes have fewer developing follicles after a blood meal. (a) Average number of normally developing *Aedes triseriatus* follicles 48h post blood meal. "Not injected" mosquitoes were cold anesthetized. dsEGFP and dsTOR mosquitoes were cold anesthetized and injected with 1 µg dsRNA dissolved in 0.2 µL sterile water. All mosquitoes were injected 24 h post-eclosion and allowed 3 days to recover before feeding on sheep blood. dsTOR-injected mosquitoes had fewer follicles than both control groups *($F_{2,26}$ = 3.37, p = 0.008). (b) A neutral red stained, dsTOR-injected *Ae. triseriatus* ovary 48 hpbm. The arrow shows an atretic follicle, characterized by bright red staining and smaller size.

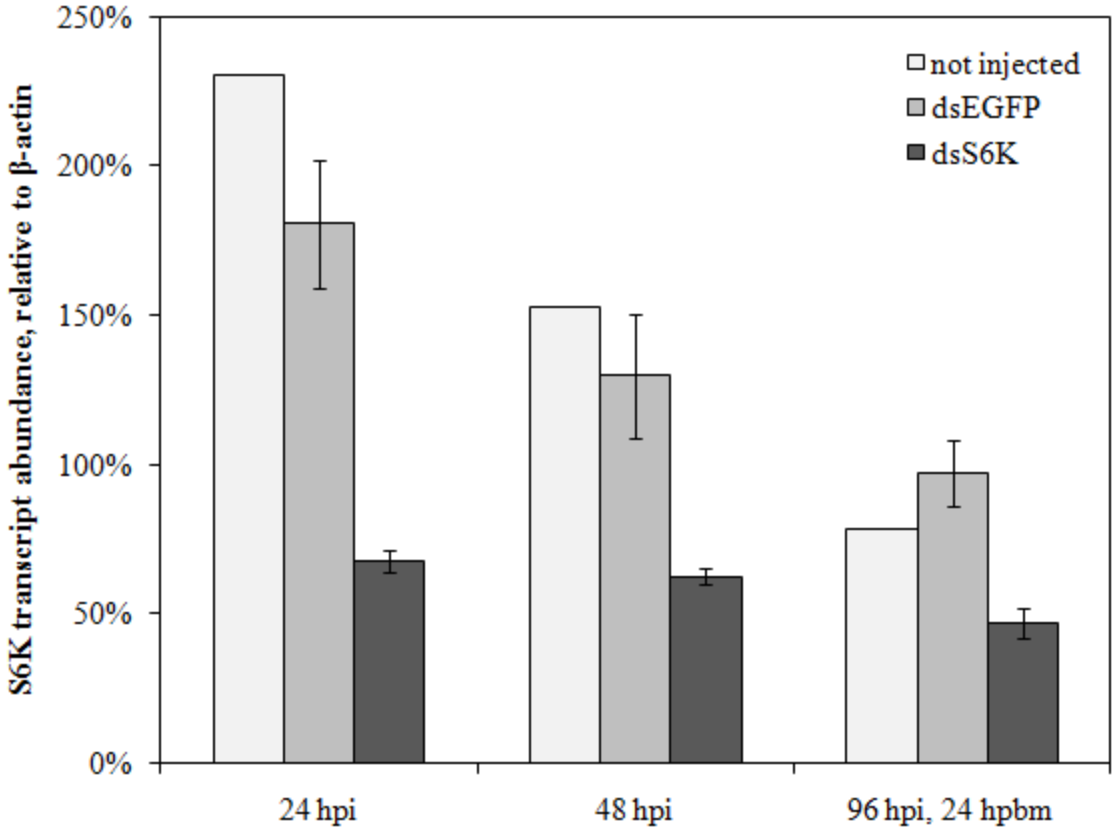


Figure 4. Quantitative RT-PCR analysis of *Aedes triseriatus* (*n* = 5 mosquitoes) whole-body S6K mRNA transcript abundance at 24, 48, and 96 hours post injection of dsRNA (mock injection, control dsEGFP injection, and target gene dsS6K injection). Mosquitoes were fed a blood meal at 72 hours post injection (hpi). Values shown represent S6K transcript copy number relative to *Aedes triseriatus* β-actin copy number.

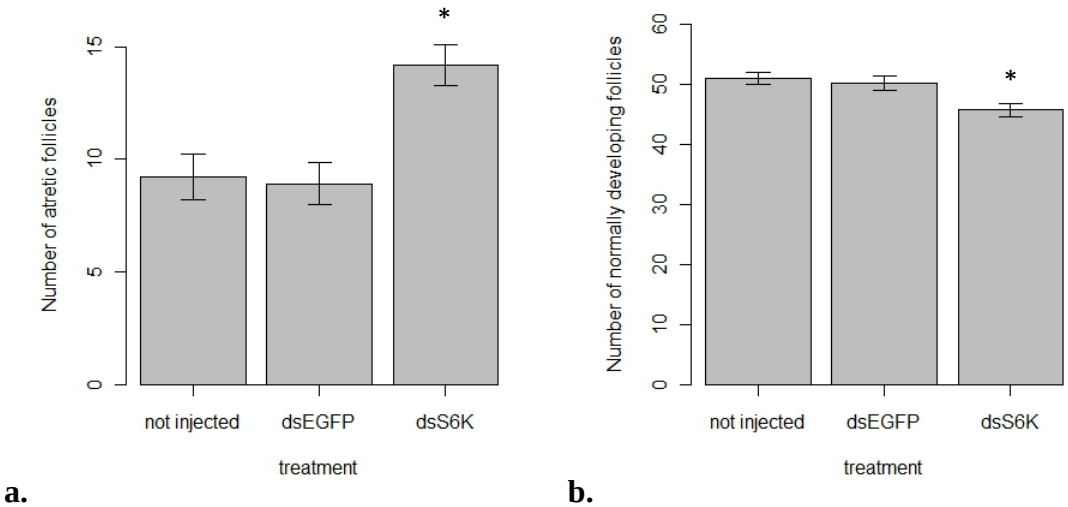


Figure 5. dsS6K increased follicular atresia and decreased the number of normally developing *Aedes triseriatus* follicles *in vivo*. (a) The number of atretic follicles 48 hpbm increased by 59% in dsS6K injected mosquitoes *($F_{2,151}$ = 9.774, p = 0.0001). There was no significant difference between the number of atretic follicles in dsEGFP-injected mosquitoes and mosquitoes that were not injected. (b) The number of normally developing follicles 48 hpbm decreased by 8.9% in dsS6K injected mosquitoes *($F_{2,151}$ = 7.426, p = 0.0008). There was no significant difference in the number of normally developing follicles between dsEGFP-injected mosquitoes and mosquitoes that were not injected.

Table 1. Oligonucleotide primers used for real-time PCR and dsRNA synthesis. Bold nucleotides represent T7 promoter regions and other nucleotides are gene-specific.

AtrTOR-F	5′-GATTTCAACACCAACGATCCG
AtrTOR-R	5′-GAGAGCCACTTTCTACCAGAAC
T7AtrTOR-F	5′-**TAATACGACTCACTATAGGGAGA**AAGATCCCGTTCCGTTTGAC
T7AtrTOR-R	5′-**TAATACGACTCACTATAGGGAGA**CGGTCAGTTTCTGCTTCACA
AtrS6K-F	5′-CACAGTTCGACACCAAATTCAC
AtrS6K-R	5′-TCGTGGCTGTTGCATCTC
T7AtrS6K-F	5′-**TAATACGACTCACTATAGGGAGA**GGCACCTGAAATCCTAACAA
T7AtrS6K-R	5′-**TAATACGACTCACTATAGGGAGA**TTGGTGAATTTGGTGTCGAA
AtrActin-F	5′-ATCATTGCCCCGCCAGAGCG
AtrActin-R	5′-AAGGTGGACAGGAAGCCAGG
EGFP RNAi f	5′-**TAATACGACTCACTATAGGGAGA**ATGGTGAGCAAGGGCGAGGAGCTGT
EGFP RNAi r	5′-**TAATACGACTCACTATAGGGAGA**TTACTTGTACAGCTCGTCCATGCCG

References

Attardo, Geoffrey M., Immo A. Hansen, Shin-Hong Shiao, and Alexander S. Raikhel. 2006. Identification of two cationic amino acid transporters required for nutritional signaling during mosquito reproduction. *Journal of Experimental Biology* 209 (16):3071-3078.

Attardo, Geoffrey M., Stephen Higgs, Kimberley A. Klingler, Dana L. Vanlandingham, and Alexander S. Raikhel. 2003. RNA interference-mediated knockdown of a GATA factor reveals a link to anautogeny in the mosquito *Aedes aegypti*. *Proceedings of the National Academy of Sciences of the United States of America* 100 (23):13374-13379.

Avruch, Joseph, Xiaomeng Long, Sara Ortiz-Vega, Joseph Rapley, Angela Papageorgiou, and Ning Dai. 2009. Amino acid regulation of TOR complex 1. *American Journal of Physiology - Endocrinology And Metabolism* 296 (4):E592-E602.

Beck, Eric T., Carol D. Blair, William C. Black Iv, Barry J. Beaty, and Bradley J. Blitvich. 2007. Alternative splicing generates multiple transcripts of the inhibitor of apoptosis protein 1 in *Aedes* and *Culex* spp. mosquitoes. *Insect Biochemistry and Molecular Biology* 37 (11):1222-1233.

Boisson, Bertrand, Jean Claude Jacques, Valérie Choumet, Estelle Martin, Jiannong Xu, Ken Vernick, and Catherine Bourgouin. 2006. Gene silencing in mosquito salivary glands by RNAi. *FEBS Letters* 580 (8):1988-1992.

Boon, Kathy, Huib N. Caron, Ronald van Asperen, Linda Valentijn, Marie-Christine Hermus, Peter van Sluis, Ilja Roobeek, Isabel Weis, P. A. Voûte, Manfred Schwab, and Rogier Versteeg. 2001. N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J* 20 (6):1383-1393.

Brandon, Michelle C., James E. Pennington, Jun Isoe, Jorge Zamora, Anne-Sophie Schillinger, and Roger L. Miesfeld. 2008. TOR signaling is required for amino acid stimulation of early trypsin protein synthesis in the midgut of *Aedes aegypti* mosquitoes. *Insect Biochemistry and Molecular Biology* 38 (10):916-922.

Christophers, S. R. 1911. The development of the egg follicle in anophelines. *Paludism* 2:73-88.

Clements, A. N., and M. R. Boocock. 1984. Ovarian development in mosquitoes: stages of growth and arrest, and follicular resorption. *Physiological Entomology* 9 (1):1-8.

Dong, Yuemei, Ruth Aguilar, Zhiyong Xi, Emma Warr, Emmanuel Mongin, and George Dimopoulos. 2006. *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. *PLoS Pathogens* 2 (6):e52.

Erickson, S. M., K. B. Platt, B. J. Tucker, R. Evans, S. Tiawsirisup, and W. A. Rowley. 2006. The potential of *Aedes triseriatus* (Diptera: Culicidae) as an enzootic vector of West Nile virus. *Journal of Medical Entomology* 43 (5):966-970.

Guo, Qingbin M., Renae L. Malek, Sunkyu Kim, Chia Chiao, Mei He, Mauro Ruffy, Krishna Sanka, Norman H. Lee, Chi V. Dang, and Edison T. Liu. 2000. Identification of c-Myc responsive genes using rat cDNA microarray. *Cancer Research* 60 (21):5922-5928.

Hansen, IA, GM Attardo, JH Park, Q Peng, and AS Raikhel. 2004. Target of rapamycin-mediated amino acid signaling in mosquito anautogeny. *Proc Natl Acad Sci U S A* 101 (29):10626-31.

Hansen, Immo A., Geoffrey M. Attardo, Saurabh G. Roy, and Alexander S. Raikhel. 2005. Target of rapamycin-dependent activation of S6 kinase is a central step in the transduction of nutritional signals during egg development in a mosquito. *Journal of Biological Chemistry* 280 (21):20565-20572.

Hay, Nissim, and Nahum Sonenberg. 2004. Upstream and downstream of mTOR. *Genes & Development* 18 (16):1926-1945.

Johnston, Laura A., David A. Prober, Bruce A. Edgar, Robert N. Eisenman, and Peter Gallant. 1999. *Drosophila* myc regulates cellular growth during development. *Cell* 98 (6):779-790.

Klowden, Marc J. 1987. Distention-mediated egg maturation in the mosquito, *Aedes aegypti*. *Journal of Insect Physiology* 33 (2):83-87.

LaFever, Leesa, Alexander Feoktistov, Hwei-Jan Hsu, and Daniela Drummond-Barbosa. 2010. Specific roles of Target of rapamycin in the control of stem cells and their progeny in the *Drosophila* ovary. *Development* 137 (13):2117-2126.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10 (3):R25.

Liao, Xin-Hua, Amit Majithia, Xiuli Huang, and Alan Kimmel. 2008. Growth control via TOR kinase signaling, an intracellular sensor of amino acid and energy availability, with crosstalk potential to proline metabolism. *Amino Acids* 35 (4):761-770.

Maines, Jean Z., Leslie M. Stevens, Xianglan Tong, and David Stein. 2004. *Drosophila* dMyc is required for ovary cell growth and endoreplication. *Development* 131 (4):775-786.

McMichael, A. J., D. H. Campbell-Lendrum, C. F. Corvalán, K. L. Ebi, A. Githeko, J. D. Scheraga, and A. Woodward. 2003. *Climate change and human health - risks and responses*.

Nezis, I. P., D. J. Stravopodis, L. H. Margaritis, and I. S. Papassideri. 2006. Autophagy is required for the degeneration of the ovarian follicular epithelium in higher Diptera. *Autophagy* 2 (4):297-8.

Orian, Amir, Bas van Steensel, Jeffrey Delrow, Harmen J. Bussemaker, Ling Li, Tomoyuki Sawado, Eleanor Williams, Lenora W.M. Loo, Shaun M. Cowley, Cynthia Yost, Sarah Pierce, Bruce A. Edgar, Susan M. Parkhurst, and Robert N. Eisenman. 2003. Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network. *Genes & Development* 17 (9):1101-1114.

Park, Jong-Hwa, Geoffrey M. Attardo, Immo A. Hansen, and Alexander S. Raikhel. 2006. GATA factor translation is the final downstream step in the amino acid/target-of-rapamycin-mediated vitellogenin gene expression in the anautogenous mosquito *Aedes aegypti*. *Journal of Biological Chemistry* 281 (16):11167-11176.

Pritchett, Tracy L., Elizabeth A. Tanner, and Kimberly McCall. 2009. Cracking open cell death in the *Drosophila* ovary. *Apoptosis* 14 (8):969-979.

Raikhel, Alexander S., Vladimir A. Kokoza, Jinsong Zhu, David Martin, Sheng-Fu Wang, Chao Li, Guoqiang Sun, Abdoulaziz Ahmed, Neal Dittmer, and Geoff Attardo. 2002. Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of antipathogen immunity. *Insect Biochemistry and Molecular Biology* 32:1275-1286.

Roy, Saurabh G., Immo A. Hansen, and Alexander S. Raikhel. 2007. Effect of insulin and 20-hydroxyecdysone in the fat body of the yellow fever mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* 37 (12):1317-1326.

Rutledge, L. C., R. A. Ward, and D. J. Gould. 1964. Studies on feeding response of mosquitoes to nutritive solutions in a new membrane feeder. *Mosquito News* 24:407-419.

Shiao, Shin-Hong, Immo A. Hansen, Jinsong Zhu, Douglas H. Sieglaff, and Alexander S. Raikhel. 2008. Juvenile hormone connects larval nutrition with target of rapamycin signaling in the mosquito *Aedes aegypti*. *Journal of Insect Physiology* 54 (1):231-239.

Teleman, Aurelio A., Ville Hietakangas, Aram C. Sayadian, and Stephen M. Cohen. 2008. Nutritional control of protein biosynthetic capacity by insulin via Myc in *Drosophila*. *Cell Metabolism* 7 (1):21-32.

Uchida, K., M. Nishizuka, D. Ohmori, T. Ueno, Y. Eshita, and A. Fukunaga. 2004. Follicular epithelial cell apoptosis of atretic follicles within developing ovaries of the mosquito *Culex pipiens pallens*. *J Insect Physiol* 50 (10):903-12.

Watson, John D., Sara K. Oster, Mary Shago, Fereshteh Khosravi, and Linda Z. Penn. 2002. Identifying genes regulated in a Myc-dependent manner. *Journal of Biological Chemistry* 277 (40):36921-36930.

Wilson, N., D. Lush, and M. G. Baker. 2010. Meteorological and climate change themes at the 2010 International Conference on Emerging Infectious Diseases. *Eurosurveillance* 15 (30).

Zerbino, DR, and E Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18 (5):821 - 829.

CHAPTER 3. *AEDENOVO*: A PIPELINE FOR *DE NOVO* TRANSCRIPT ASSEMBLY USING ORTHOLOGOUS GENES

Patrick Jennings and Lyric Bartholomay

Abstract

High throughput mRNA sequencing (RNA-seq) is a powerful tool for transcriptome analysis of non-model organisms, but accurate quantification of transcripts requires full-length reference transcript assemblies. The pipeline program presented here, *Aedenovo*, allows fast and accurate assembly of full-length transcripts from RNA-seq data. Previously annotated transcripts from a closely related organism are used as a template, and *de novo* assembled contigs from RNA-seq short reads are mapped to the template sequences and joined. This approach creates full-length *de novo* transcript assemblies that are orthologs of the known input sequences. The utility and significance of the *Aedenovo* assembly tool are discussed.

