ABSTRACT

Genomic Analysis of the Diapause Program in the West Nile Virus Vector *Culex pipiens*

David S. Kang, Ph.D.

Mentor: Cheolho Sim, Ph.D.

The *Culex pipiens* complex of mosquitoes is a major vector for several pathogens responsible for infectious human diseases including West Nile virus in North America. With the growth of the American population, the prevalence of vector-borne diseases has become an increasing threat to human safety. This complex of mosquitoes includes the northern house mosquito *Culex pipiens* form pipiens, the underground mosquito *Culex pipiens* form molestus, and the southern house mosquito *Culex quinquefasciatus*. While each of these closely related mosquitoes share similar morphology, they all employ unique life strategies suited to their unique ecological niches, including overwintering diapause, host biting preference, ecological niche distribution and reproductive strategies. Despite these similarities and differences historically being obstacles to the control of these mosquitoes, new molecular technologies have enabled us to use these factors as a foundation for novel forms of vector control.

Diapause is an overwintering dormancy characterized by a suite of adaptations in response to seasonal changes. As survival mechanisms in anticipation of harsh winter conditions, these adaptations include alterations in the female *Culex pipiens* form pipiens

life cycle such as increased stress tolerance, increased nutrient allocation and alternative reproductive development. In *Culex pipiens* form pipiens, it is initiated by the shortened day lengths of fall.

One of the main obstacles in vector control of the *Culex pipiens* complex of mosquitoes is accurate identification of specimens, which relied on the historically unreliable measurement of the dorsal and ventral arms of the male phallosome (genetalia). Here we offer high resolution melting curve analysis which allows the use of fixed, single nucleotide polymorphisms (SNPs) to cheaply and quickly validate the identities of mosquito specimens.

Isolating and disruption of genes contributing to the life strategies of *Culex pipiens* complex mosquitoes is central to control of these vectors. Fortunately, the shared ancestry and divergent genetics of these mosquitoes provides an excellent opportunity for mapping studies, and consequently identification of candidate genes for disruption. The next chapters describe development of SNP markers for genetic mapping, and transcriptomic analysis utilizing RNAseq technologies to simultaneously identify and quantify expression differences between *Culex pipiens* biotypes and diapausing and nondiapausing specimens.

Genomic Analysis of the Diapause Program in the West Nile Virus Vector *Culex pipiens*

by

David S. Kang, B.S., M.S.

A Dissertation

Approved by the Department of Biology

_____

Robert D. Doyle, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

_____

Cheolho Sim, Ph.D., Chairperson

_____

Richard E. Duhrkopf, Ph.D.

_____

Bessie W. Kebaara, Ph.D.

_____

Myeongwoo Lee, Ph.D.

_____

Sung Joon Kim, Ph.D.

Accepted by the Graduate School

December 2015

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Cx. pipiens	*Culex pipiens*
ace-2	acetylcholinesterase-2
COI	cytochrome oxidase I
HRM	high-resolution melting
ND	non-diapause
D	diapause
DV/D ratio	dorsal-ventral/dorsal ratio
SNP	single nucleotide polymorphism

ACKNOWLEDGMENTS

I am deeply grateful to my friends and family for supporting me through my research and dissertation. I thank Cheolho Sim, Ph.D. for his guidance and for serving as a model of patience and balance during this hectic time. I thank my committee, Richard Durhrkopf, Ph.D., Bessie Kebaara, Ph.D., Myeongwoo Lee, Ph.D, and Sung Joon Kim, Ph.D. for their excellent advice on my research. Finally, I thank the current and past members of the Sim research group that facilitated this research.

DEDICATION

To my wonderful parents, John and Susan Kang, my sister, Esther Kang,
my loving wife, Julie Stauffer, and our fur babies

EPIGRAPH

"Bugs are gross." – Abraham Lincoln, Nov 19, 1863

CHAPTER ONE

Introduction

Overview of Culex pipiens *Complex Mosquitoes*

Comprised of *Culex pipiens* form pipiens, *Culex pipiens* form molestus, and *Culex quinquefasciatus* the *Culex pipiens* complex mosquitoes are vectors for globally significant viruses and pathogens that result in diseases such as West Nile encephalitis, Rift Valley Fever, Lymphatic filariasis and other human pathogens (Diamond 2009; Lai *et al.* 2000; Meegan *et al.* 1980; Monath 1988). Unfortunately, while they differ in regard to vector competence, geographical distribution, and behavioral/reproductive traits, similar morphology hinders accurate identification of *Culex pipiens* complex mosquitoes (Barr 1957; Clements 1992; Harbach *et al.* 1984; Spielman 1967). While these unique adaptations have profound implications in the transmission of human diseases, they may serve as the key to the control and suppression of these vectors.