Introduction

High-throughput sequencing technologies (*e.g.* Illumina sequencing, 454 pyrosequencing, SOLiD sequencing) have revolutionized the field of genomics and transcriptomics. RNA-seq can be used to generate highly reproducible and accurate transcriptome data (Marioni et al. 2008) by purifying and fragmenting the extracted RNA, ligating adapters to the fragments, and sequencing a portion of every fragment simultaneously. Accurate quantification of transcript abundance ultimately depends on matching each read to a unique transcript. If no genome sequence is available, *de novo*

transcripts can be assembled from short reads, and then accurate mapping and quantification of reads is possible.

Bowtie, Tophat, and Cufflinks are widely-used programs for estimation of transcript abundance in RNA-seq experiments (Trapnell et al. 2010; Langmead et al. 2009). These three bioinformatics tools can be used to report a normalized relative expression level for each transcript, expressed in units of FPKM (expected fragments per kilobase of transcript per million fragments sequenced), but the programs require a reference genome or exome. In order to use these tools for normalized transcript expression analysis of non-model organisms, short reads should first be assembled into transcripts, and several programs (Velvet, Oases, Trans-ABySS) have been created for such *de novo* assembly (Zerbino and Birney 2008; Robertson et al. 2010). However, these programs cannot always assemble full-length transcript sequences. Unfortunately, biases exist in RNA-seq transcript coverage due to GC-rich regions (Dohm et al. 2008) and incomplete reverse transcription of RNA (Oshlack and Wakefield 2009); as a result of the non-uniform read coverage in RNA-seq, the accuracy of transcriptome analysis can be improved if a full-length transcript sequence is available for each gene of interest as the reference sequence against which transcript fragments can be compared.

Long contiguous sequences (contigs) can be assembled quickly from short reads without the need for a reference genome by using the program Velvet. Velvet uses De Bruijn graphs to assemble contigs quickly and with a low memory cost by connecting nodes of a fixed length, termed k-mers. Because of the nature of transcriptome data, a single k-mer value is not sufficient to assemble all transcripts. Transcripts that are expressed in high numbers are more accurately assembled with a high k-mer value, but certain rare transcripts

can only be assembled at lower k-mer values. The best *de novo* transcriptome assembly can be accomplished by combining unique contigs from multiple Velvet assemblies, each with a different k-mer value. This strategy has been called the additive multiple-k Velvet assembly method (Surget-Groba and Montoya-Burgos 2010). This assembly method generates high-quality contigs, but transcript sequences are often still highly fragmented.

If an orthologous sequence for the gene of interest is known, then the contigs from each Velvet assembly can be matched to it with the basic local alignment search tool (BLAST). BLAST is an ideal tool to use in this situation because of its adaptability when matching dissimilar nucleotide sequences from closely related species or translated nucleotide sequences from more distantly related species. In addition, BLAST meets the necessity for gapped alignment to account for insertions and deletions in different species, and it is a familiar tool to researchers in molecular biology and bioinformatics. BLAST parameters can be fine-tuned to the appropriate amount of similarity between query sequence and subject sequence. If the species are closely related, blastn (nucleotide-nucleotide matching) can be used, but in distantly related species, tblastx is a better option.

The pipeline described here, called *Aedenovo*, navigates the existing toolsets and associated shortfalls described above and assembles many highly accurate, full-length transcripts for organisms that lack a reference genome. Additionally, full functionality of standalone BLAST is preserved to assemble contigs of even distantly-related genes. The pipeline (Fig. 1) begins with removal of adapter sequences—artifacts of the sequencing process—from short reads and filtering of reads that would interfere with assembly. The additive multiple-k Velvet assembly method (Surget-Groba and Montoya-Burgos 2010) is then used to assemble contigs. Redundant contigs are identified and grouped, and the longest

contig in each group is kept. The remaining unique contigs are matched to known query sequences with BLAST based on user-defined settings. Next, a Perl script aligns the full length of the contigs against the query sequence. Overlapping contigs are connected if they share a specified amount of overlap (75 nt by default), and the process is repeated, using the newly assembled transcripts as the query sequences. With each iteration of the process, the 5′ and 3′ ends of each transcript are extended.

The pipeline program was originally developed to assemble a *de novo* transcriptome for analysis of three different stages of ovarian follicle development after blood feeding in the Eastern treehole mosquito, *Aedes* (*Ochlerotatus*) *triseriatus* (Say), and the closely related yellow fever mosquito, *Aedes aegypti*. Annotated genomic sequence is available for *Ae. aegypti* (Nene et al. 2007), which was used for validation of the *de novo* assembly methods. RNA was extracted from ovaries at each time point, and poly-A selected mRNA was sequenced on an Illumina GA_{II} platform to generate a complete ovary transcriptome for both species and identify genes that are transcriptionally upregulated or downregulated at each stage of interest. The program, *Aedenovo*, was developed to quickly assemble transcript sequences from *Ae. triseriatus* that are one-to-one orthologs of known *Ae. aegypti* transcripts and facilitate cross-species comparisons between differentially regulated transcripts. In one day, more than 20,000 transcripts were assembled for *Ae. triseriatus* using the *Aedenovo* pipeline program.

To validate the *Aedenovo* assembly method, RNA-seq short reads from *Ae. aegypti* were assembled against known transcripts from the Southern house mosquito, *Culex quinquefasciatus* (Arensburger et al. 2010), and compared to known transcript sequences for *Ae. aegypti*.

Materials and Methods

Short reads for *de novo* assembly were generated by Illumina RNA-seq. RNA was extracted and purified from *Ae. aegypti* ovaries and *Ae. triseriatus* ovaries during four unique developmental stages after a blood meal. Total RNA was extracted by a single-step acid guanidinium thiocyanate-phenol-chloroform extraction method (TRIzol®, Invitrogen) and sequenced on an Illumina GA_{II} platform using the mRNA-Seq 8-Sample Prep and TruSeq SBS v5-GA kits (Illumina, Inc., San Diego, CA, USA) at the Iowa State University DNA Facility. This protocol generated an average of 26,231,834 reads of up to 104 nt each in each of eight sequencing lanes.

Our approach was to use known *Ae. aegypti* transcripts as the query sequences to build orthologous *Ae. triseriatus* transcripts. For validation of the *Aedenovo* assembly method, *Ae. aegypti* transcripts were assembled with *Aedenovo* using a set of known transcripts from another mosquito. We identified candidate genes from *Cx. quinquefasciatus* that have orthologs in our *Ae. aegypti* Illumina short read dataset. The *Cx. quinquefasciatus* genome was sequenced in 2006 and recently annotated (Arensburger et al. 2010), so this closely related species was a good candidate to use for validation purposes. Bowtie v.0.12.7 (Langmead et al. 2009) was used to map short reads against the currently annotated *Ae. aegypti* transcripts (Geneset-AaegL1.2), and 4394 *Ae. aegypti* transcripts were selected with the criterion of having at least 500 Bowtie hits. One-to-one orthologs for the 4394 selected *Ae. aegypti* genes were identified in *Cx. quinquefasciatus* using BioMart (Kasprzyk et al. 2004) on http://www.vectorbase.org. These 4394 *Cx. quinquefasciatus* genes were used as query sequences in the pipeline program, and a *de novo* assembly was attempted from the *Ae. aegypti* Illumina mRNA-seq short reads.

The Perl program *Aedenovo* was used (Fig. 2) to automate the following steps in the assembly. Adapter sequence was trimmed from approximately six percent of the short reads, and reads that were trimmed by more than 40% of their total length were removed. Contigs were assembled by the Multiple-k Velvet Method (Surget-Groba and Montoya-Burgos 2010). Briefly, five separate Velvet assemblies were performed on the Illumina mRNA-seq output FASTQ file, using five different k-mer values (59, 55, 45, 39, and 21), with each assembly filtered to a coverage cutoff value of 8. The five k-mer values were selected based on an experiment to determine the optimum k-mer values for *Ae. triseriatus* transcriptome assembly (Fig. 3). The k-mer value was added to the header line of each contig, and the reads were concatenated into a single FASTA file. Redundant sequences were removed with CD-HIT-EST, with default parameters. A BLAST database (NCBI, formatdb version 2.2.24) was created from the non-redundant contig file, and a local blastn search was performed with E-value 0.1 and all other parameters set at default values. Alignments with bit scores below 99 were discarded, and the BLAST output table was sorted by query name (lexical sort) and bit score (descending numerical sort). Each line of the BLAST output table was read and used to create a temporary human-readable file showing alignments between Velvet contigs (subject sequence) and the original nucleotide query. If the subject start position was greater than the stop position, the reverse complement of the Velvet contig was printed. The contigs that aligned to each query sequence were connected by direct nucleotide matching of 75 nt near each end of the contig and searching the following contig for the saved sequence. If that 75 nt overlap was identified between contigs, they were joined together. The assembled genes were compiled into a single FASTA file, which was used as the query file for a second iteration of the *Aedenovo* pipeline program to extend the assemblies. The output FASTA file

name was changed to avoid overwriting, and in the second iteration of *Aedenovo*, the blastn E-value was adjusted to 0.01 to ensure accurate assembly (Fig. 2).

To analyze the accuracy of the transcript assembly, *de novo* assembled transcripts from *Ae. aegypti* were analyzed by blastx against the non-redundant protein database at NCBI, and the transcripts were also analyzed locally by blastn against the same *Ae. aegypti* transcripts that were used to find orthologous *Cx. quinquefasciatus* gene sequences.

After validation of the program, *Ae. triseriatus* transcriptome assembly was performed with *Aedenovo*, using for the query sequences every available mosquito transcript sequence from http://www.vectorbase.org (18,760 transcripts from *Ae. aegypti*, 18,883 from *Cx. quinquefasciatus*, and 14,324 from *Anopheles gambiae*). The first iteration of *Ae.* program used tblastx for mapping Velvet contigs to known transcript sequences, and the second iteration used blastn to match Velvet contigs to the *de novo* assembly of *Ae. triseriatus* transcripts.

Results

Transcriptome assembly results

The initial assembly of the *Ae. triseriatus* transcriptome used tblastx matching to map Velvet contigs against known mosquito transcripts from three different species, and the first iteration of *Aedenovo* generated 36,061 *Ae. triseriatus* transcripts. Removal of redundant contigs reduced the initial assembly to 20,160 transcripts. The second iteration of the program used blastn matching and extended the transcripts. Removal of redundant contigs reduced the second assembly to 17,743 transcripts that were longer than 104 base pairs. The

median length of *de novo* transcripts was 927 nt, with a maximum of 11,776 nt (ortholog of *Ae. aegypti* transformation/transcription domain-associated protein AAEL012951-RA).

Aedenovo validation results

After *de novo* assembly of *Ae. aegypti* genes using 4,394 known *Cx. quinquefasciatus* transcripts as a query, 3,277 *Ae. aegypti* transcripts were assembled, and 98.8% of the assembled transcripts matched back to the predicted one-to-one orthologous *Ae. aegypti* genes. The mean length of assembled contigs was 1.937 Kb, which is similar to the mean length of the predicted orthologs. Of the transcripts assembled by *Aedenovo*, 43% were longer than the currently annotated transcript, and nearly all of the longer transcripts represent an extension of the known 5′ and 3′ untranslated regions.

Discussion

Aedenovo is a versatile program that can be used for *de novo* assembly of orthologous transcripts from any organism, regardless of the availability of a reference genome or transcript data. One of the main advantages of the program is its ease of use. The programs that are driven by *Aedenovo* (Velvet, CD-HIT-EST, and BLAST) are open source, have excellent user manuals, and are easy to compile on any UNIX system. Because *Aedenovo* assembles orthologs of known genes, putative function of many genes can be inferred from the annotated FASTA header line of the query file.

During the validation of *Aedenovo*, it became obvious that this program is very useful for annotating existing transcriptomes. This program was designed to be a fast, easy to use solution for *de novo* transcriptome assembly from an organism without a reference genome

or transcriptome, but it has equal utility in annotation of existing transcriptomes or confirmation of predicted gene sequences by extending untranslated regions, identifying exons that are missed by gene prediction software, or searching for new genes that are known to be expressed in a closely related species. The source code for *Aedenovo* (Appendix A) is available as an open-source Perl script at http://aedenovo.sourceforge.net/.

Acknowledgments

Perl programming training was provided by an excellent teacher, Hui-Hsien Chou, and suggestions for performance optimization were kindly provided by Daniel Standage. We would also like to thank Sijun Liu for his helpful discussions about transcriptome assembly. Transcriptome sequencing was funded by a grant from the Iowa State University Center for Integrated Animal Genomics.

Figures

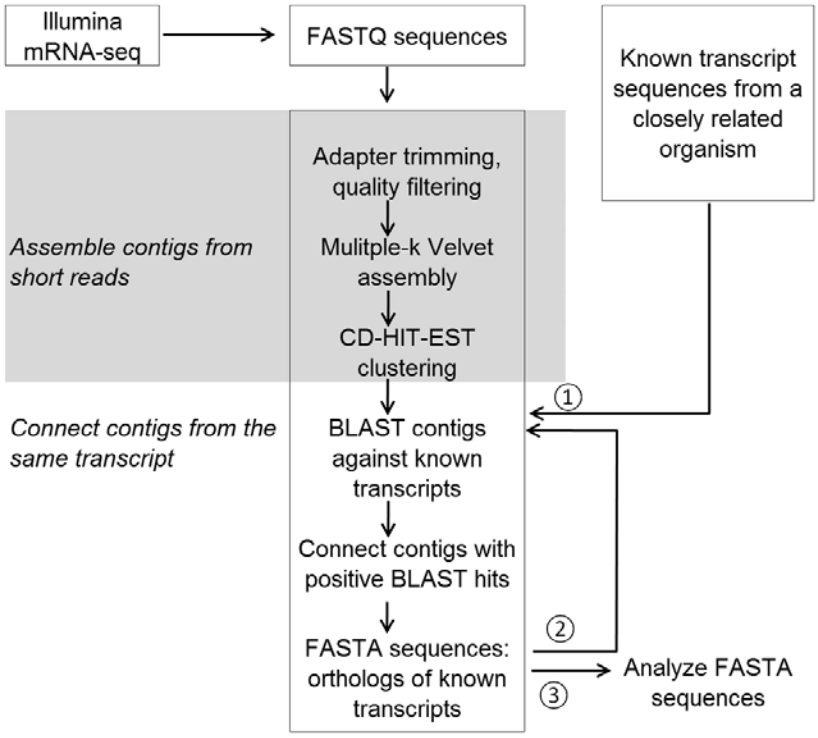
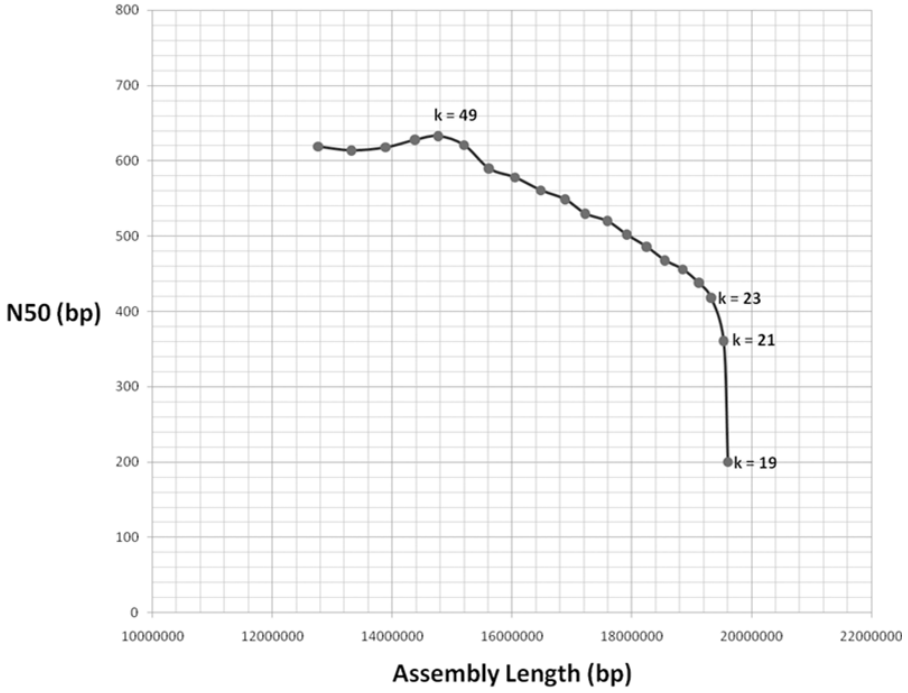


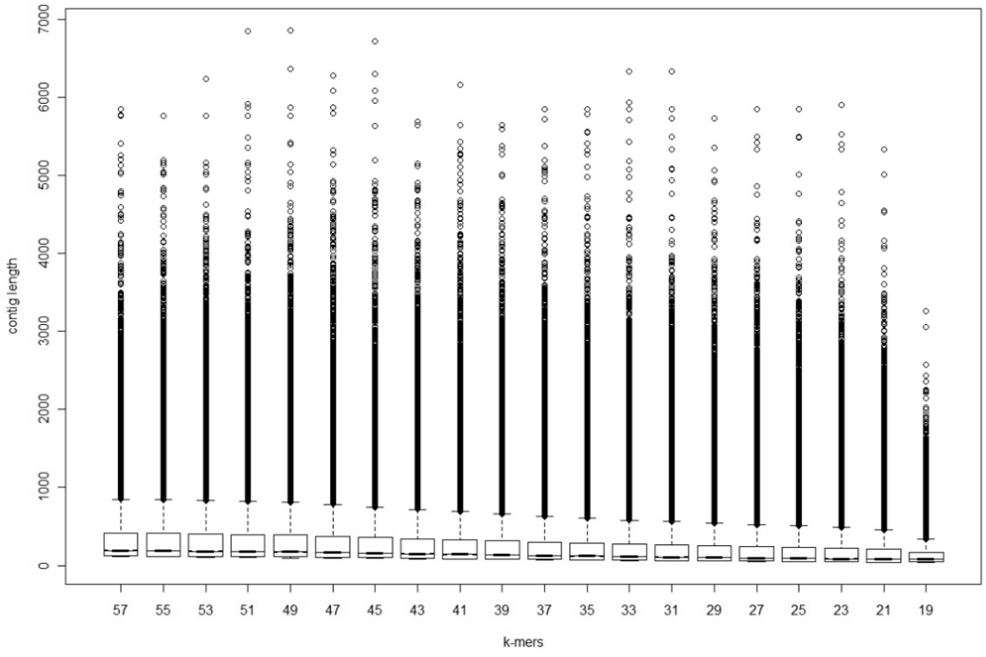
Figure 1. Overview of the *Aedenovo* transcriptome assembly process based on orthologous transcript sequences. On the first iteration of the program, known transcript sequences from a closely related organism (1) are used as tblastx or blastn query templates for identifying contigs that might belong to the same transcript, with Velvet contigs as the BLAST database. On the second iteration of the program, the *de novo* assembled FASTA sequences from the first iteration (2) are used as the BLAST query, and additional Velvet contigs in the database can be incorporated to extend the transcripts. This process can be repeated (2), or the reads can be analyzed (3) by traditional mRNA bioinformatics tools.

```
$  Aedenovo.pl queryfile=cquinxgenesunique.fa shortreads=s_5_sequence.txt
   trim=yes kmers=59,55,45,39,21 blastall=-p#blastn#-a#6#-e#0.1 cutoff=99

$  mv genebuild.txt genebuild1.txt

$  Aedenovo.pl queryfile=genebuild1.txt subject=contigs_5_All.fa
   blastdb=contigs_5_All.fa blastall=-p#blastn#-a#8#-e#0.01
```

Figure 2. UNIX command-line arguments used for validation of *Aedenovo*.



a.



b.