Gonotrophic Cycle

Like most mosquitoes, *Culex pipiens* form pipiens and *Culex quinquefasciatus* require protein from a blood meal in order to produce eggs, while the autogenous *Culex pipiens* form molestus does not for initial oviposition. The cycle of blood feeding and egg development in *Culex pipiens* mosquitoes is known as a gonotrophic cycle, with adult females able to take multiple blood meals during a cycle and undergo multiple cycles in their life spans (Clements 1992; Vinogradova 2000a). In optimal conditions (23-25˚C)

Culex pipiens complex mosquitoes are capable of laying eggs after a little less than 3 days, and ovarian development is categorized into five Christophers stages based on follicle length. Consequently those eggs consequently develop to larvae, then to pupae to adults 11 days. In contrast, at lower temperatures (15-16˚C) egg production took over 8 days and an additional 36 days to develop fully into adults (Christophers 1911; Eldridge 1966; Jobling 1938).

Geographic Distribution

Given the importance of temperature on the life cycle of *Culex pipiens* mosquitoes, geographic distribution serves as an important factor in their capacity to transmit disease. *Culex pipiens* complex mosquitoes have a global distribution, and are found domestically in the continental United States. *Culex pipiens* form pipiens, also known as the northern house mosquito is found in the northern temperate regions. A sister species, the *Culex quinquefasciatus* or the southern house mosquitoes is found in southern, more tropical regions. A third biotype, *Culex pipiens* form molestus, also known as the underground mosquito is closely adjacent with the aboveground form pipiens, but is found in protected drainages and subways (Byrne & Nichols 1999; Harbach *et al.* 1984; Knight & Stone 1977; Vinogradova 2000a). While the taxonomy of these mosquitoes has long been in debate, it is clear that these mosquitoes are able to hybridize and are closely related (Amraoui *et al.* 2012; Fonseca *et al.* 2004).

Mosquito Identification

Historically, a morphometric approach has been used to identify mosquitoes. Unfortunately, many of the common taxonomic characteristics such as thorax patterning, wing shape and abdominal banding are conserved across the *Culex pipiens* complex, requiring biologists to largely rely on geographic location when identifying these mosquitoes (Clements 1992). Methods comparing the dorsal and ventral arms of the male phallosome (DV/D ratio), fell short in that they could not identify females of the species and that they have been reported to be difficult and unreliable (Barr 1957; Bourguet *et al.* 1998; Darsie *et al.* 1981; Sundararaman 1949; Wilton & Jakob 1985).

Life Strategies

Despite sharing similar physiology the unique characteristics of *Culex pipiens* complex mosquitoes greatly influence their life strategies and consequently, their behaviors as vectors of disease. The eurygamous (requiring large space for mating) northern *Culex pipiens* f. pipiens is adapted to exposure to the harsh northern climates, while the *Culex pipiens* f. molestus and *Culex quinquefacsiatus* are stenogamous and are capable of surviving and reproducing in small, manmade spaces, just as storm drains (Clements 1992; Spielman 1967; Vinogradova 2000b). Additionally, each of these species exhibit different blood meal preferences for the creation of eggs. *Culex pipiens* f. pipiens prefers avian blood, while *Culex quinquefasciatus* is more of a generalized blood feeder. In contrast, *Culex pipiens* f. molestus is autogenous and does not require a bloodmeal in order to product an initial batch of eggs (Clements 1992; Kent *et al.* 2007; Vinogradova 2000b). Furthermore, the heterodynamic diapausing behavior of *Culex*

pipiens f. pipiens allows for survival of harsh above ground exposure of winter, while the homodynamic *Culex pipiens* f. molestus relies on manmade subterranean infrastructure and the homodynamic *Culex quinquefasciatus* relies on mild southern temperatures (Byrne & Nichols 1999; Harbach *et al.* 1985; Harbach *et al.* 1984; Sirivanakarn & White 1978; Spielman & Wong 1973).