Figure 3. Assembly characteristics of various k-mer values (k = 57 to k = 19) on Velvet assembly of *Aedes triseriatus* ovary mRNA-seq 104-nt short reads. (a) N50 values are highest at k = 49 and decrease precipitously at k < 21. As k decreases, the total assembly length increases. (b) Boxplots of contig nucleotide lengths show the median length of contigs decreases with decreasing k, but lower k-mer values improve assembly of low-copy-number transcripts.

References

Arensburger, Peter, Karine Megy, Robert M. Waterhouse, Jenica Abrudan, Paolo Amedeo, Beatriz Antelo, Lyric Bartholomay, Shelby Bidwell, Elisabet Caler, Francisco Camara, Corey L. Campbell, Kathryn S. Campbell, Claudio Casola, Marta T. Castro, Ishwar Chandramouliswaran, Sinead B. Chapman, Scott Christley, Javier Costas, Eric Eisenstadt, Cedric Feschotte, Claire Fraser-Liggett, Roderic Guigo, Brian Haas, Martin Hammond, Bill S. Hansson, Janet Hemingway, Sharon R. Hill, Clint Howarth, Rickard Ignell, Ryan C. Kennedy, Chinnappa D. Kodira, Neil F. Lobo, Chunhong Mao, George Mayhew, Kristin Michel, Akio Mori, Nannan Liu, Horacio Naveira, Vishvanath Nene, Nam Nguyen, Matthew D. Pearson, Ellen J. Pritham, Daniela Puiu, Yumin Qi, Hilary Ranson, Jose M. C. Ribeiro, Hugh M. Roberston, David W. Severson, Martin Shumway, Mario Stanke, Robert L. Strausberg, Cheng Sun, Granger Sutton, Zhijian Tu, Jose Manuel C. Tubio, Maria F. Unger, Dana L. Vanlandingham, Albert J. Vilella, Owen White, Jared R. White, Charles S. Wondji, Jennifer Wortman, Evgeny M. Zdobnov, Bruce Birren, Bruce M. Christensen, Frank H. Collins, Anthony Cornel, George Dimopoulos, Linda I. Hannick, Stephen Higgs, Gregory C. Lanzaro, Daniel Lawson, Norman H. Lee, Marc A. T. Muskavitch, Alexander S. Raikhel, and Peter W. Atkinson. 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* 330 (6000):86-88.

Dohm, Juliane C., Claudio Lottaz, Tatiana Borodina, and Heinz Himmelbauer. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* 36 (16):e105.

Kasprzyk, Arek, Damian Keefe, Damian Smedley, Darin London, William Spooner, Craig Melsopp, Martin Hammond, Philippe Rocca-Serra, Tony Cox, and Ewan Birney. 2004. EnsMart: A generic system for fast and flexible access to biological data. *Genome Research* 14 (1):160-169.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10 (3):R25.

Marioni, John C., Christopher E. Mason, Shrikant M. Mane, Matthew Stephens, and Yoav Gilad. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* 18 (9):1509-1517.

Nene, Vishvanath, Jennifer R. Wortman, Daniel Lawson, Brian Haas, Chinnappa Kodira, Zhijian Tu, Brendan Loftus, Zhiyong Xi, Karyn Megy, Manfred Grabherr, Quinghu Ren, Evgeny M. Zdobnov, Neil F. Lobo, Kathryn S. Campbell, Susan E. Brown, Maria F. Bonaldo, Jingsong Zhu, Steven P. Sinkins, David G. Hogenkamp, Paolo Amedeo, Peter Arensburger, Peter W. Atkinson, Shelby Bidwell, Jim Biedler, Ewan Birney, Robert V. Bruggner, Javier Costas, Monique R. Coy, Jonathan Crabtree, Matt Crawford, Becky deBruyn, David DeCaprio, Karin Eiglmeier, Eric Eisenstadt, Hamza

El-Dorry, William M. Gelbart, Suely L. Gomes, Martin Hammond, Linda I. Hannick, James R. Hogan, Michael H. Holmes, David Jaffe, J. Spencer Johnston, Ryan C. Kennedy, Hean Koo, Saul Kravitz, Evgenia V. Kriventseva, David Kulp, Kurt LaButti, Eduardo Lee, Song Li, Diane D. Lovin, Chunhong Mao, Evan Mauceli, Carlos F. M. Menck, Jason R. Miller, Philip Montgomery, Akio Mori, Ana L. Nascimento, Horacio F. Naveira, Chad Nusbaum, Sinéad O'Leary, Joshua Orvis, Mihaela Pertea, Hadi Quesneville, Kyanne R. Reidenbach, Yu-Hui Rogers, Charles W. Roth, Jennifer R. Schneider, Michael Schatz, Martin Shumway, Mario Stanke, Eric O. Stinson, Jose M. C. Tubio, Janice P. VanZee, Sergio Verjovski-Almeida, Doreen Werner, Owen White, Stefan Wyder, Qiandong Zeng, Qi Zhao, Yongmei Zhao, Catherine A. Hill, Alexander S. Raikhel, Marcelo B. Soares, Dennis L. Knudson, Norman H. Lee, James Galagan, Steven L. Salzberg, Ian T. Paulsen, George Dimopoulos, Frank H. Collins, Bruce Birren, Claire M. Fraser-Liggett, and David W. Severson. 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316 (5832):1718-1723.

Oshlack, Alicia, and Matthew Wakefield. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct* 4 (1):14.

Robertson, Gordon, Jacqueline Schein, Readman Chiu, Richard Corbett, Matthew Field, Shaun D. Jackman, Karen Mungall, Sam Lee, Hisanaga Mark Okada, Jenny Q. Qian, Malachi Griffith, Anthony Raymond, Nina Thiessen, Timothee Cezard, Yaron S. Butterfield, Richard Newsome, Simon K. Chan, Rong She, Richard Varhol, Baljit Kamoh, Anna-Liisa Prabhu, Angela Tam, YongJun Zhao, Richard A. Moore, Martin Hirst, Marco A. Marra, Steven J. M. Jones, Pamela A. Hoodless, and Inanc Birol. 2010. De novo assembly and analysis of RNA-seq data. *Nat Meth* 7 (11):909-912.

Surget-Groba, Y, and JI Montoya-Burgos. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* 20 (10):1432 - 1440.

Trapnell, Cole, Brian A. Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J. van Baren, Steven L. Salzberg, Barbara J. Wold, and Lior Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotech* 28 (5):511-515.

Zerbino, DR, and E Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18 (5):821 - 829.

CHAPTER 4. TRANSCRIPTOME SEQUENCING AND ANALYSIS OF MOSQUITO OVARIAN DEVELOPMENT USING ILLUMINA RNA-SEQ TECHNOLOGY

Patrick Jennings and Lyric Bartholomay

Abstract

The eastern treehole mosquito, *Aedes triseriatus* (Say), is the primary vector of La Crosse virus (LACV) in the United States. LACV is maintained in natural populations by transovarial transmission from infected females to their progeny. Our understanding of the innate immune response in mosquito ovaries is a fundamental step toward understanding the transmission cycle of LACV and other vertically-transmitted arthropod-borne viruses. We examined the transcriptomes of *Ae. triseriatus* and *Ae. aegypti* ovaries by Illumina RNA-seq, and we identified genes that are differentially regulated during the processes of follicular atresia, including apoptosis- and autophagy-related genes. We assembled 17,743 novel transcript sequences from the medically significant species, *Ae. triseriatus*, and compared transcriptional profiles of egg development in *Ae. triseriatus* and *Ae. aegypti*.

Introduction

La Crosse virus (LACV; family *Bunyaviridae*, genus *Orthobunyavirus*) is one of the leading causes of pediatric viral encephalitis in the United States. The Eastern treehole mosquito, *Aedes* (*Ochlerotatus*) *triseriatus* (Say), is the primary vector of LACV, and *Ae. triseriatus* can transmit LACV horizontally between vertebrates and vertically (via transovarial transmission) to its offspring. Transovarial transmission of LACV is critical to its persistence in the northern United States, but the molecular mechanisms of LACV

infection and regulation of cell death in the ovaries of *Ae. triseriatus* are unknown. The yellow fever mosquito, *Aedes aegypti*, is an important vector of a wide variety of arthropod-borne viruses (arboviruses), most notably dengue virus, chikungunya virus, and yellow fever virus (Gubler 1998; Weaver and Reisen 2010). The *Ae. aegypti* genome has been sequenced and annotated (Nene et al. 2007), and provides an excellent reference organism for transcriptomic analysis of mosquito oogenesis.

Hematophagous mosquitoes have polytrophic meroistic ovaries composed of many individual ovarioles, each of which can produce one viable egg after each blood meal. The anterior part of each ovariole contains follicle stem cells and germline cells, and the rest of the ovariole contains differentiated follicles, each comprising an oocyte and seven nurse cells surrounded by a follicular epithelium (FE). The FE is made up of many syncytia of cells with intracellular bridges formed by incomplete cytokinesis (Fiil 1978), and these cells are important for regulating growth and development of the oocyte. In the vitellogenic stage of oogenesis (after a blood meal is taken), damaged follicles may degrade in a process called follicular atresia, which involves both apoptosis and autophagy (Uchida et al. 2004; Malagoli et al. 2010). If the mosquito did not acquire sufficient protein for the development of every oocyte, follicular atresia occurs in a higher proportion of ovarian follicles and reallocates protein from atretic follicles to the other developing follicles (Volozina 1967; Uchida et al. 2004).

During normal mosquito oogenesis, the nurse cells undergo endoreplication and synthesize ribosomal RNA, which is transported into the oocyte by intercellular canals (Cave 1982). Approximately two days after the blood meal, nurse cells degrade by apoptosis and release their contents into the developing oocyte. A third cell death event occurs just before

oviposition. The FE cells secrete an extracellular matrix (chorion) that surrounds the oocyte, and then the FE degrades by autophagy and apoptosis (Nezis et al. 2006), releasing the primary follicle for oviposition. Surprisingly little is known about genes that regulate cell death in the mosquito ovary during oogenesis, and characterizing the processes of apoptosis and autophagy in the mosquito ovary are important to our understanding of virus transmission, especially in the case of transovarially transmitted arboviruses.

Autophagy and apoptosis are both used as forms of innate immunity to viral infection (Lee *et al*. 2007; Talloczy *et al*. 2002), but viruses that are transmitted transovarially do not typically reduce fecundity in their mosquito hosts (Lambrechts and Scott 2009). Therefore, vertically transmitted arboviruses must have evolved a mechanism to avoid or alter the host's innate immune responses (Keene et al. 2004; Reese et al. 2010), and alteration of cell death pathways are likely to be involved in persistent arboviral infections (Shelly et al. 2009; Girard et al. 2007; Wang et al. 2008).

We used next-generation sequencing technology to elucidate the genetic mechanisms of cell death in *Ae*. *aegypti* and *Ae*. *triseriatus*. Illumina RNA-seq technology is a powerful tool for transcriptome analysis of organisms with or without a reference genome. Current RNA-seq methods can generate tens of millions of reads, each more than 100 base pairs long, from a single sequencing reaction. These reads can be assembled into full-length transcripts or compared to known transcript sequences to generate highly replicable, statistically meaningful transcriptome data from a single flow cell (Marioni et al. 2008). We used Illumina RNA-seq to sequence the *Ae. triseriatus* ovary transcriptome and identify genes that are differentially regulated during mosquito oogenesis, with a focus on the molecular mechanisms that regulate autophagy and apoptosis.

Materials and Methods

Mosquito maintenance and blood feeding.

Ae. triseriatus Ames strain (Erickson et al. 2006) larvae and *Ae. aegypti* Liverpool strain larvae were hatched from dry oviposition papers in deoxygenated, deionized water, and then transferred to enamel rearing pans filled with deionized water at a density of 150 larvae per pan (*Ae*. *triseriatus*) or 300 per pan (*Ae*. *aegypti*). Larvae were fed TetraMin® Tropical Flakes daily. Female pupae were separated into half-liter cartons. Adults were maintained at 27°C and 70% relative humidity and were fed 10% sucrose from a cotton pad for 3 days. Three- to five-day-old mosquitoes were fed 35% packed-cell-volume defibrinated sheep blood (Hemostat Laboratories, Dixon, CA, USA). One group of *Ae. triseriatus* and one group of *Ae. aegypti* were fed blood diluted with an equal volume of 0.9% saline in order to induce more follicular atresia (Uchida et al. 2004).

Ovary dissections.

Mosquito ovaries were dissected to analyze three developmental processes: follicular atresia, nurse cell death, and follicular epithelial cell death. Mosquito ovaries were removed and stained with neutral red for 10 seconds to determine the correct stage of development (Clements and Boocock 1984). The first collections were made from diluted- and undiluted blood-fed mosquitoes when the oocyte occupied 50% of the follicle (Christophers' stage IIIb) and follicles showed signs of atresia. In the undiluted blood-fed group, approximately 10% of follicles in both species were undergoing follicular atresia, and in the diluted blood-fed group, approximately 20% of follicles were atretic at stage IIIb. The next collections were started when nurse cells occupied less than 10% of the total length of the follicle and the

nurse cells began to stain dark red (stages IVa-IVb). The final collections were made when nurse cells were completely degraded and follicular epithelial cells were beginning to stain dark red (stage V). The number of mosquitoes dissected (*n*) and the time points in hours post blood meal (hpbm) are presented in Table 1.

Ovaries for RNA-seq analysis were dissected into a drop of sterile 0.9% saline with special care to avoid collecting fat body or malpighian tubule tissue. Each ovary was immediately rinsed once in sterile saline and transferred to an RNA stabilization solution (RNA*later*®, Ambion) in an RNase-free microcentrifuge tube at room temperature. Ovaries were stored at 4°C overnight in RNA*later* before RNA extractions.

RNA extractions.

RNA stabilization solution was removed from the collection tubes, and total RNA was extracted by a single-step acid guanidinium thiocyanate-phenol-chloroform extraction method (TRIzol®, Invitrogen). RNA integrity and purity was confirmed by spectrophotometry (NanoDrop ND-2000c, Thermo Scientific), agarose gel electrophoresis, and Agilent 2100 bioanalyzer RNA nanochip analysis. The 26S rRNA of *Ae*. *aegypti*, *Ae*. *triseriatus* and many other insects contains a 'hidden break,' causing the rRNA to split into two equal-sized subunits upon heating. The resulting products co-migrate with the 18S rRNA during electrophoresis (Shine and Dalgarno 1973). Because of the hidden break, Agilent 2100 bioanalyzer RNA integrity numbers and concentration values were unreliable, and spectrophotometry was used to quantify RNA.

RNA-seq.