Diapause

It is crucial for animals to be able to detect changes in their environment and adapt accordingly. Nutrient availability, seasonal differences in photoperiod and temperature, moisture, and breeding site availability are all important factors to which organisms commonly respond. In particular, response to seasonal variation is an important factor in the success of many living organisms. One such adaptation, diapause is an anticipated, preprogrammed response to low temperatures and shortened photoperiods, characterized by cell cycle and developmental delay or arrest, suppressed metabolism, fat hypotrophy, and enhanced stress tolerance (Sim & Denlinger 2013).

Not to be confused with the temporary dormancy of quiescence, in female *Culex pipiens* f. pipiens, this alternative developmental pathway allows mosquitoes to enter a long term dormant state in which they may overwinter, and resume transmission of disease, such as West Nile Virus, during the following spring (Nasci *et al.* 2001). Due to this long-term dormancy, mosquitoes anticipating diapause delay reproductive development and blood feeding and rather prioritize the accumulation of nutrients. Altering the initiation or termination of diapause insects would significantly lower their

fitness. Thus, understanding diapause is critical for both the control of mosquito populations and preventing the proliferation of disease from year to year.

Environmental Initiation of Diapause

A key advantage of insect diapause is that it is an anticipatory rather than reactionary response. Thus, when harsh conditions arrive insects have already prepared for stressful conditions. In *Culex pipiens* mosquitoes, photoperiod is a crucial initiator of diapause, as it is the most reliable indicator of seasonal variation. Nonetheless, in many mosquitoes temperature acts to fine tune the intensity of the diapause response initiated by day length, indicating complex genetic interactions (Christophers 1911; Eldridge 1966; Jobling 1938). In other insects diapause may be initiated at any stage between embryo and adult, but in *Culex pipiens* complex mosquitoes photosensitivity persists from larval to adult development with the last larval instar being the most sensitive (Sanburg & Larsen 1973; Spielman & Wong 1973).

Endocrine Control of Diapause

The genes behind diapause are associated with endocrine effectors resulting in a myriad of adaptations such as enhanced lipid stores, reduced metabolism, increased innate immunity and resistance to the rigors of low temperatures (Denlinger 2002). Observation of metabolic pathways has implicated insulin peptides as a major upstream regulator of diapause. In *Culex pipiens* f. pipiens, the forkhead transcription factor or (FOXO) acts as a developmental switch, initiating or priming the production of insulin by the corpora allata, which controls molting and vitellogenesis in addition to diapause

(Hahn & Denlinger 2007; Ragland *et al.* 2010; Sim & Denlinger 2008, 2013; Tatar *et al.* 2001; Williams *et al.* 2010). While much is known about diapause, the full basis behind the syndrome has yet to be fully examined from genetic and molecular perspectives. It is our hope that elucidating the genetic and molecular basis behind diapause will serve as a platform for the development of new methods of vector control.

Thesis Overview

It is clear the eco-physiological differences between *Culex pipiens* complex biotypes is rooted in genetic architecture, and that these differences may be exploited by vector biologists in the control of disease transmission. In particular, the environmental cues that initiate diapause in *Culex pipiens* complex mosquitoes is well documented, but the mechanisms behind the diapause syndrome are not yet completely understood at a molecular level.

The obvious first step in vector control is correct identification of mosquito biotypes. As simple as this seems the classic morphometric approach is difficult to employ and yields unreliable data at best. In the following chapter (Chapter Two), "Identification of *Culex* complex species using SNP markers based on high resolution melting analysis" I utilize melting curve technology to classify the *Culex pipiens* form pipiens, *Culex pipiens* form molestus, and *Culex quinquefasciatus* mosquitoes based on single nucleotide polymorphisms of the acetylcholinesterase-2 (*ace-2)* locus.

In Chapter Three, entitled "Identification and assessment of single nucleotide polymorphisms (SNPs) between *Culex* complex mosquitoes" we develop an array of 28 SNP markers for mapping studies. It is our expectation that these markers will improve

the resolution of existing microsatellite marker sets, and will consequently employed in quantitative trait loci (QTL) studies, which in turn will narrow down candidate genes contributing to traits unique to *Culex pipiens* complex biotypes.

Chapter Four, entitled "A comparative examination of key expression differences between the *Culex pipiens* form pipiens and *Culex pipiens* form molestus biotypes" further examines expression profile differences between females via Illumina RNAseq, and then sorts candidate genes by statistical significance, fold change differences, and biological relevance. Chapter Five, entitled "A transcriptomic approach to identification of candidate genes for vector control of *Culex pipiens* mosquitoes" further narrows down screening of expression profiles by comparing diapausing and nondiapausing female mosquitoes of the same *Culex pipiens* form pipiens biotype.