RNA was diluted in nuclease-free water to a concentration of 1.0 μg/μl, and 10 μl were used for Illumina mRNA-seq analysis. Briefly, poly-A mRNA was isolated using Illumina's mRNA-seq prep kit, mRNA was fragmented by addition of divalent cations under elevated temperature, double-stranded cDNA (dscDNA) was synthesized (SuperScript II/DNA polymerase I), blunt ends were created with T4 DNA polymerase and Klenow DNA polymerase, 3′ ends were adenylated, and Illumina PE PCR adapters (5′-AATGATACGGCG ACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′ and 5′-CAA GCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCG ATCT-3′) were ligated to the fragmented dscDNA. The cDNA templates were size selected by agarose gel electrophoresis and amplified with 15 cycles of PCR. Sequencing was performed on an Illumina GA_{II} System. A total of 108,230,642 reads and 101,624,029 reads were sequenced from *Ae. triseriatus* and *Ae. aegypti*, respectively.

Bioinformatics.

Adapter sequences were removed from all FASTQ formatted reads using a custom perl script, which was incorporated into the pipeline program, *Aedenovo*. Contigs were assembled from *Ae*. *aegypti* and *Ae*. *triseriatus* reads with Velvet v.1.0.13 (http://www.ebi.ac.uk /~zerbino/velvet/), and genome-referenced assembly was performed for *Ae*. *aegypti* against the *Ae*. *aegypti* genome build AaegL1.2 (http://aaegypti.vectorbase.org). Bowtie v.0.12.7, Tophat v.1.2.0, and Cufflinks v.0.9.3 (http://www.cbcb.umd.edu/) were used to map reads to the *Ae. aegypti* genome and estimate transcript abundance by splice-junction mapping. The abundance of each transcript

expressed in FPKM (expected fragments per kilobase of transcript per million fragments sequenced) is normalized based on the amount of starting reads and the length of each transcript, allowing comparisons between transcripts of different lengths and across multiple sequencing lanes. Cuffdiff v.0.9.3 was used to identify statistically significant changes in transcript abundance in *Ae. aegypti*.

To optimize *de novo* assembly for *Ae. triseriatus* with Velvet, hashes were built with stepwise *k*-mer values (appendix B) and reads were assembled against each hash with stepwise coverage cutoff values. For *Ae. triseriatus* transcriptome assembly with Velvet, k-mer values of 59, 55, 45, 39, and 21 and coverage cutoff of 8 were used. The five assemblies were combined and redundant sequences were removed with CD-HIT-EST v.4.3. Contigs were mapped to known mosquito transcript sequences with a custom Perl program called *Aedenovo* (see Chapter 3).

Bowtie v.0.12.7 was used to map *Ae. triseriatus* reads to the generated transcripts, and FPKM normalization was done manually by dividing the number of hits per transcript by the read length (in kilobases) and by the number of reads in each assembly (in millions).

Functional classes of genes (Bartholomay et al. 2010; Nene et al. 2007) were matched to *Ae. aegypti* transcripts using a Microsoft Access query. Functional classes for the *Ae. triseriatus* transcripts were predicted by matching the functional group of the best *Ae. aegypti* blastn hit to *Ae. triseriatus* transcripts. One problem that arose when analyzing the complete transcriptome was that a high proportion (78.4%) of transcripts did not show any significant change in transcript abundance (Figs. 1d and 2d), and these transcripts obscured the effects of upregulated or downregulated transcripts on functional group analysis. To that end, transcripts with constant expression in each of the four experimental

groups were excluded from analysis of functional groups. For each transcript, if the standard deviation of the relative proportions of FPKM values was less than 0.1 (Equation 1), the transcript was considered to have constant expression, and was excluded from whole-transcriptome analysis of functional groups (Figs. 3, 4, and 5).

$$\text{exclude if } 0.1 > \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{FPKM_j}{\sum_{j=1}^{N}(FPKM_j)}\right)_i - \mu\right]^2}, \text{where } \mu = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{FPKM_j}{\sum_{j=1}^{N} FPKM_j}\right)_i \quad (1)$$

A more complete transcriptomic analysis was performed that did not exclude any high-copy-number transcripts. Transcripts that shared similar expression patterns during oogenesis were grouped together, and the numbers of transcripts represented in each functional group were counted, per expression pattern group (Fig. 1 and 2). For these pattern-based analyses, only low-copy-number transcripts (represented by less than 2 FPKM across the three developmental stages) were filtered out before analysis. Each remaining transcript was classified into discrete groups based on the relative percent change of each transcript across the three developmental stages, and these groups can be represented in a Venn diagram fashion as depicted in Figures 1a – 1g and 2a – 2g.

Results

Illumina sequencing generated a total of 21.4 billion nucleotides of mRNA sequence in 209.8 million short reads from eight sequencing lanes (Table 1). Adapter trimming and subsequent removal of reads that were trimmed to < 60 nt reduced the total number of reads to 205 million (97.7%).

For *Ae. aegypti*, the program Cuffdiff identified significant changes in the abundance of 893 transcripts between the three developmental stages of oogenesis (follicular atresia, nurse cell death, and follicular epithelium death), and we also identified significant changes in the abundance of 93 transcripts between undiluted- and diluted blood-fed mosquitoes at 22-25 hpbm (Fig. 3). The filtering procedure (Equation 1) identified 2362 transcripts in *Ae. aegypti* and 1992 transcripts in *Ae. triseriatus* that showed differential transcription during oogenesis, Of those differentially regulated transcripts, the total transcript abundance per functional group changes dramatically across the developmental stages of oogenesis in *Ae. triseriatus* (Fig. 4) and *Ae. aegypti* (Fig. 5). Interestingly, the total number of differentially regulated transcripts expressed in *Ae. aegypti* increased over time, whereas an opposite effect was observed in *Ae. triseriatus*. However, orthologs were not identified for every *Ae. aegypti* transcript, and the effects of unidentified transcripts in *Ae. triseriatus* may contribute to the differences observed.

Discussion

In *Ae. aegypti*, the 93 transcripts that show expression changes between undiluted- and diluted blood-fed mosquitoes are likely to be involved in transduction of nutritional signals and induction of follicular atresia. The 50% diluted blood meal caused the number of atretic follicles to approximately double (~10% of follicles were atretic in the undiluted blood-fed group and ~20% of follicles were atretic in the diluted blood-fed group). A large proportion of the upregulated transcripts during increased follicular atresia function in translation, indicating that a select subgroup of ribosomal proteins may be involved in follicular atresia. Ribosomal proteins S2, L44/L41, L36, S25, and mitochondrial ribosomal

protein L42 showed increased transcript abundance (+11%, 12%, 13%, 22%, and 46%, respectively) in the mosquito ovaries that were undergoing more follicular atresia, but the transcript abundance of ribosomal protein S5 showed a decrease in transcript abundance (– 11%). Ribosomal protein S5 is necessary for small ribosomal subunit assembly (Roy-Chaudhuri, Kirthi, and Culver 2010), and its upregulation is anticipated during periods of increased protein synthesis when nutrient levels are high. Ribosomal protein S2 is essential for oogenesis in *Drosophila melanogaster* (Cramton and Laski 1994) and plays a role in regulating reproductive diapause in the mosquito *Culex pipiens* (Kim and Denlinger 2010), so it may play an important role in the translation of proteins related to nutritional signaling in mosquitoes.

Ae. aegypti transcriptome analysis revealed interesting insights into cell death events that occur during early vitellogenic phase oogenesis (follicular atresia), early post-vitellogenic phase oogenesis (nurse cell death), and late post-vitellogenic phase oogenesis (follicular epithelium death). For example, the caspase activator ARK (Apaf-1 related killer) is downregulated during nurse cell death, but upregulated during follicular atresia and follicular epithelium death. ARK regulates stress-induced apoptosis in *Drosophila* (Zimmermann et al. 2002) Many of these differentially regulated genes have diverse or unknown function, and the expression changes observed here will provide valuable insight into the function of these genes in oogenesis and cell death. A small representative group of highly differentially regulated genes with normalized expression values are presented in Table 2. Many mosquito gene products with previously unknown function displayed an extraordinary degree of transcriptional regulation in this transcriptome, which demonstrates

the utility of RNA-seq to identify targets for advanced analysis of cellular signaling pathways.

Acknowledgments

Transcriptome sequencing was funded by a grant from the Iowa State University Center for Integrated Animal Genomics. Many thanks to Dr. John VanDyk for technical support and software installation.

Figures/Tables

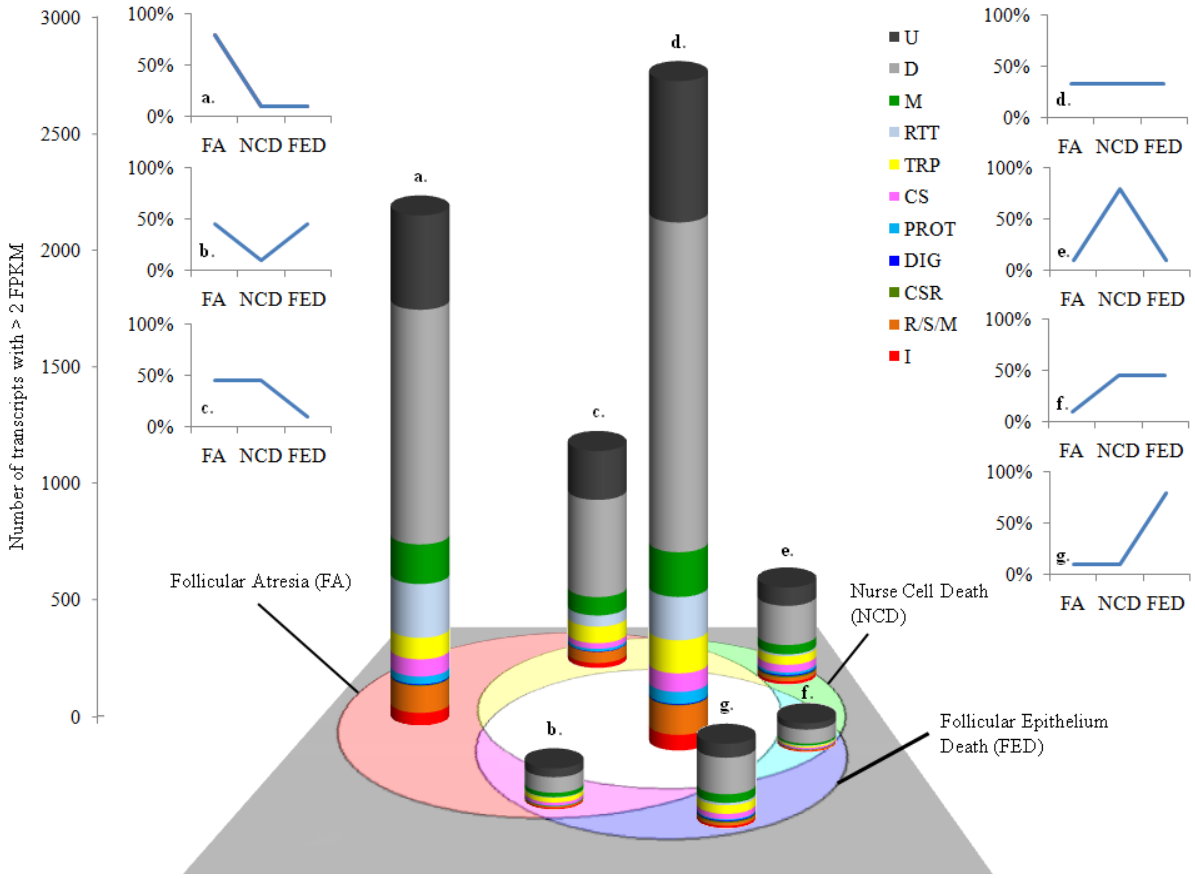


Figure 1. Number of discrete transcripts expressed in the *Aedes triseriatus* ovary during three stages of development. Transcripts that were found at numbers below 2 FPKM were excluded from this analysis. The area-proportional Venn diagram shows the total proportion of transcripts present at numbers above 2 FPKM, and the height of the bars represent the same information, but divided by functional group. Letters (a – g) correspond with the typical expression pattern (percent of the total abundance by transcript) represented in the Venn diagram structure. Abbreviations: U, unknown functions; D, diverse functions; M, metabolism; RTT, replication, transcription, and translation; TRP, transport; CS, cytoskeletal and structural; PROT, proteolysis; DIG, blood digestive; CSR, chemosensory reception; R/S/M, redox, stress, and mitochondrial; I, immunity; FA, follicular atresia; NCD, nurse cell death; FED, follicular epithelium death.

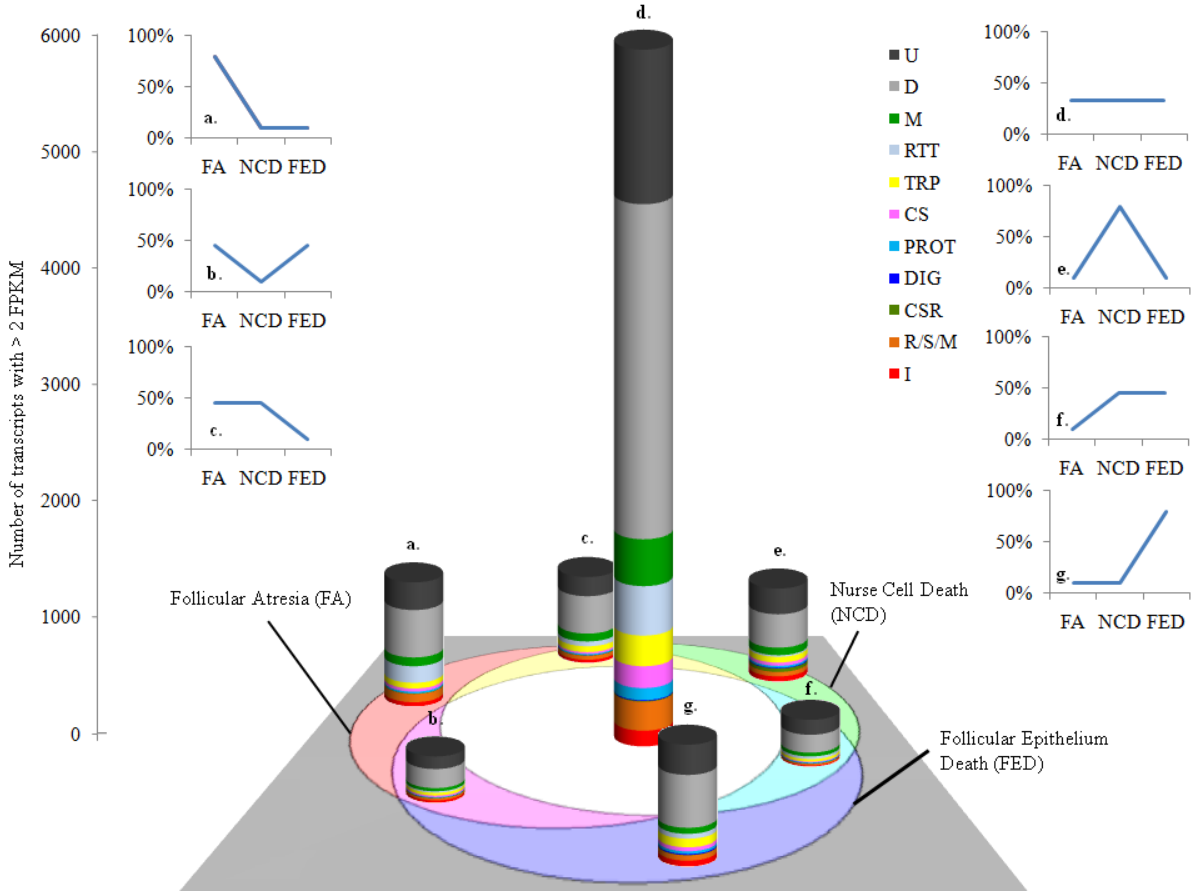


Figure 2. Number of discrete transcripts expressed in the *Aedes aegypti* ovary during three stages of development. Transcripts that were found at numbers below 2 FPKM were excluded from this analysis. The area-proportional Venn diagram shows the total proportion of transcripts present at numbers above 2 FPKM, and the height of the bars represent the same information, but divided by functional group. Letters (a – g) correspond with the typical expression pattern (percent of the total abundance by transcript) represented in the Venn diagram structure. Abbreviations: U, unknown functions; D, diverse functions; M, metabolism; RTT, replication, transcription, and translation; TRP, transport; CS, cytoskeletal and structural; PROT, proteolysis; DIG, blood digestive; CSR, chemosensory reception; R/S/M, redox, stress, and mitochondrial; I, immunity; FA, follicular atresia; NCD, nurse cell death; FED, follicular epithelium death.