The goal of this dissertation is further the understanding of gene expression affects the many divergent traits of *Culex pipiens* complex mosquitoes, and resulted in the identification of several candidate genes for the manipulation of their life strategies, providing a strong launching point for further functional study of these genes.

This dissertation and the experiments herein was designed, performed and written by David S. Kang under the supervision of Cheolho Sim, Ph.D. who also helped edit this manuscript.

CHAPTER TWO

Identification of *Culex* complex species using SNP markers based on high resolution melting analysis

This chapter published as Kang D, Sim C (2013) Identification of *Culex* complex species using SNP markers based on high-resolution melting analysis. *Molecular Ecology Resources*. 13: 369-376.

Abstract

Mosquitoes belonging to the *Culex pipiens* complex are primary vectors for diseases such as West Nile encephalitis, Eastern equine encephalitis, many arboviruses, as well as lymphatic filariases. Despite sharing physiological characteristics, each mosquito species within the *Culex* complex has unique behavioral and reproductive traits that necessitate a proper method of identification. Unfortunately, morphometric methods of distinguishing members of this complex have failed to yield consistent results, giving rise to the need for molecular methods of identification. In this study we propose a novel identification method using high-resolution melting analysis by examining single nucleotide polymorphisms in the acetylcholinesterase-2 (*ace-2)* locus. Our method provides a high confidence for species determination among the three *Culex* complex mosquitoes.

Introduction

Culex pipiens complex mosquitoes are the primary vectors for West Nile encephalitis, Eastern equine encephalitis, lymphatic filariasis, as well as many arboviruses (Diamond 2009; Hubalek & Halouzka 1999; Lai et al. 2000; Lanciotti et al. 2000; Margulies 2004; Monath 1988). While these mosquitoes share similar morphology, divergent behavioral,and reproductive traits result in limited geographical distribution or unique ecological niches (Barr 1957; Clements 1992; Harbach et al. 1984; Spielman 1967).

C. pipiens f. pipiens is eurygamous (requires a mating flight before copulation) whereas C. pipiens f. molestus and C. quinquefasciatus are stenogamous (capable of mating in small spaces such as sewer drains) (Clements 1992; Kent et al. 2007; Roth 1948; Spielman 1967; Vinogradova 2000). In addition, these three species have varied blood meal preferences for the process of vitellogenesis. C. pipiens f. pipiens and C. quinquefasciatus are both anautogenous (require a blood meal to produce eggs), yet C. pipiens f. pipiens prefers avian blood hosts while the C. quinquefasciatus has been documented as generalized blood feeders that prey on both mammals and avian hosts (Clements 1992; Kent et al. 2007; Molaei et al. 2007; Spielman 1967; Vinogradova 2000). In contrast C. pipiens f. molestus is autogenous and does not require a blood meal to produce their initial batch of offspring (Kent et al. 2007). C. pipiens f. pipiens undergoes an adult reproductive diapause to overwinter while both C. quinquefasciatus and C. pipiens f. molestus are homodynamic (nondiapausing) despite the aboveground dwelling f. pipiens and the belowground dwelling f. molestus being considered sympatric while the C. quinquefasciatus is geographically remote (Bourguet et al. 1998; Clements

1992; Kent et al. 2007; Roth 1948; Sim & Denlinger 2008; Vinogradova 2000). The implications of these diverse behavioral differences in the transmission of disease and host-shifting are profound. From these ecological and life history differences arises the need for methods to correctly identify these mosquitoes before an accurate assessment can be made on their roles as vectors. Unfortunately, despite their unique life-history traits, the morphological differences are very limited and species identification remains difficult, which complicates attempts to study and subsequently control these mosquitoes.

Early attempts to identify the different Culex pipiens complex mosquitoes have been based on morphological methods. Unfortunately, many of the taxonomic characteristics describing these mosquitoes such as thorax markings, wing shape, and abdominal banding are shared among Culex pipiens complex species, forcing researchers to posit identity largely based on geographic location (Clements 1992; National Communicable Disease Center (U.S.) & Communicable Disease Center (U.S.). 1969). The advent of the DV/D ratio method, a comparison of the dorsal and ventral arms of the male phallosome, seemed to be a breakthrough in discerning between C. pipiens f. pipiens and C. quinquefasciatus. However, this method could only identify the non-blood feeding males and has proven to be a difficult and inconsistent diagnostic tool(Barr 1957; Bourguet et al. 1998; Smith & Fonseca 2004; Sundararaman 1949; Wilton & Jakob 1985).