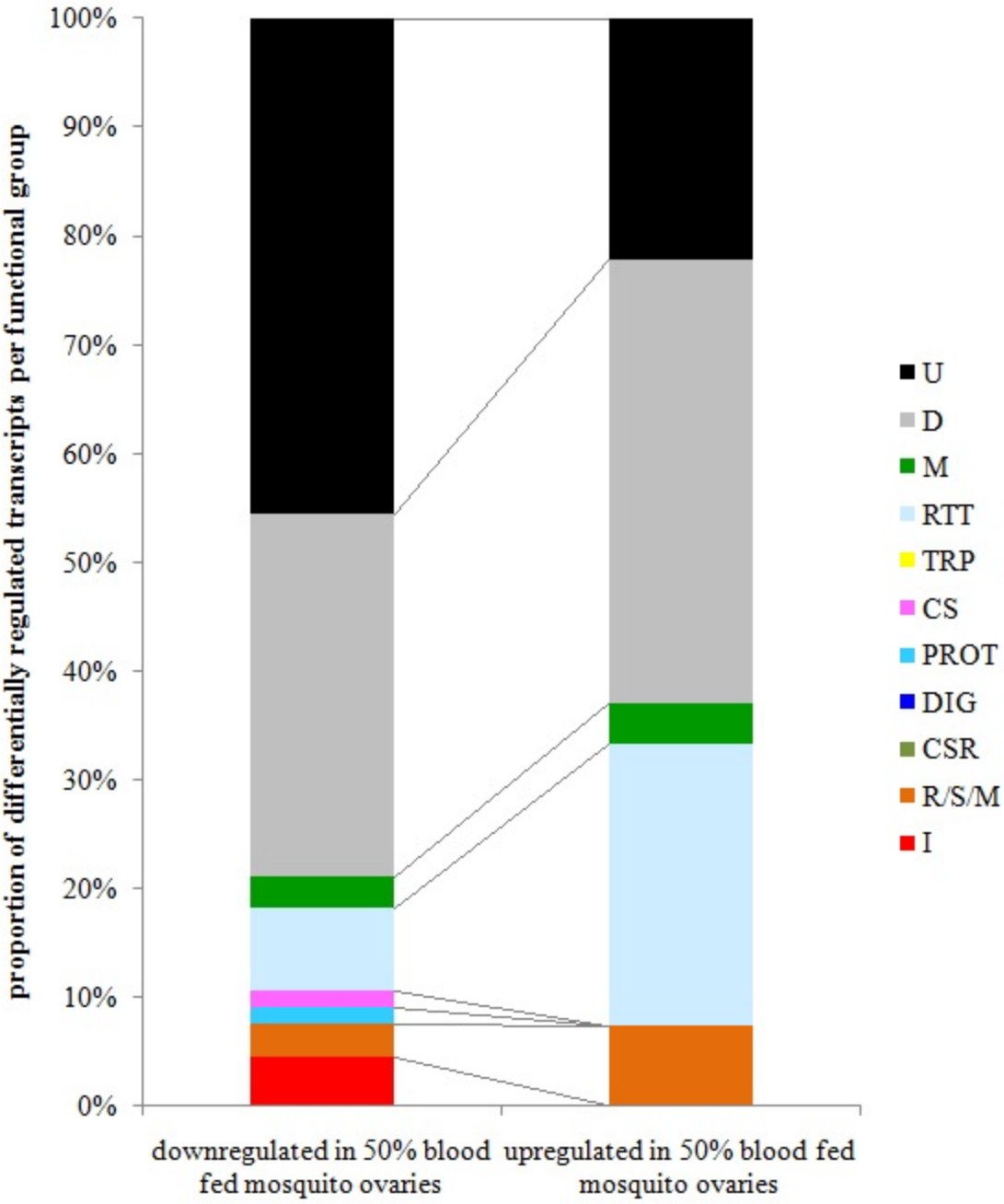


Figure 3. Proportion of *Aedes aegypti* functional groups of genes that are differentially regulated when mosquitoes are fed a 50% diluted blood meal and experience increased follicular atresia. Cuffdiff identified 66 transcripts that were downregulated and 27 transcripts that were upregulated. Abbreviations: U, unknown functions; D, diverse functions; M, metabolism; RTT, replication, transcription, and translation; TRP, transport; CS, cytoskeletal and structural; PROT, proteolysis; DIG, blood digestive; CSR, chemosensory reception; R/S/M, redox, stress, and mitochondrial; I, immunity.

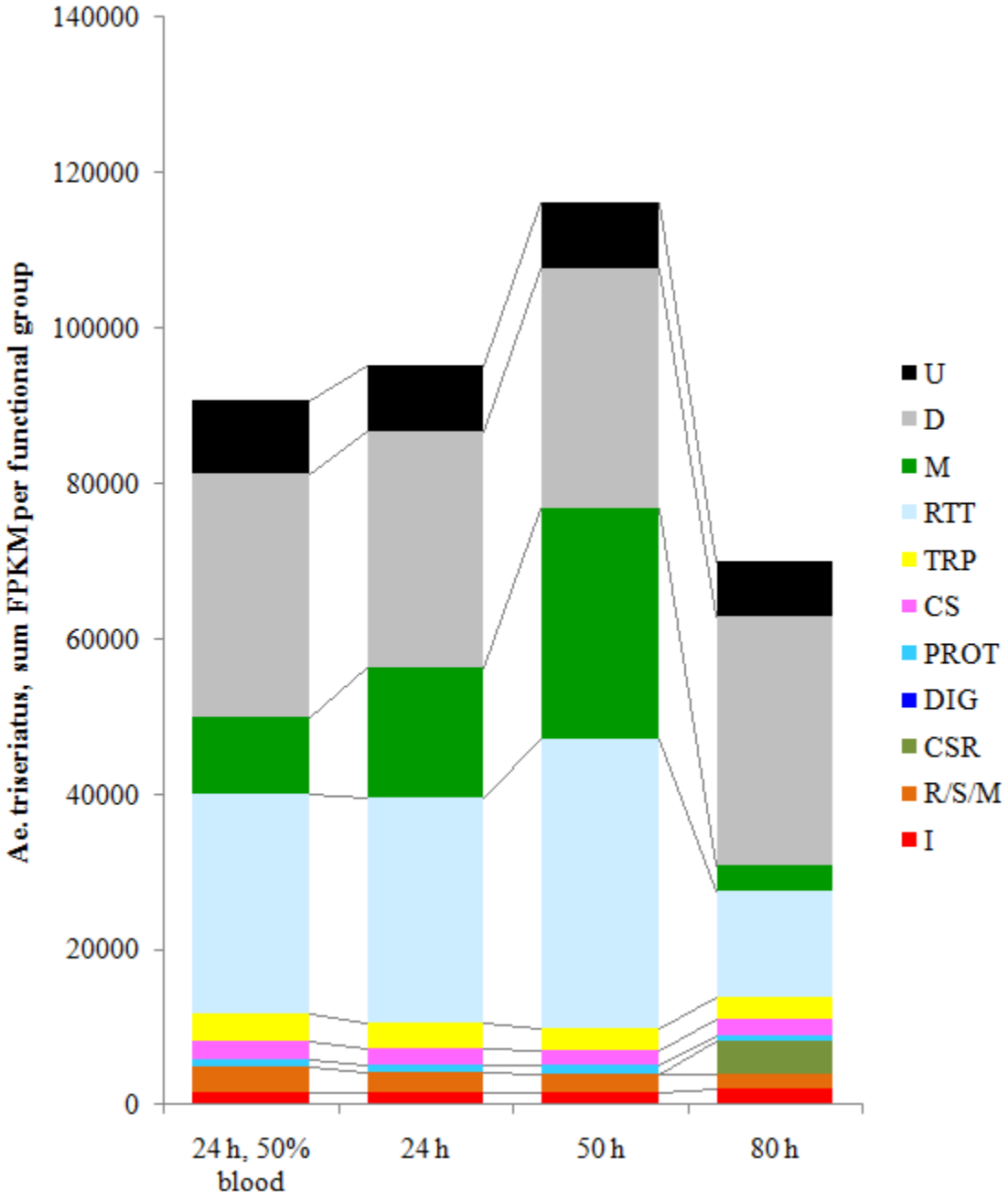


Figure 4. Sum normalized abundance of differentially regulated *Aedes triseriatus* transcripts per functional group during oogenesis. Units are in FPKM (see text), and transcripts that showed relatively small changes in abundance (S.D. < 0.1 for percent FPKM values of transcript across all four experimental groups) were excluded. Abbreviations: U, unknown functions; D, diverse functions; M, metabolism; RTT, replication, transcription, and translation; TRP, transport; CS, cytoskeletal and structural; PROT, proteolysis; DIG, blood digestive; CSR, chemosensory reception; R/S/M, redox, stress, and mitochondrial; I, immunity.

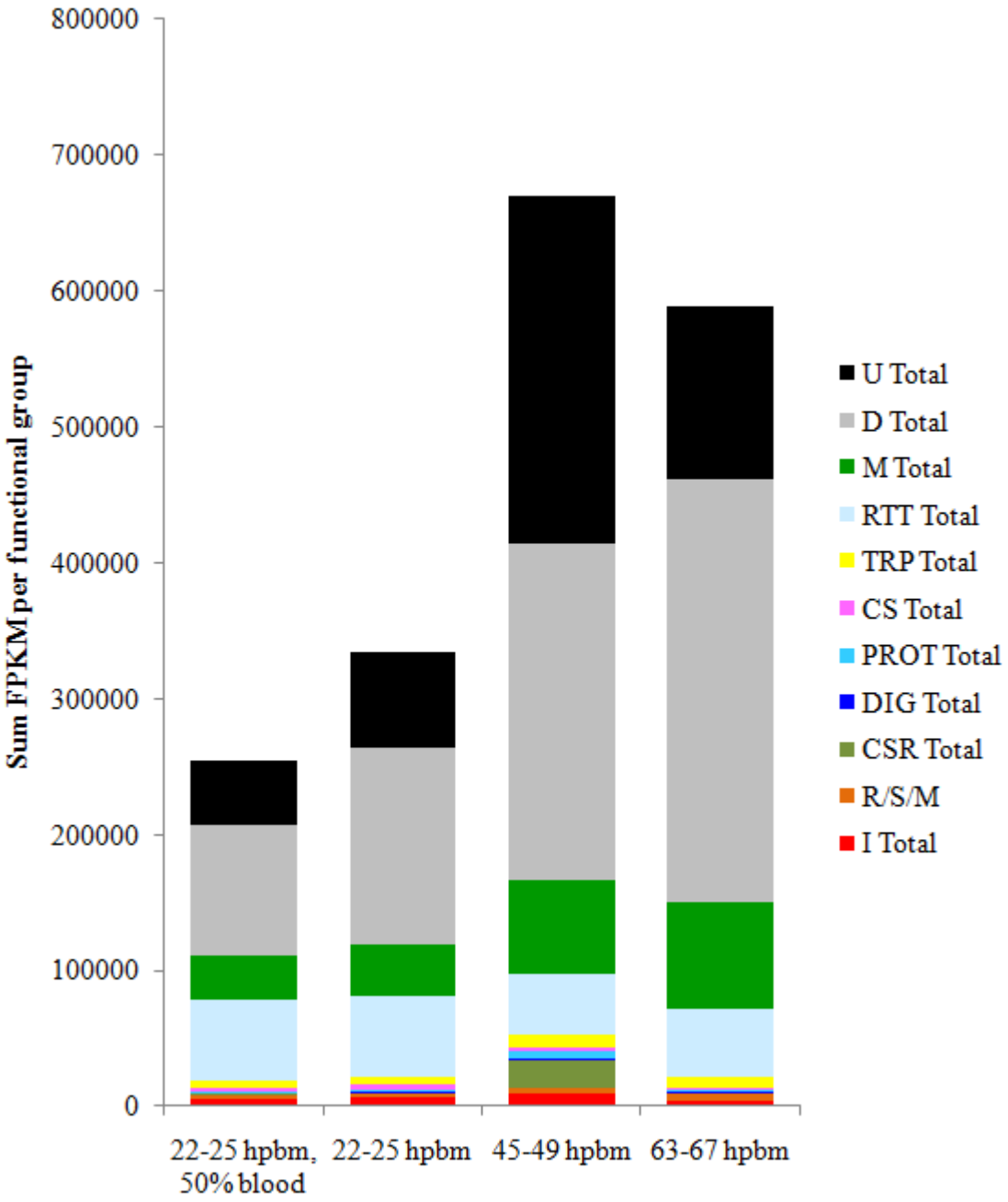


Figure 5. Sum normalized abundance of differentially regulated *Aedes aegypti* transcripts per functional group during oogenesis. Units are in FPKM (see text), and transcripts that showed relatively small changes in abundance (S.D. < 0.1 for percent FPKM values of transcript across all four experimental groups) were excluded. Abbreviations: U, unknown functions; D, diverse functions; M, metabolism; RTT, replication, transcription, and translation; TRP, transport; CS, cytoskeletal and structural; PROT, proteolysis; DIG, blood digestive; CSR, chemosensory reception; R/S/M, redox, stress, and mitochondrial; I, immunity.

Table 1. Illumina mRNA-seq experimental design and sequencing output from eight separate flow cell lanes.

Developmental stage	n	Species	Hours post blood meal	Blood meal concentration	Read length (nt)	No. Reads sequenced	No. reads after adapter trimming
Follicular atresia (20% atretic)	20	*Aedes triseriatus*	24-28	50%	104	24,961,983	24,723,150
Follicular atresia (10% atretic)	60	*Aedes triseriatus*	24-28	100%	104	25,690,481	25,231,826
Nurse cell death	40	*Aedes triseriatus*	51-55	100%	104	29,078,218	28,468,615
Follicular epitheilum death	20	*Aedes triseriatus*	79-82	100%	104	28,499,960	27,751,775
					***Aedes triseriatus* total**	**108,230,642**	**106,175,366**
Follicular atresia (20% atretic)	75	*Aedes aegypti*	22-25	50%	104	27,015,032	26,071,720
Follicular atresia (10% atretic)	75	*Aedes aegypti*	22-25	100%	104	27,803,900	26,374,091
Nurse cell death	21	*Aedes aegypti*	45-49	100%	104	28,595,563	28,326,455
Follicular epitheilum death	15	*Aedes aegypti*	63-67	100%	82	18,209,534	18,081,610
					***Aedes aegypti* total**	**101,624,029**	**98,853,876**

Table 2. Differential regulation of select transcripts during *Aedes aegypti* oogenesis. Values shown are normalized transcript abundance in FPKM (see text for description), and the background shading corresponds with transcript abundance per gene across all four groups.

VectorBase GeneID	Description	Group	22-25 hpbm, 50% blood	22-25 hpbm	45-49 hpbm	63-67 hpbm
AAEL004230	hypothetical protein	U	391.97	253.13	30.49	27.92
AAEL009444	hypothetical protein	U	6.95	0.60	0.65	1.65
AAEL007879	hypothetical protein	U	1563.42	1549.63	904.03	188.39
AAEL010327	six/sine homebox transcription factors	D	458.71	358.92	66.43	88.38
AAEL008802	conserved hypothetical protein	U	2978.32	3598.80	919.67	228.02
AAEL004498	hypothetical protein	U	1106.54	1237.59	332.59	295.75
AAEL006617	conserved hypothetical protein	U	613.22	868.32	407.62	32.63
AAEL001308	CRAL/TRIO domain-containing prot.	TRP	412.60	514.32	106.39	22.55
AAEL005763	lysosomal alpha-mannosidase	M	1241.59	1763.92	565.09	64.68
AAEL007555	acyl-coa dehydrogenase	M	732.46	952.25	321.66	219.76
AAEL003443	threonine dehydrogenase	D	409.37	510.65	5.16	27.68
AAEL008789	apolipophorin-III, putative	D	2046.60	3536.87	610.62	538.90
AAEL000961	conserved hypothetical protein	D	1531.57	1559.23	927.48	185.15
AAEL007686	conserved hypothetical protein	D	467.93	666.54	362.92	61.77
AAEL010344	SEC14, putative	D	546.09	1026.15	583.09	69.13
AAEL008303	calponin/transgelin	CS	653.64	693.70	296.32	68.78
AAEL000566	conserved hypothetical protein	U	19.71	18.64	802.45	17.02
AAEL011297	hypothetical protein	U	16.27	25.46	883.09	27.99
AAEL013027	conserved hypothetical protein	U	12136.60	31605.70	159849.00	35.97
AAEL003766	hypothetical protein	U	45.98	69.78	1469.33	23.13
AAEL001146	n-acetylgalactosaminyltransferase	M	12.04	10.75	489.20	4.72
AAEL002662	elongase, putative	M	192.76	320.22	1949.77	7.72
AAEL007213	delta(9)-desaturase, putative	M	46.57	76.93	1192.70	127.12
AAEL007302	conserved hypothetical protein	D	16.63	27.97	444.85	27.48
AAEL013585	conserved hypothetical protein	D	21.67	20.67	657.43	62.92
AAEL011491	Odorant-binding protein 50d, putative	CSR	14.22	10.34	773.44	14.62
AAEL005212	hypothetical protein	U	744.73	1033.18	2206.60	33787.00
AAEL004521	conserved hypothetical protein	U	22.83	27.04	54.48	503.62
AAEL003687	histone h2a	RTT	67.55	94.87	258.49	787.16
AAEL003706	histone h2a	RTT	66.30	90.74	258.72	823.65
AAEL005752	lysosomal alpha-mannosidase	M	10.41	10.55	74.65	1473.22
AAEL005337	carbonic anhydrase	M	117.54	162.21	87.46	663.95
AAEL003742	glucosylceramidase	M	50.66	53.03	51.85	282.90
AAEL011088	conserved hypothetical protein	D	18.93	15.51	14.55	617.77
AAEL006143	conserved hypothetical protein	D	35.08	29.98	41.58	1111.16
AAEL014128	hypothetical protein	D	529.75	583.40	1865.38	8304.72
AAEL005593	conserved hypothetical protein	D	58.39	50.40	49.11	278.01

References

Bartholomay, Lyric C., Robert M. Waterhouse, George F. Mayhew, Corey L. Campbell, Kristin Michel, Zhen Zou, Jose L. Ramirez, Suchismita Das, Kanwal Alvarez, Peter Arensburger, Bart Bryant, Sinead B. Chapman, Yuemei Dong, Sara M. Erickson, S. H. P. Parakrama Karunaratne, Vladimir Kokoza, Chinnappa D. Kodira, Patricia Pignatelli, Sang Woon Shin, Dana L. Vanlandingham, Peter W. Atkinson, Bruce Birren, George K. Christophides, Rollie J. Clem, Janet Hemingway, Stephen Higgs, Karine Megy, Hilary Ranson, Evgeny M. Zdobnov, Alexander S. Raikhel, Bruce M. Christensen, George Dimopoulos, and Marc A. T. Muskavitch. 2010. Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* 330 (6000):88-90.

Cave, M. 1982. Morphological manifestations of ribosomal DNA amplification during insect oogenesis. In *Insect Ultrastructure*, edited by R. King and H. Akai. New York: Plenum Press.

Clements, A. N., and M. R. Boocock. 1984. Ovarian development in mosquitoes: stages of growth and arrest, and follicular resorption. *Physiological Entomology* 9 (1):1-8.

Cramton, S. E., and F. A. Laski. 1994. *string of pearls* encodes *Drosophila* ribosomal protein S2, has *Minute*-like characteristics, and is required during oogenesis. *Genetics* 137 (4):1039-1048.

Erickson, S. M., K. B. Platt, B. J. Tucker, R. Evans, S. Tiawsirisup, and W. A. Rowley. 2006. The potential of *Aedes triseriatus* (Diptera: Culicidae) as an enzootic vector of West Nile virus. *Journal of Medical Entomology* 43 (5):966-970.

Fiil, A. 1978. Follicle cell bridges in the mosquito ovary: syncytia formation and bridge morphology. *J Cell Sci* 31 (1):137-143.

Girard, Yvette A., Bradley S. Schneider, Charles E. McGee, Julie Wen, Violet C. Han, Vsevolod Popov, Peter W. Mason, and Stehpen Higgs. 2007. Salivary gland morphology and virus transmission during long-term cytopathologic West Nile virus infection in *Culex* mosquitoes. *Am J Trop Med Hyg* 76 (1):118-128.

Gubler, Duane J. 1998. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* 11 (3):480-496.

Keene, Kimberly M., Brian D. Foy, Irma Sanchez-Vargas, Barry J. Beaty, Carol D. Blair, and Ken E. Olson. 2004. RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America* 101 (49):17240-17245.

Kim, Mijung, and David Denlinger. 2010. A potential role for ribosomal protein S2 in the gene network regulating reproductive diapause in the mosquito *Culex pipiens*. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* 180 (2):171-178.

Malagoli, Davide, Fabio Abdalla, Yang Cao, Qili Feng, Kozo Fujisaki, Ales Gregorc, Tomohide Matsuo, Ioannis Nezis, Issidora Papassideri, Mikls Sass, Elaine C. M. Silva-Zacarin, Gianluca Tettamanti, and Rika Umemiya-Shirafuji. 2010. Autophagy and its physiological relevance in arthropods: Current knowledge and perspectives. *Autophagy* 6 (5).

Marioni, John C., Christopher E. Mason, Shrikant M. Mane, Matthew Stephens, and Yoav Gilad. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* 18 (9):1509-1517.

Nene, Vishvanath, Jennifer R. Wortman, Daniel Lawson, Brian Haas, Chinnappa Kodira, Zhijian Tu, Brendan Loftus, Zhiyong Xi, Karyn Megy, Manfred Grabherr, Quinghu Ren, Evgeny M. Zdobnov, Neil F. Lobo, Kathryn S. Campbell, Susan E. Brown, Maria F. Bonaldo, Jingsong Zhu, Steven P. Sinkins, David G. Hogenkamp, Paolo Amedeo, Peter Arensburger, Peter W. Atkinson, Shelby Bidwell, Jim Biedler, Ewan Birney, Robert V. Bruggner, Javier Costas, Monique R. Coy, Jonathan Crabtree, Matt Crawford, Becky deBruyn, David DeCaprio, Karin Eiglmeier, Eric Eisenstadt, Hamza El-Dorry, William M. Gelbart, Suely L. Gomes, Martin Hammond, Linda I. Hannick, James R. Hogan, Michael H. Holmes, David Jaffe, J. Spencer Johnston, Ryan C. Kennedy, Hean Koo, Saul Kravitz, Evgenia V. Kriventseva, David Kulp, Kurt LaButti, Eduardo Lee, Song Li, Diane D. Lovin, Chunhong Mao, Evan Mauceli, Carlos F. M. Menck, Jason R. Miller, Philip Montgomery, Akio Mori, Ana L. Nascimento, Horacio F. Naveira, Chad Nusbaum, Sinéad O'Leary, Joshua Orvis, Mihaela Pertea, Hadi Quesneville, Kyanne R. Reidenbach, Yu-Hui Rogers, Charles W. Roth, Jennifer R. Schneider, Michael Schatz, Martin Shumway, Mario Stanke, Eric O. Stinson, Jose M. C. Tubio, Janice P. VanZee, Sergio Verjovski-Almeida, Doreen Werner, Owen White, Stefan Wyder, Qiandong Zeng, Qi Zhao, Yongmei Zhao, Catherine A. Hill, Alexander S. Raikhel, Marcelo B. Soares, Dennis L. Knudson, Norman H. Lee, James Galagan, Steven L. Salzberg, Ian T. Paulsen, George Dimopoulos, Frank H. Collins, Bruce Birren, Claire M. Fraser-Liggett, and David W. Severson. 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316 (5832):1718-1723.

Nezis, Ioannis, Dimitrios Stravopodis, Lukas Margaritis, and Issidora Papassideri. 2006. Autophagy is required for the degeneration of the ovarian follicular epithelium in higher Diptera. *Autophagy* 2 (4):297-298.

Reese, Sara, Eric Mossel, Meaghan Beaty, Eric Beck, Dave Geske, Carol Blair, Barry Beaty, and William Black. 2010. Identification of super-infected *Aedes triseriatus* mosquitoes collected as eggs from the field and partial characterization of the infecting La Crosse viruses. *Virology Journal* 7 (1):1-27.

Roy-Chaudhuri, Biswajoy, Narayanaswamy Kirthi, and Gloria M. Culver. 2010. Appropriate maturation and folding of 16S rRNA during 30S subunit biogenesis are critical for translational fidelity. *Proceedings of the National Academy of Sciences* 107 (10):4567-4572.

Shelly, Spencer, Nina Lukinova, Shelly Bambina, Allison Berman, and Sara Cherry. 2009. Autophagy is an essential component of *Drosophila* immunity against vesicular stomatitis virus. *Immunity* 30 (4):588-598.

Shine, J., and L. Dalgarno. 1973. Occurrence of heat-dissociable ribosomal RNA in insects: The presence of three polynucleotide chains in 26 S RNA from cultured *Aedes aegypti* cells. *Journal of Molecular Biology* 75 (1):57-72.

Uchida, K., M. Nishizuka, D. Ohmori, T. Ueno, Y. Eshita, and A. Fukunaga. 2004. Follicular epithelial cell apoptosis of atretic follicles within developing ovaries of the mosquito *Culex pipiens pallens*. *J Insect Physiol* 50 (10):903-12.

Volozina, N. V. 1967. The effect of the amount of blood taken and additional carbohydrate nutrition on oogenesis in females of blood-sucking mosquitoes of the genus *Aedes* (Diptera, Culicidae) of various weights and ages. *Entomological Review* 46:27-32.

Wang, Hua, Carol D. Blair, Ken E. Olson, and Rollie J. Clem. 2008. Effects of inducing or inhibiting apoptosis on Sindbis virus replication in mosquito cells. *J Gen Virol* 89 (11):2651-2661.

Weaver, Scott C., and William K. Reisen. 2010. Present and future arboviral threats. *Antiviral Research* 85 (2):328-345.

Zimmermann, Katja C., Ricci Jean-Ehrland, Nathalie M. Droin, and Douglas R. Green. 2002. The role of ARK in stress-induced apoptosis in *Drosophila c*ells. *The Journal of Cell Biology* 156 (6):1077-1087.

CHAPTER 5. CONCLUSIONS

The RNA-seq transcriptome and RNAi experiments presented here give us many new insights into the molecular biology of mosquito oogenesis, particularly related to nutritional signaling mechanisms involved in follicular atresia. The RNAi experiments presented in Chapter 2 were supported by RNA-seq analysis of TOR expression, which appears to be transcriptionally regulated during amino acid deprivation.

RNA-seq analysis described in Chapter 4 showed that TOR transcript levels decreased by approximately 40% in *Ae. triseriatus* when mosquitoes were fed a 50% diluted blood meal compared to an undiluted blood meal at the same time point. As mentioned in Chapter 4, the diluted blood meal resulted in approximately a twofold increase in the number of atretic follicles compared to the undiluted blood-fed mosquitoes. Coincidentally, the same knockdown of TOR mRNA was accomplished by injection of dsTOR, as described in Chapter 2. This 40% knockdown of TOR transcript by RNAi caused a nearly twofold increase in follicular atresia when mosquitoes were fed an undiluted blood meal. The similarity of the phenotypes observed between these two experiments highlights the importance of TOR as an upstream activator in the process of follicular atresia. Taken together, these data suggest that TOR may be transcriptionally downregulated during periods of amino acid starvation (as in Chapter 4), and this downregulation of TOR induces folicular atresia through activation of S6K (as shown in Chapter 2), decreasing vitellogenesis and/or follicle cell growth. A different response to starvation was reported in the silkworm moth (*Bombyx mori*), in which transcription of TOR mRNA increased in *B. mori* fat body tissue compared to a well-fed control after 24 hours of starvation (Zhou et al. 2010). Complete

starvation elicits a very different response than amino acid deprivation, and the regulation of Ser(2448) phosphorylation is a convergence point for the counteracting regulatory effects of amino acid levels and growth factors (Navé et al. 1999). There is much to learn about mosquito nutritional signaling pathways, and the ovary transcriptomes of *Ae. aegypti* and *Ae. triseriatus* presented here will provide invaluable insights into the molecular mechanisms that govern programmed cell death during oogenesis.

These data may translate into applied methodologies for mosquito control. For example, RNAi targeting critical mosquito physiologies could provide new mosquitocidal products (Zhang, Zhang, and Zhu 2010), and genes in the TOR pathway could be potential targets for such a control strategy. The effects of amino acid deprivation should first be determined in larval mosquitoes, which are more amenable to RNAi-based control strategies, and the duration of transcript suppression into the adult stage should be investigated.

References

Navé, Barbara T., D. Margriet Ouwens, Dominic J. Withers, Dario R. Alessi, and Shepherd Peter R. 1999. Mammalian target of rapamycin is a direct target for protein kinase B: identification of a convergence point for opposing effects of insulin and amino-acid deficiency on protein translation. *Biochem J.* 344 (2):427-431.

Zhang, X., J. Zhang, and K. Y. Zhu. 2010. Chitosan/double-stranded RNA nanoparticle-mediated RNA interference to silence chitin synthase genes through larval feeding in the African malaria mosquito (*Anopheles gambiae*). *Insect Molecular Biology* 19 (5):683-693.

Zhou, S., Q. Zhou, Y. Liu, S. Wang, D. Wen, Q. He, W. Wang, W. G. Bendena, and S. Li. 2010. Two Tor genes in the silkworm *Bombyx mori*. *Insect Molecular Biology* 19 (6):727-735.

APPENDIX A. *AEDENOVO* USER MANUAL AND SCRIPT

User Manual

```
### Aedenovo ###

Written by Patrick D. Jennings
Iowa State University
patrick.d.jennings@gmail.com



This program was created to use the popular, open-source programs
Velvet (http://www.ebi.ac.uk/~zerbino/velvet/), BLAST
(http://blast.ncbi.nlm.nih.gov/), CD-HIT-EST
(http://www.bioinformatics.org/cd-hit/), and Perl
(http://www.perl.org/) to assemble a de novo transcriptome from
known transcript sequences.  This program is useful if you have
short-read data (such as Illumina RNA-seq) and you are trying to
assemble homologs of known genes.

For full functionality, you should have these programs in your
PATH: velvetg, velveth, cd-hit-est, blastall.  You should also
modify the first line of denovo.pl (the "shebang") to reflect
your perl binary directory.  For example:

#!/usr/bin/perl

Successful assembly was achieved with mosquito (Aedes aegypti)
transcripts using Illumina mRNA-seq 104 base pair reads in FASTQ
format.  In most cases, the assembly produced assembled contigs that
were longer than the currently annotated gene sequences.

For more complete assembly, short reads were searched for adapter
sequences that are an artifact of the Illumina sequencing reaction,
and the adapter sequences were trimmed, and sequences were removed
if more than 40% of the sequence was adapter sequence.  For Illumina
mRNA-seq, the adapter sequences removed were:

5'-CCGCTCTTCCGATCT-[cDNA sequence]-3'
5'-[cDNA sequence]-AGATCGGAAGAGCGG-5'

If you wish to remove these same adapter sequences from your FASTQ
files, simply use the option trim=yes.  This will trim adapter
sequences and remove reads that are shorter than 60% of the original
length of the reads, and save the new FASTQ files with the extension
_trimmed60cln.fq.

For a list of all standard options, type Aedenovo.pl at the command
prompt with no arguments, and a description of available arguments
will be printed.
```