Recently, molecular biology has provided new tools to identify species (Hebert et al. 2003). Early attempts at utilizing restriction fragment length polymorphisms of cytochrome oxidase I (COI) and actylcholinesterase (ACE) provided inaccurate results, yet recent advances in molecular techniques have provided reliable methods of

identification (Bourguet et al. 1998; Kothera et al. 2010). Researchers have attempted to create rapid, cost effective methods of genotype identification using the polymerase chain reaction (PCR) but have been met with limited success, yielding ambiguous results that are susceptible to the slightest changes in reaction conditions (Bahnck & Fonseca 2006; Kothera et al. 2010; McAbee et al. 2008; Smith & Fonseca 2004). Attempts to use traditional barcoding to distinguish between Culex pipiens f. molestus and Culex pipiens f. pipiens via mtDNA (small subunit ribosomal 12S, CO1, and Nd4) exhibited no genetic differences, possibly because lineages have not had enough time to gain unique genetic identity (Kent et al. 2007). Microsatellite markers have quickly become the gold standard in genotyping, proving to be highly accurate at a population level, but requiring the use of an experienced sequencing core which can be time consuming, expensive, and difficult to interpret (Ewing et al. 1998).

Our findings support that high-resolution melting (HRM) analysis on a real-time PCR platform provides a rapid, economic alternative to other molecular assays with a higher level of sensitivity and specificity (Heid et al. 1996; Wittwer et al. 2003). We have developed an HRM assay based on the sex-linked ace-2 gene utilizing primers at the second and third exon, spanning across the second intron (Malcolm et al. 1998; Smith & Fonseca 2004). We then assayed 20 female C. pipiens f. pipiens, 20 female C. pipiens f. molestus, and 20 female C. quinquefasciatus with DNA sequencing confirmation of SNP identities.

Materials and Methods

Mosquito Collection and Rearing

The stock colony of C. pipiens form pipiens (Buckeye Strain) was provided from Dr. David L. Denlinger's lab at the Ohio State University. The colony was established in September of 2000 from larvae collected in Columbus, Ohio. Larvae are reared in de-chlorinated tap water and fed on Tetramin fish food (Tetra holding (US) Inc., Blacksburg, VA, USA). Adults are maintained on honey sponges and kept in large screened cages. Culex pipiens f. pipiens are reared at 25 oC, 75% relative humidity, with a 15 hour light: 9 hour dark (L:D) daily light cycle.

The stock colony of C. pipiens form molestus was provided by Dr. Linda Kothera's lab in the Centers for Disease Control and Prevention Division of Vector-Borne Infectious Diseases at Fort Collins, Colorado. The Culex pipiens f. molestus colony was established from belowground individuals from the Calumet Water Reclamation Plant in Chicago (Mutebi & Savage 2009). Similar to Culex pipiens f. pipiens, Culex pipiens f. molestus specimens are reared at 25°C and 75% relative humidity under a 15 h light: 9 h dark (L:D) photoperiod.

Individual field samples of Culex quinquefasciatus were trapped from September 21, 2011 to July 1, 2012, at sites in the greater Waco, TX area. Alfalfa infused gravity traps were placed at sites during the late afternoon and retrieved the following morning. Date, time, humidity, temperature, wind speed, dew point, and global positioning system (GPS) coordinates were recorded upon collection. Adult mosquitoes were preserved in 95% ethanol and kept chilled until identification on a stereomicroscope (Darsie et al. 1981; Sundararaman 1949).

Morphological Identification

Mosquitoes were next keyed to identification as members of the Culex complex. Specimens were examined for standard identifiers such as U-shaped markings on the thorax, a clear wing with a palp shorter than the proboscis, and an M-shaped banding on the dorsal abdomen (Barr 1957; Bourguet et al. 1998; Darsie et al. 1981; Smith & Fonseca 2004; Sundararaman 1949; Wilton & Jakob 1985). Mosquitoes were then subjected to DV/D analysis as described by Sundararaman in which the distance of tips of dorsal arms (D) of the male phallosomes are compared to the distance between the tips of the ventral and dorsal arms of the male phallosome. Literature suggests that mosquitoes possessing a ratio of less than 0.2 are C. pipiens form pipiens, mosquitoes with a ratio greater than 0.4 are C. quinquefasciatus and those with a ratio ranging from 0.2-0.4 are a hybrid of the two forms (Sundararaman 1949).