Perl Script

```
#!/usr/bin/perl
use warnings;
use strict;

############################################################
## Instructions if no command line arguments are given ##
############################################################

sub commandLineOptions {
print('                                                          '."\n");
print('  ## Aedenovo ## - Patrick Jennings, 2011                 '."\n");
print('                                                          '."\n");
print('  The programs Velvet, CD-HIT-EST, and BLAST must be installed
'."\n");
print('  locally for BLASTnovo.pl to be fully functional.     '."\n");
print('  All input options are given in this format: option=value
'."\n");
print('  The following options exist:                            '."\n");
print('                                                          '."\n");
print('  queryfile=knownSequences.fa   <required; known sequences>'."\n");
print('                                                          '."\n");
print('  subject=velvetcontigs.fa                                '."\n");
print('   The subject file should be the assembled velvet contigs '."\n");
print('   using the additive multiple-k method. FASTA format only.'."\n");
print('  --OR--                                                  '."\n");
print('  shortreads=s_1_sequence.txt                             '."\n");
print('  trim=(yes|no) <OPTIONAL; default: trim=no>              '."\n");
print('    trims adapter sequences from the 3\' and 5\' ends, and '."\n");
print('    removes sequences that contained more than 40% adapter.'."\n");
print('  kmers=59,55,45,39,21  <OPTIONAL; default values shown here>
'."\n");
print('    If Velvet assembly has not been completed, this program'."\n");
print('    will do the assembly with these options specified.    '."\n");
print('                                                          '."\n");
print('  blastdb=velvetcontigs                                   '."\n");
print('    The local BLAST database, assembled from velvet contigs'."\n");
print('    with the command \"formatdb -p F -o T\".              '."\n");
print('  --OR--                                                  '."\n");
print('  formatdb=T  <OPTIONAL; default if "shortreads" is specified>
'."\n");
print('    If this option is given, a BLAST database will be created
'."\n");
print('    from the subject FASTA file specified.  This option is '."\n");
print('    mutually exclusive with blastdb=<velvetcontigs>.      '."\n");
print('                                                          '."\n");
print('  blastall=-p#blastn#-a#1  <OPTIONAL>                     '."\n");
print('    blastall command line agruments that can be modified,  '."\n");
print('    separated by # signs where spaces would go on the     '."\n");
print('    command prompt.  -i, -d, -m, and -o cannot be changed. '."\n");
print('    Example:  blastall=-a#8#-e#0.01                       '."\n");
print('       this example uses 8 threads and an e-value of 0.01  '."\n");
print('      Default = -p blastn -e 0.1 -a 1                     '."\n");
print('                                                          '."\n");
```

```
print('  cutoff=100  <OPTIONAL>                                        '."\n");
print('    This is the minimum cutoff value for a BLAST alignment '."\n");
print('      in bits.  Default=100                                     '."\n");
print('                                                                '."\n");
print('                                                                '."\n");
}

#########################################################
## The following defines the command line arguments ##
#########################################################

my %args; # this hash table holds the comand line arguments

# parse the comand line arguments to the %args hash. If no arguments are
given, print out the above help message.
if($#ARGV == -1){
        &commandLineOptions();
        die "\n";
}
for(my $i=0; $i <= $#ARGV; $i++){
        my @f = split(/=/,$ARGV[$i]);
        $args{"$f[0]"} = $f[1];
}


my $subjectfile;
my $trimmedout;
my $shortreads;
my $removeblastdb = "no";
my $formatdb="T";

unless($args{"blastdb"}){
    print "Do you want to remove the BLAST database after alignment is
finished\? \<yes or no\>\n";
    $removeblastdb = <STDIN>;
}

my $in_seq = 0;
if($args{"shortreads"}){
    $shortreads = $args{"shortreads"};

###############################################
## Trim sequences, if trim=yes is specified. ##
###############################################

    if($args{"trim"} eq 'yes'){

        open SHORTREADSFILE, "$shortreads" or die $!;
        if($shortreads =~ m/(.+)\.(fastq|fq|FASTQ|FQ)$/){
            $shortreads = $1;
        }
        $trimmedout = $shortreads . '_trimmed60cln.fq';
        open TRIMMEDOUT, ">>", "$trimmedout" or die $!;
```

```
# Defines the length of the reads and where the 40% cutoff should be.
        my $readlength = 0;
        while(<SHORTREADSFILE>){
            if (/^\@(ILLUMINA|HWI)/){
                $in_seq = 1;
            }elsif($in_seq){
                $readlength = length($_);
                last;
            }
        }
        print "Read length\: $readlength bp\n";
        my $fortypercentexact = ($readlength * 0.4);
        my $fortypercent = sprintf("%.0f", $fortypercentexact);


# Removes adapter sequences from the 5' end of the reads.
        my $in_seq = 0;
        my $len = 0;
        my $in_qual = 0;
        print "removing 5\' adapter sequences...\n";
        while(<SHORTREADSFILE>){
            chomp;
            if (/^(\@ILLUMINA|\@HWI)/){
                $in_seq = 1;
                print TRIMMEDOUT $_, "\n";
            }elsif($in_seq){
                if (/^(.*)ACGCTCTTCCGATCT(.+)$/){
                    $len = (15 + length($1));
                    print TRIMMEDOUT $2;
                } elsif(/^CGCTCTTCCGATCT(.+)$/){
                    $len = 14;
                    print TRIMMEDOUT $1;
                } elsif(/^GCTCTTCCGATCT(.+)$/){
                    $len = 13;
                    print TRIMMEDOUT $1;
                } elsif(/^CTCTTCCGATCT(.+)$/){
                    $len = 12;
                    print TRIMMEDOUT $1;
                } elsif(/^TCTTCCGATCT(.+)$/){
                    $len = 11;
                    print TRIMMEDOUT $1;
                } elsif(/^CTTCCGATCT(.+)$/){
                    $len = 10;
                    print TRIMMEDOUT $1;
                } elsif(/^TTCCGATCT(.+)$/){
                    $len = 9;
                    print TRIMMEDOUT $1;
                } elsif(/^TCCGATCT(.+)$/){
                    $len = 8;
                    print TRIMMEDOUT $1;
                } elsif(/^CCGATCT(.+)$/){
                    $len = 7;
                    print TRIMMEDOUT $1;
                } elsif(/^CGATCT(.+)$/){
                    $len = 6;
```

```
            print TRIMMEDOUT $1;
        } elsif(/^GATCT(.+)$/){
            $len = 5;
            print TRIMMEDOUT $1;
        } elsif(/^ATCT(.+)$/){
            $len = 4;
            print TRIMMEDOUT $1;
        } else {
            $len = 0;
            print TRIMMEDOUT $_;
        }
        print TRIMMEDOUT "\n";
        $in_qual = 1;
        $in_seq = 0;
    }elsif($in_qual==1 && /^(\+ILLUMINA|\+HWI)/){
        $in_qual = 2;
        print TRIMMEDOUT $_, "\n";
    }elsif($in_qual==2){
        print TRIMMEDOUT substr($_, $len);
        print TRIMMEDOUT "\n";
        $len = 0;
        $in_qual = 0;
    }
}
close(SHORTREADSFILE);
close(TRIMMEDOUT);


open SHORTREADSFILE2, "<", "$trimmedout" or die $!;
open TRIMMEDOUT2, ">>", "trimmedout2temp.fq" or die $!;
$in_seq = 0;
$len = 0;
$in_qual = 0;
my $first_len = 0;
print "Removing 3\' adapter sequences...\n";
while(<SHORTREADSFILE2>){
    chomp;
    if (/^(\@ILLUMINA|\@HWI)/){
        $in_seq = 1;
        print TRIMMEDOUT2 $_, "\n";
    }elsif($in_seq){
        if (/^(.+)AGATCGGAAGAGCG(.*)$/){
            $len = (15 + length($2));
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGAAGAGCG$/){
            $len = 14;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGAAGAGC$/){
            $len = 13;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGAAGAG$/){
            $len = 12;
```

```
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGAAGA$/){
            $len = 11;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGAAG$/){
            $len = 10;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGAA$/){
            $len = 9;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGGA$/){
            $len = 8;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCGG$/){
            $len = 7;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATCG$/){
            $len = 6;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGATC$/){
            $len = 5;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } elsif(/^(.+)AGAT$/){
            $len = 4;
            $first_len = length($1);
            print TRIMMEDOUT2 $1;
        } else {
            $len = 0;
            $first_len = length($_);
            print TRIMMEDOUT2 $_;
        }
        print TRIMMEDOUT2 "\n";
        $in_qual = 1;
        $in_seq = 0;
    }elsif($in_qual==1 && /^(\+ILLUMINA|\+HWI)/){
        $in_qual = 2;
        print TRIMMEDOUT2 $_, "\n";
    }elsif($in_qual==2){
        print TRIMMEDOUT2 substr($_, 0, $first_len);
        print TRIMMEDOUT2 "\n";
        $len = 0;
        $in_qual = 0;
    }
}
close(SHORTREADSFILE2);
close(TRIMMEDOUT2);
my $removefile = 'rm ' . $trimmedout;
```

```
        system($removefile);


        open SHORTREADSFILE3, "<", "trimmedout2temp.fq" or die $!;
        open TRIMMEDOUT3, ">>", "$trimmedout" or die $!;
        my $flag = 0;
        my $line1;
        my $line2;
        my $line3;
        my $line4;
        print "Removing reads shorter than $fortypercent bp...\n";
        while(<SHORTREADSFILE3>){
            chomp;
            if(/^\@(ILLUMINA|HWI)/){
                $flag = 1;
                $line1 = $_;
            }elsif($flag==1){
                $line2 = $_;
                $flag = 0;
            }elsif(/^\+(ILLUMINA|HWI)/){
                $flag = 2;
                $line3 = $_;
            }elsif($flag==2){
                $line4 = $_;
                if(length($line2)>$fortypercent){
                    print TRIMMEDOUT3 "$line1\n$line2\n$line3\n$line4\n";
                }
                $flag = 0;
            }
        }
        $shortreads = $trimmedout;
        close(SHORTREADSFILE3);
        close(TRIMMEDOUT3);
        my $removetempfile = 'rm trimmedout2temp.fq';
        system($removetempfile);
    }
}

########################################
## Additive multiple-k velvet assembly ##
########################################

# Define the k-mer values to use for additive multiple-k velvet assembly
my $kmerscombined;
if($args{"kmers"}){
    $kmerscombined = $args{"kmers"};
}else{
    $kmerscombined = "59,55,45,39,21";
}
my @kmers = split(",", $kmerscombined);


# Perform the additive multiple-k velvet assembly
my $velvetruncommand;
if($args{"shortreads"}){
```

```
    print "Performing Velvet Additive Multiple-k assembly...\n";
    for(my $ind=0; $ind < scalar(@kmers); $ind++){
        $velvetruncommand = 'velveth k' . $kmers[$ind] . ' ' .
$kmers[$ind] . ' -fastq -short ' . $shortreads . ' ';
        print "   $velvetruncommand\n";
        system($velvetruncommand);
        $velvetruncommand = 'velvetg k' . $kmers[$ind] . ' -cov_cutoff 8 -
very_clean yes ; cd k' . $kmers[$ind] . ' ; rm Graph2 Graph PreGraph
stats.txt ; cd .. ; ';
        print "   velvetg k$kmers[$ind] -cov_cutoff 8 -very_clean yes\n";
        system($velvetruncommand);
        print "   Removing unnecessary files...\n";
    }
    $velvetruncommand = 'cat ';
    for(my $ind=0; $ind < scalar(@kmers); $ind++){
        $velvetruncommand .= 'k' . $kmers[$ind] . '/Log ';
    }
    $velvetruncommand .= '> MasterLog1.txt';
    system($velvetruncommand);
    print "   Combining contig files into \"contigs_All\.fa\"...\n";
    for(my $ind=0; $ind < scalar(@kmers); $ind++){
        my $contigfilename = 'k' . $kmers[$ind] . '/contigs.fa';
        open VELVETCONTIGFILE, "<", "$contigfilename";
        open VELVETOUTALL, ">>", "contigs_All.fa";
        while(<VELVETCONTIGFILE>){
            chomp;
            if(/^\>(.+)$/){
                print VELVETOUTALL "\>$1_k$kmers[$ind]\n";
            }else{
                print VELVETOUTALL "$_\n";
            }
        }
        close(VELVETOUTALL);
        close(VELVETCONTIGFILE);
    }
    $velvetruncommand = 'rm -r k*';
    system($velvetruncommand);
    $subjectfile = 'contigs_All.fa';

############################
## CD-HIT-EST filtering ##
############################

    my $cdhitestcommand = 'cd-hit-est -i contigs_All.fa -d 0 -o
contigs_All_cdhitest.fa -c 0.9 -n 8 -T 0 -r 1 -G 1 -g 1 -b 20 -s 0.0 -aL
0.0 -aS 0.9';
    system($cdhitestcommand);
}

############
## BLAST ##
############

# blast minimum score (in bits)
my $cutoff;
```

```
if($args{"cutoff"}){
    $cutoff = $args{"cutoff"};
}else{
# default is 100 bits. After BLAST is used, sequence alignments below
$cutoff bits will be discarded.
    $cutoff = 100;
}

# subject FASTA file (your assembled velvet contigs)

if($args{"subject"}){
    $subjectfile = $args{"subject"};
}elsif($args{"shortreads"}){
    $subjectfile = "contigs_All_cdhitest.fa";
}else{
    print "WARNING\: NO SUBJECT FILE SPECIFIED\!\nTerminating program.";
    exit;
}

# Specify or format the blast database
my $blastdb;
if($args{"blastdb"}){
    $blastdb = $args{"blastdb"};
}elsif($args{"shortreads"} || $args{"formatdb"}){
    if($args{"formatdb"}){
        if($args{"formatdb"} =~ m/^[Ff]/){
            last;
        }
    }
    print "Formatting local BLAST database\.\.\.\n";
    my $runcommand = 'formatdb -p F -o T -i ' . $subjectfile;
    system($runcommand);
    $blastdb = $subjectfile;
}else{
    $blastdb = $subjectfile;
}

# the FASTA file containing known sequences
my $queryfile = $args{"queryfile"};
my $queryformatted = "TEMP01AQUERYFORMAT.fa";

# this replaces spaces with underscores in query file
print "Formatting query file $queryfile\.\.\.\n";
open QUERY, "$queryfile" or die $!;
open QUERY2, ">", "$queryformatted" or die $!;
while(<QUERY>){
    chomp;
    $_ =~ s/\ /\_/g;
    print QUERY2 "$_\n";
}
close(QUERY);
close(QUERY2);

# save blastall command line arguments as variable $blastallCommands.
my $blastallCommands;
```

```
if($args{"blastall"}){
    my @blastargs = split(/#/, $args{"blastall"});
    $blastallCommands = ' ' . join(" ", @blastargs) . ' ';
    # BLASTn is default unless otherwise specified.
    unless($blastallCommands =~ m/\ \-p\ ([^\ ]+)\ /){
        $blastallCommands .= '-p blastn ';
    }
    # -e is 0.1 by default.
    unless($blastallCommands =~ m/\ \-e\ ([^\ ]+)\ /){
        $blastallCommands .= '-e 0.1 ';
    }
    # this automatically fixes errors that will cause problems later on.
    $blastallCommands =~ s/\ \-m\ \d\ /\ /g;
    $blastallCommands =~ s/\ \-i\ ([^\ ]+)\ /\ /g;
    $blastallCommands =~ s/\ \-d\ ([^\ ]+)\ /\ /g;
    $blastallCommands =~ s/\ \-o\ ([^\ ]+)\ /\ /g;
    $blastallCommands .= ' -m 8 -o tableout1.txt -i ' . $queryformatted .
' -d ' . $blastdb;

}else{
    # if no blast arguments are given, this sets the default arguments.
    $blastallCommands = '-p blastn -m 8 -d ' . $blastdb . ' -e 0.1 -i ' .
$queryformatted . ' -a 1 -o tableout1.txt';
}


# BLAST alignment
    print "Executing BLAST alignment\.\.\.\n";
    my $runblastcommand = "blastall " . $blastallCommands;
    print "$runblastcommand\n";
    system($runblastcommand);

    if($removeblastdb =~ m/^[yY]/){
        my $blastRemoveCommand = "rm $blastdb\.n\* formatdb.log";
        system($blastRemoveCommand);
    }

##############################################
## sort alignments by query start position ##
##############################################
my $alignmentfileflag=0;
# This creates a human-readable alignment file, but decreases the accuracy
of assembly. This may be advantageous for manual assembly of a small
number of transcripts. To enable, remove "##" from the following lines.
##    print "Sorting BLAST hits\.\.\.\n";
##    my $sortcommand = 'sort -k 1,1 -k 7,7n --temporary-directory=./
tableout1.txt > alignment_file.txt';
##    system($sortcommand);
##    my $alignmentfileflag = 1;
if($alignmentfileflag==1){
    my $command = 'rm tableout1.txt ; mv alignment_file.txt
tableout1.txt';
    system($command);
    $alignmentfileflag = 2;
}
```

```
###################################
## remove low quality alignments ##
###################################

print "Removing alignments with scores below $cutoff bits\.\.\.\n";
open TABLEOUT2, "<", "tableout1.txt" or die $!;
open TABLEOUT3, ">", "tableout3.txt" or die $!;

my $repeated = "";
while(<TABLEOUT2>){
        chomp;
        my @ar = split(/\t/, $_);
        unless($ar[1] eq $repeated){   # removes duplicate entries from
the same contig
            if($ar[11]>$cutoff){
                print TABLEOUT3 join("\t", @ar), "\n";
            }
            $repeated = $ar[1];
        }
}
close(TABLEOUT3);
close(TABLEOUT2);
my $removecommand = 'rm tableout1.txt';
system($removecommand);

###############################################
## The following sets up the progress meter ##
###############################################

my $linecountA=0;

open TABLEOUT3, "tableout3.txt" or die $!;
while(<TABLEOUT3>){
    $linecountA++;
}
close(TABLEOUT3);
my $currentline = 0;
my $lineflag = 0;
my $part1 = ($linecountA * 0.1);
my $part2 = ($linecountA * 0.2);
my $part3 = ($linecountA * 0.3);
my $part4 = ($linecountA * 0.4);
my $part5 = ($linecountA * 0.5);
my $part6 = ($linecountA * 0.6);
my $part7 = ($linecountA * 0.7);
my $part8 = ($linecountA * 0.8);
my $part9 = ($linecountA * 0.9);

#########################################################
## align sequences into a human readable text file ##
#########################################################

my $tableout3 = "tableout3.txt";
```

```
# Load query sequences and velvet contigs into memory
my $contigshash = {};
open VELVETCONTIGS, "<", "$subjectfile" or die $!;
my $sequence="";
my $header="";
while(<VELVETCONTIGS>){
    chomp;
    if(/^\>(.*)$/ && $sequence){
        $contigshash->{"$header"} =  "$sequence";
        $sequence = "";
        $header = $1;
    }elsif(/^\>(.*)$/){
        $header = $1;
    }else{
        $sequence .= $_;
    }
}
$contigshash->{"$header"} = "$sequence";
close(VELVETCONTIGS);
undef $sequence;


my $queryhash = {};
open QUERYHASH, "<", "$queryformatted" or die $!;
my $inseq=0;
my $seq="";
$header="";
while(<QUERYHASH>){
    chomp;
    if(/^\>(.*)$/ && $seq){
        $queryhash->{"$header"} = "$seq";
        $seq = "";
        $header = $1;
    }elsif(/^\>(.*)$/){
        $header = $1;
        $seq = "";
    }else{
        $seq.=$_;
    }
}
$queryhash->{"$header"} = "$seq";
close(QUERYHASH);

# Now start the progress meter and executing the alignment
print "Aligning full length Velvet contigs with query sequences\.\.\.\n";
print "   Progress\:   0\%\n";

open OUTPUT, ">", "alignmentout.txt" or die $!;
open INPUT, "<", "$tableout3" or die $!;

my $match;
my $start;
my $stop;
my $stopq;
my $startq;
```

```
my $query;
my $seq3;
my $querysave = "x";

while(<INPUT>){
    chomp;
    $currentline++;
    if($currentline > $part1 && $lineflag==0){
        print "              10\%\n";
        $lineflag++;
    }elsif($currentline > $part2 && $lineflag==1){
        print "              20\%\n";
        $lineflag++;
    }elsif($currentline > $part3 && $lineflag==2){
        print "              30\%\n";
        $lineflag++;
    }elsif($currentline > $part4 && $lineflag==3){
        print "              40\%\n";
        $lineflag++;
    }elsif($currentline > $part5 && $lineflag==4){
        print "              50\%\n";
        $lineflag++;
    }elsif($currentline > $part6 && $lineflag==5){
        print "              60\%\n";
        $lineflag++;
    }elsif($currentline > $part7 && $lineflag==6){
        print "              70\%\n";
        $lineflag++;
    }elsif($currentline > $part8 && $lineflag==7){
        print "              80\%\n";
        $lineflag++;
    }elsif($currentline > $part9 && $lineflag==8){
        print "              90\%\n";
        $lineflag++;
    }elsif($currentline==($linecountA - 1)){
        print "             100\%\n";
    }
    my @array = split(/\t/, $_);
    $query = $array[0];
    $match = $array[1];
    $startq = $array[6];
    $stopq = $array[7];
    $start = $array[8];
    $stop = $array[9];
    if($query eq $querysave){
        &getseq();
    }else{
        &getquery();
        $querysave = $query;
        &getseq();
    }
}
close(INPUT);
close(OUTPUT);
```

```
# This part prints the query sequence. The vertical bar "|" should line up
at the start of the spaces. For visual comparisons, add tab characters to
make it line up.
sub getquery {
    if($queryhash->{"$query"}){
        print OUTPUT "$query\ \t\|";
        for(my $i=0; $i<8299; $i++){
            print OUTPUT "\ ";
        }
        print OUTPUT $queryhash->{"$query"}, "\n";
    }
}


# This prints the full length subject sequences, aligned to the query
sequence.

sub getseq {
    if($contigshash->{"$match"}){
        print OUTPUT "$match\t$startq\t$stopq\t$start\t$stop\t";
        if($start > $stop){
            my $rseq = reverse $contigshash->{"$match"};
            $rseq =~ tr/ACGT/TGCA/;
            my $seq3 = $rseq;
            for(my $i=0; $i<(8299+$startq-(length($seq3)-$start)); $i++){
                print OUTPUT "\ ";
            }
            print OUTPUT $seq3, "\n";
            undef $seq3;
        }else{
            for(my $i=0; $i<((8300 - $start) + $startq); $i++){
                print OUTPUT "\ ";
            }
            print OUTPUT $contigshash->{"$match"}, "\n";
        }
    }
}

$queryhash = {};
$contigshash = {};
$removecommand = 'rm tableout3.txt';
system($removecommand);


#####################################################
## compare overlapping ends and extend contigs ##
#####################################################

my $alignmentoutsave = "alignmentout.txt";      # input file from previous
step, temporary
my $filepath = "genebuild.txt";                 # output FASTA file (de
novo sequences)

my $flag1 = 0;
my $flag2 = 0;
```

```
my $totallen = 0;
my $spacelen = 0;
$header="";
$seq="";

print "Building consensus sequences \(genebuild\.txt\)\.\.\.\n";
print "   Progress\:   0\%\n";

# Another progress meter

open ALIGNMENTOUTTXT, "<", "$alignmentoutsave" or die $!;
$linecountA=0;
while(<ALIGNMENTOUTTXT>){
    $linecountA++;
}
$currentline = 0;
$lineflag = 0;
$part1 = ($linecountA * 0.1);
$part2 = ($linecountA * 0.2);
$part3 = ($linecountA * 0.3);
$part4 = ($linecountA * 0.4);
$part5 = ($linecountA * 0.5);
$part6 = ($linecountA * 0.6);
$part7 = ($linecountA * 0.7);
$part8 = ($linecountA * 0.8);
$part9 = ($linecountA * 0.9);
close(ALIGNMENTOUTTXT);

open ALIGNMENTOUTTXT, "<", "$alignmentoutsave" or die $!;
open TEXT, ">>", "$filepath" or die "$filepath cannot be opened.";
while(<ALIGNMENTOUTTXT>){
    chomp;
    $currentline++;
    if($currentline > $part1 && $lineflag==0){
        print "              10\%\n";
        $lineflag++;
    }elsif($currentline > $part2 && $lineflag==1){
        print "              20\%\n";
        $lineflag++;
    }elsif($currentline > $part3 && $lineflag==2){
        print "              30\%\n";
        $lineflag++;
    }elsif($currentline > $part4 && $lineflag==3){
        print "              40\%\n";
        $lineflag++;
    }elsif($currentline > $part5 && $lineflag==4){
        print "              50\%\n";
        $lineflag++;
    }elsif($currentline > $part6 && $lineflag==5){
        print "              60\%\n";
        $lineflag++;
    }elsif($currentline > $part7 && $lineflag==6){
        print "              70\%\n";
        $lineflag++;
    }elsif($currentline > $part8 && $lineflag==7){
```

```
        print "              80\%\n";
        $lineflag++;
    }elsif($currentline > $part9 && $lineflag==8){
        print "              90\%\n";
        $lineflag++;
    }elsif($currentline==($linecountA - 1)){
        print "             100\%\n";
    }
    if(/^(.*)\ *\t\|(\s+)([ACGTN]+)$/){
        $flag2 = 0;
        $header = "\>$1";
        if($seq){
            print TEXT "$seq\n$header\n";
            $seq = "";
        }else{
            print TEXT "$header\n";
        }
    }elsif(/^(.+)\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(\ +)([ACGTN]+)$/ &&
$flag2==0){
        $spacelen = $6;
        $seq = $7;
        $totallen = (length($6)+length($7));
        $flag2 = 1;
    }elsif(/^(.+)\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(\ +)([ACGTN]+)$/ &&
$flag2==1){
        my $seqcompare = $7;
        my $templen = $6;
        my $temptotallen = (length($6)+length($7));
        my $num=0;
        my $sub=0;
        if(length($seqcompare)>99){
            $num = 100;
            $sub = 75;
        }else{
            $num = length($seqcompare);
            $sub = (3 * $num / 4);
        }
        if(length($spacelen)>length($6)){
            my $compare = substr($seq, 25, $sub);
            if($seqcompare =~ m/^([ACGTN]+)$compare([ACGTN]*)$/){
                $seq = $1 . substr($seq, 25);
                $spacelen = $templen;
            }else{
                $compare = substr($seq, 2, $sub);
                if($seqcompare =~ m/^([ACGTN]+)$compare([ACGTN]*)$/){
                    $seq = $1 . substr($seq, 2);
                    $spacelen = $templen;
                }
            }
        }
        if($temptotallen>$totallen){
            my $compare2 = substr($seq, -$num, $sub);
            if($seqcompare =~ m/^([ACGTN]*)$compare2([ACGTN]+)$/){
                $seq = substr($seq, 0, -$num) . $compare2 . $2;
            }else{
```

```
                $compare2 = substr($seq, (-$sub - 2), $sub);
                if($seqcompare =~ m/^([ACGTN]*)$compare2([ACGTN]+)$/){
                    $seq = substr($seq, 0, (-$sub - 2)) . $compare2 . $2;
                }
            }
        }
    }
}
print TEXT "$seq\n";
close(TEXT);
close(ALIGNMENTOUTTXT);

if($args{"clean"}){
    if($args{"clean"} =~ m/^[Yy]/){
        my $removecommand = 'rm error.log MasterLog1.txt contigs_All*';
        system($removecommand);
    }
}
$removecommand = 'rm TEMP01AQUERYFORMAT.fa';
system($removecommand);
unless($alignmentfileflag==2){
    $removecommand = 'rm alignmentout.txt';
    system($removecommand);
}
```