SNP Selection

Based on C. quinquefasciatus sequence data for the acetylcholinesterase-2 (ace-2, GenBank accession no. AY196911) gene obtained from Vectorbase (http://cquinquefasciatus.vectorbase.org/), the Primer3 tool (http://frodo.wi.mit.edu/primer3/) was used to design an amplicon containing intron 2 flanked by primers at exon 2 (F 5'-TTTGGGTACCAACGAAGACC-3') and exon 3 (R 5'-AAGTCCTCATCGTCCTGGAA-3'). Genomic DNA was extracted from five females of each of the investigated forms, C. quinquefasciatus, C. pipiens f. pipiens, and C. pipiens f. molestus, with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. After measuring purity against the elution buffer

on a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) at 260/280 nm samples we amplified the intron 2 via polymerase chain reactions (PCR) using the Taq PCR Kit (New England BioLabs, Ipswich, MA). 50 µl PCR amplifications were assayed utilizing 200 µM dNTPs, 10 x PCR buffer, 2 mM MgSO4, 10 mM primers and 10 ng of gDNA. Amplifications were then performed on a T100 thermal cycler (Bio-Rad, Hercules, California) with an initial denaturation at 94°C for 2 minutes, followed by 40 cycles at 94°C for 15 seconds, 42°C for 15 seconds, 72° for 15 seconds, then one cycle of 72°C for 5 minutes. PCR products and a negative control were then visualized on 1.5% agarose with a 100 bp molecular weight ladder (Invitrogen, Carlsbad, CA). Amplified PCR products were purified using ExoSAP-IT (GE Healthcare, Little Chalfont, United Kingdom), removing unincorporated primers, nucleotides and salts, before being submitted to the DNA Analysis Facility at Yale University for DNA sequencing.

Sequences were aligned and analyzed using CLC Main Workbench 6 (CLC bio, Aarhus, Denmark). Candidate SNP positions were examined for fixation within members of the Culex pipiens complex and variability between members of the different forms. SNP data was submitted to the NCBI SNP database (dbSNP accession no. ss539004799).

Quantitative PCR and High-Resolution-Melting Analysis

A primer set (F 5'-AATCGCTTATTGGTTCTTCG-3', R 5'-GCACAGCTCGACAGAAAA-3') was created based on consensus regions between the different Culex pipiens complex members with a 76 bp amplicon with unique, form-specific melting curves in order to ensure maximum resolution in HRM analysis. Initial

optimization utilized PCR amplification and then standard resolution melting curves to confirm calculated Tm and assess appropriate reagent concentrations and conditions for standardized amplification of test samples.

Genomic DNA from 15 positive control mosquitoes previously identified via sequencing, 45 known mosquitoes and blank negative controls were then subjected to pre-amplification and high-resolution-melting (HRM) analysis utilizing the Type-IT HRM PCR Kit (Qiagen, Hilden, Germany), with the saturating dye Eva Green, on the Rotor-Gene Q real-time thermal cycler (Qiagen, Hilden, Germany). Pre-amplification was performed with an initial denaturation at 95°C for 5 minutes, followed by 40 cycles at 95°C for 10 seconds, 51°C for 30 seconds, and 72° for 10 seconds with 2 x Type-IT HRM PCR Master Mix, 10 mM forward and reverse HRM primers and 10 ng of gDNA. After a 90 second pre-melt hold dissociation was performed by ramping temperatures at 0.1°C increments every 2 seconds from 65°C to 95°C. Changes in Eva Green fluorescence were monitored at the green channel with 470±10 nm excitation and 510±5 nm emission. Fluorescence normalization was then performed by the Rotor-Gene software which then automatically assigned genotypes to specimens with a minimum of 95% confidence against sequenced positive controls.

Results

Morphological Identification

Previous studies suggest members of the genus Culex can be keyed by U-shaped markings on the thorax, a clear wing with a palp shorter than the proboscis, and an M-shaped banding on the dorsal abdomen (Darsie et al., 1981). Unfortunately, as current literature suggests, upon examination of our specimens we found that these characteristics were not unique to a specific form or were difficult to distinguish across the three Culex pipiens complex mosquitoes that we examined (Fig 1) (Barr 1957; Bourguet et al. 1998; Smith & Fonseca 2004; Sundararaman 1949; Wilton & Jakob 1985).

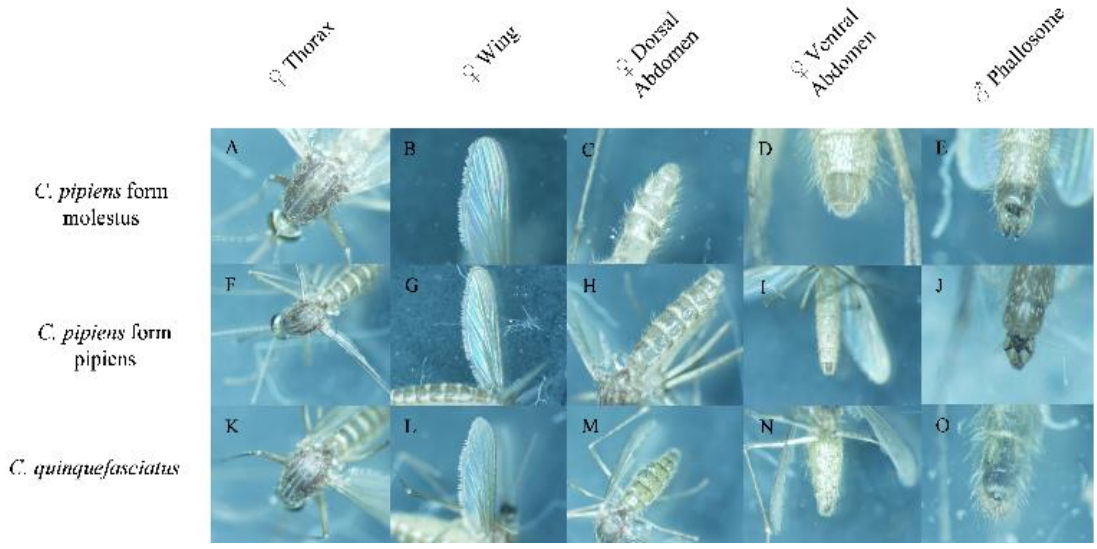

Figure 1. Morphological comparisons between adult females of the *Culex* complex. Classic morphological markers associated with the identification of *Culex* complex mosquitoes (u-shaped thorax markings, wing fringing, abdominal banding patterns, and genitalia) prove unreliable within three *Culex* complex species. **(A-D)** *C. pipiens* f. molestus, **(E-H)** *C. pipiens* f. pipiens, and **(I-L)** *C. quinquefasciatus.*

We next measured the DV/D ratio of the male phallosome as described by Sundararaman in the identification of C. quinquefasciatus and C. pipiens f. pipiens, and

found that with the addition of C. pipiens f. molestus the method became inconsistent (Sundararaman 1949). Whether the specimens were lab raised or trapped from the wild they rarely conformed to the DV/D ratios specified by literature of less than 0.2 for C. pipiens f. pipiens, and over 0.4 for C. quinquefasciatus with their hybrids ranging between 0.2-0.4. Interestingly, C. pipiens f. molestus exhibited ratios spanning across both ranges as well as that of the intermediate hybrid form confounding the results of this classic morphometric method (Fig 1. E, J, O).

PCR Amplification and Sequencing

We investigated the possibility of a common HRM primer set able to differentiate between Culex pipiens f. pipiens, Culex pipiens f. molestus, and Culex quinquefasciatus. The intron 2 of the ace-2 gene was amplified from gDNA extracted from complex members and sequenced. We found that the primers based on the Exon 2 and 3 of ace-2 gene (Fig 2. A), had reliably amplified the 565 bp PCR products of the different Culex complex members, with each sample exhibiting strong bands when visualized by 1.5 % agarose electrophoresis (Fig 2. B). The sequencing revealed only a single class IV (A/T) transversion between C. pipiens f. pipiens and C. quinquefasciatus, but a high level of polymorphism when either is compared to C. pipiens f. molestus (Fig 3).