APPENDIX B. BIOINFORMATICS PARAMETERS

The assembly programs were run from a UNIX environment on a Mac Pro with an 8-core 2.93GHz Quad Xeon 5570 CPU, 64GB 1066MHz DDR3 ECC SDRAM, and a 6 TB RAID 0 array, using remote connection through SSH Secure Shell version 3.2.9 build 282.

Bowtie/Tophat

```
$  bowtie-build -f aaegypti.SUPERCONTIGS-Liverpool.AaegL1.fa
   aaegypti.SUPERCONTIGS-Liverpool.AaegL1

$  tophat --phred64-quals --GTF aaegypti.BASEFEATURES_Liverpool-
   AaegL1.2.gff3 -p 8 --library-type fr-unstranded --rg-id Aedes_aegypti -
   -rg-sample 24hpbm50pct aaegypti.SUPERCONTIGS-Liverpool.AaegL1.fa
   s_5_trimmed60cln.fq
```

Velvet

The following chain of commands was used to perform the additive Multiple-k Velvet assembly as described by Surget-Groba and Montoya-Burgos (2010) and generate the values observed in Figure 3 of Chapter 3.

```
$  velveth60 k59 59 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg60 k59 -
   cov_cutoff 8 -very_clean yes ; rm k59/Graph2 k59/PreGraph k59/stats.txt
   ; velveth60 k57 57 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg60 k57
   -cov_cutoff 8 -very_clean yes ; cd k57 ; rm Graph2 PreGraph stats.txt ;
   cd .. ; velveth60 k55 55 -fastq -short ~/s_1_trimmed60cln.fq ;
   velvetg60 k55 -cov_cutoff 8 -very_clean yes ; cd k55 ; rm Graph2
   PreGraph stats.txt ; cd .. ; velveth60 k53 53 -fastq -short
   ~/s_1_trimmed60cln.fq ; velvetg60 k53 -cov_cutoff 8 -very_clean yes ;
   cd k53 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth60 k51 51 -fastq
   -short ~/s_1_trimmed60cln.fq ; velvetg60 k51 -cov_cutoff 8 -very_clean
   yes ; cd k51 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth60 k49 49
   -fastq -short ~/s_1_trimmed60cln.fq ; velvetg60 k49 -cov_cutoff 8 -
   very_clean yes ; cd k49 ; rm Graph2 PreGraph stats.txt ; cd .. ;
   velveth60 k47 47 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg60 k47 -
   cov_cutoff 8 -very_clean yes ; cd k47 ; rm Graph2 PreGraph stats.txt ;
   cd .. ; velveth60 k45 45 -fastq -short ~/s_1_trimmed60cln.fq ;
   velvetg60 k45 -cov_cutoff 8 -very_clean yes ; cd k45 ; rm Graph2
   PreGraph stats.txt ; cd .. ; velveth60 k43 43 -fastq -short
   ~/s_1_trimmed60cln.fq ; velvetg60 k43 -cov_cutoff 8 -very_clean yes ;
   cd k43 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth60 k41 41 -fastq
   -short ~/s_1_trimmed60cln.fq ; velvetg60 k41 -cov_cutoff 8 -very_clean
   yes ; cd k41 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth60 k39 39
   -fastq -short ~/s_1_trimmed60cln.fq ; velvetg60 k39 -cov_cutoff 8 -
   very_clean yes ; cd k39 ; rm Graph2 PreGraph stats.txt ; cd .. ;
   velveth60 k37 37 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg60 k37 -
   cov_cutoff 8 -very_clean yes ; cd k37 ; rm Graph2 PreGraph stats.txt ;
   cd .. ; velveth60 k35 35 -fastq -short ~/s_1_trimmed60cln.fq ;
```

```
velvetg60 k35 -cov_cutoff 8 -very_clean yes ; cd k35 ; rm Graph2
PreGraph stats.txt ; cd .. ; velveth60 k33 33 -fastq -short
~/s_1_trimmed60cln.fq ; velvetg60 k33 -cov_cutoff 8 -very_clean yes ;
cd k33 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth k31 31 -fastq -
short ~/s_1_trimmed60cln.fq ; velvetg k31 -cov_cutoff 8 -very_clean yes
; cd k31 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth k29 29 -fastq
-short ~/s_1_trimmed60cln.fq ; velvetg k29 -cov_cutoff 8 -very_clean
yes ; cd k29 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth k27 27 -
fastq -short ~/s_1_trimmed60cln.fq ; velvetg k27 -cov_cutoff 8 -
very_clean yes ; cd k27 ; rm Graph2 PreGraph stats.txt ; cd .. ;
velveth k25 25 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg k25 -
cov_cutoff 8 -very_clean yes ; cd k25 ; rm Graph2 PreGraph stats.txt ;
cd .. ; velveth k23 23 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg
k23 -cov_cutoff 8 -very_clean yes ; cd k23 ; rm Graph2 PreGraph
stats.txt ; cd .. ; velveth k21 21 -fastq -short ~/s_1_trimmed60cln.fq
; velvetg k21 -cov_cutoff 8 -very_clean yes ; cd k21 ; rm Graph2
PreGraph stats.txt ; cd .. ; velveth k19 19 -fastq -short
~/s_1_trimmed60cln.fq ; velvetg k19 -cov_cutoff 8 -very_clean yes ; cd
k19 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth k17 17 -fastq -
short ~/s_1_trimmed60cln.fq ; velvetg k17 -cov_cutoff 8 -very_clean yes
; cd k17 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth k15 15 -fastq
-short ~/s_1_trimmed60cln.fq ; velvetg k15 -cov_cutoff 8 -very_clean
yes ; cd k15 ; rm Graph2 PreGraph stats.txt ; cd .. ; velveth k13 13 -
fastq -short ~/s_1_trimmed60cln.fq ; velvetg k13 -cov_cutoff 8 -
very_clean yes ; cd k13 ; rm Graph2 PreGraph stats.txt ; cd .. ;
velveth k11 11 -fastq -short ~/s_1_trimmed60cln.fq ; velvetg k11 -
cov_cutoff 8 -very_clean yes ; cd k11 ; rm Graph2 PreGraph stats.txt ;
cd ..

$  cat k59/Log k57/Log k55/Log k53/Log k51/Log k49/Log k47/Log k45/Log
   k43/Log k41/Log k39/Log k37/Log k35/Log k33/Log k31/Log k29/Log k27/Log
   k25/Log k23/Log k21/Log k19/Log k17/Log k15/Log k13/Log k11/Log >
   MasterLog.txt

$  perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k59\n";}else{print
   "$_\n";}}' k59/contigs.fa > contigsAll.fa ; perl -e 'while(<>){chomp;
   if(/^\>(.+)$/){print "\>$1_k57\n";}else{print "$_\n";}}' k57/contigs.fa
   >> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
   "\>$1_k55\n";}else{print "$_\n";}}' k55/contigs.fa >> contigsAll.fa ;
   perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k53\n";}else{print
   "$_\n";}}' k53/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
   if(/^\>(.+)$/){print "\>$1_k51\n";}else{print "$_\n";}}' k51/contigs.fa
   >> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
   "\>$1_k49\n";}else{print "$_\n";}}' k49/contigs.fa >> contigsAll.fa ;
   perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k47\n";}else{print
   "$_\n";}}' k47/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
   if(/^\>(.+)$/){print "\>$1_k45\n";}else{print "$_\n";}}' k45/contigs.fa
   >> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
   "\>$1_k43\n";}else{print "$_\n";}}' k43/contigs.fa >> contigsAll.fa ;
   perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k41\n";}else{print
   "$_\n";}}' k41/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
   if(/^\>(.+)$/){print "\>$1_k39\n";}else{print "$_\n";}}' k39/contigs.fa
   >> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
   "\>$1_k37\n";}else{print "$_\n";}}' k37/contigs.fa >> contigsAll.fa ;
   perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k35\n";}else{print
```

```
"$_\n";}}' k35/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
if(/^\>(.+)$/){print "\>$1_k33\n";}else{print "$_\n";}}' k33/contigs.fa
>> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
"\>$1_k31\n";}else{print "$_\n";}}' k31/contigs.fa >> contigsAll.fa ;
perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k29\n";}else{print
"$_\n";}}' k29/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
if(/^\>(.+)$/){print "\>$1_k27\n";}else{print "$_\n";}}' k27/contigs.fa
>> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
"\>$1_k25\n";}else{print "$_\n";}}' k25/contigs.fa >> contigsAll.fa ;
perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k23\n";}else{print
"$_\n";}}' k23/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
if(/^\>(.+)$/){print "\>$1_k21\n";}else{print "$_\n";}}' k21/contigs.fa
>> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
"\>$1_k19\n";}else{print "$_\n";}}' k19/contigs.fa >> contigsAll.fa ;
perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k17\n";}else{print
"$_\n";}}' k17/contigs.fa >> contigsAll.fa ; perl -e 'while(<>){chomp;
if(/^\>(.+)$/){print "\>$1_k15\n";}else{print "$_\n";}}' k15/contigs.fa
>> contigsAll.fa ; perl -e 'while(<>){chomp; if(/^\>(.+)$/){print
"\>$1_k13\n";}else{print "$_\n";}}' k13/contigs.fa >> contigsAll.fa ;
perl -e 'while(<>){chomp; if(/^\>(.+)$/){print "\>$1_k11\n";}else{print
"$_\n";}}' k11/contigs.fa >> contigsAll.fa
```

Aedenovo

Aedes aegypti transcript assembly for validation

```
$  Aedenovo.pl queryfile=cquinxgenesunique.fa shortreads=s_5_sequence.txt
   trim=yes kmers=59,55,45,39,21 blastall=-p#blastn#-a#6#-e#0.1 cutoff=99

$  mv genebuild.txt genebuild1.txt

$  Aedenovo.pl queryfile=genebuild1.txt subject=contigs_5_All.fa
   blastdb=contigs_5_All.fa blastall=-p#blastn#-a#8#-e#0.01
```

Aedes triseriatus transcriptome assembly

```
$  Aedenovo.pl queryfile=Mosquito_TRANSCRIPTS subject=newnames100plus.fa
   formatdb=T blastall=-p#tblastx#-a#8#-e#1 cutoff=75 ; mv genebuild.txt
   genebuild1.txt

$  Aedenovo.pl queryfile=genebuild1.txt subject=newnames100plus.fa
   blastdb=newnames100plus.fa blastall=-p#blastn#-a#8#-e#0.1 cutoff=75
```

References

Surget-Groba, Y, and JI Montoya-Burgos. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* 20 (10):1432 - 1440.

ACKNOWLEDGMENTS

This work was made possible with grants from the Iowa State University Center for Integrated Animal Genomics and the Roy J. Carver Charitable Trust, and their generous contributions toward scientific research are always greatly appreciated.

I would like to thank my major professor, Dr. Lyric Bartholomay, for her guidance and support throughout my graduate education. She has been a great mentor and I have truly enjoyed working with her in the Iowa State University medical entomology laboratory. I would also like to acknowledge the other members of my graduate program of study committee, Drs. Bryony Bonning and Bradley Blitvich, for their helpful advice relating to my research and coursework.

Finally, I would like to thank my wife, Marissa, for her unending support and patience.