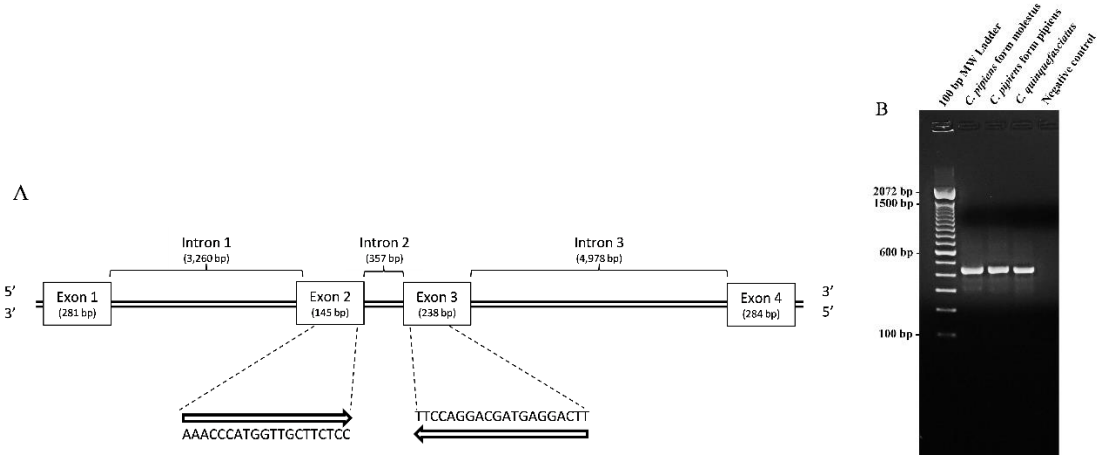

Figure 2. PCR amplification and visualization of product with Exon2/Exon3 *ace-2* primers. (A) Orientation of common forward and reverse primers relative to gDNA. **(B)** Visualization of PCR products through 1.5% agarose gel electrophoresis against a 100 bp ladder. Presence of common amplicon size confirms conservation of the *ace-2 gene* across the *Culex* complex.

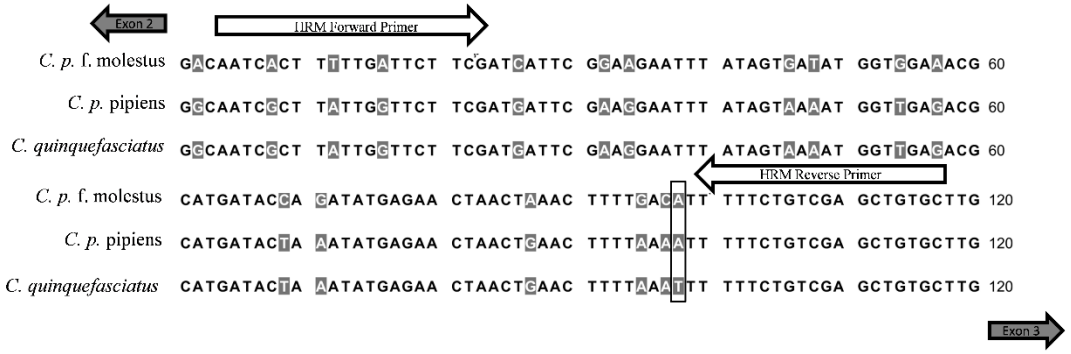

Figure 3. Single nucleotide polymorphisms (SNPs) across the *Culex* complex. The High Resolution Melting (HRM) forward primer orients at the 3' end of *ace-2* exon 2 and HRM reverse primer orients at the 5' end of exon 3. Polymorphic sites are highlighted (gray). The *C. pipiens* f. molestus diverges greatly from *C. pipiens* f. pipiens and *C. quinquefasciatus* which in turn differ by a single nucleotide polymorphism (rectangle).

Genotyping

Our method was able to correctly identify 58 out of 60 specimens with a minimum threshold of 95% confidence in comparison to sequencing results. We found that using standard melting curves on qPCR machines used in initial optimization we could differentiate between C. pipiens. f. molestus and C. quinquefasciatus or C. pipiens f. pipiens but were unable to discriminate between C. quinquefasciatus and C. pipiens f. pipiens due a genetic deviation of only a single adenine-thymine transversion (Fig. 3). In contrast, HRM analysis is capable of easily detecting this subtle difference as a 0.2 degree shift in melting temperature. Moreover HRM analysis clearly exhibits unique melting temperature (Tm) for each sample with non-overlapping, separate inflection points for each Culex pipiens complex species ranging between 74.4-75.4°C (Fig. 4), supporting our hypothesis that HRM analysis may be used to distinguish between the three groups.

Discussion

Using morphological features to identify Culex pipiens complex mosquitoes may lead to misidentification, which in turn can hinder efforts to monitor the behavior of these mosquitoes and to evaluate the threat they pose of spreading diseases. Our attempt to distinguish members of the Culex pipiens complex using classical morphometric analysis proved to be inconsistent, which combined with literature suggests that DV/D ratio is

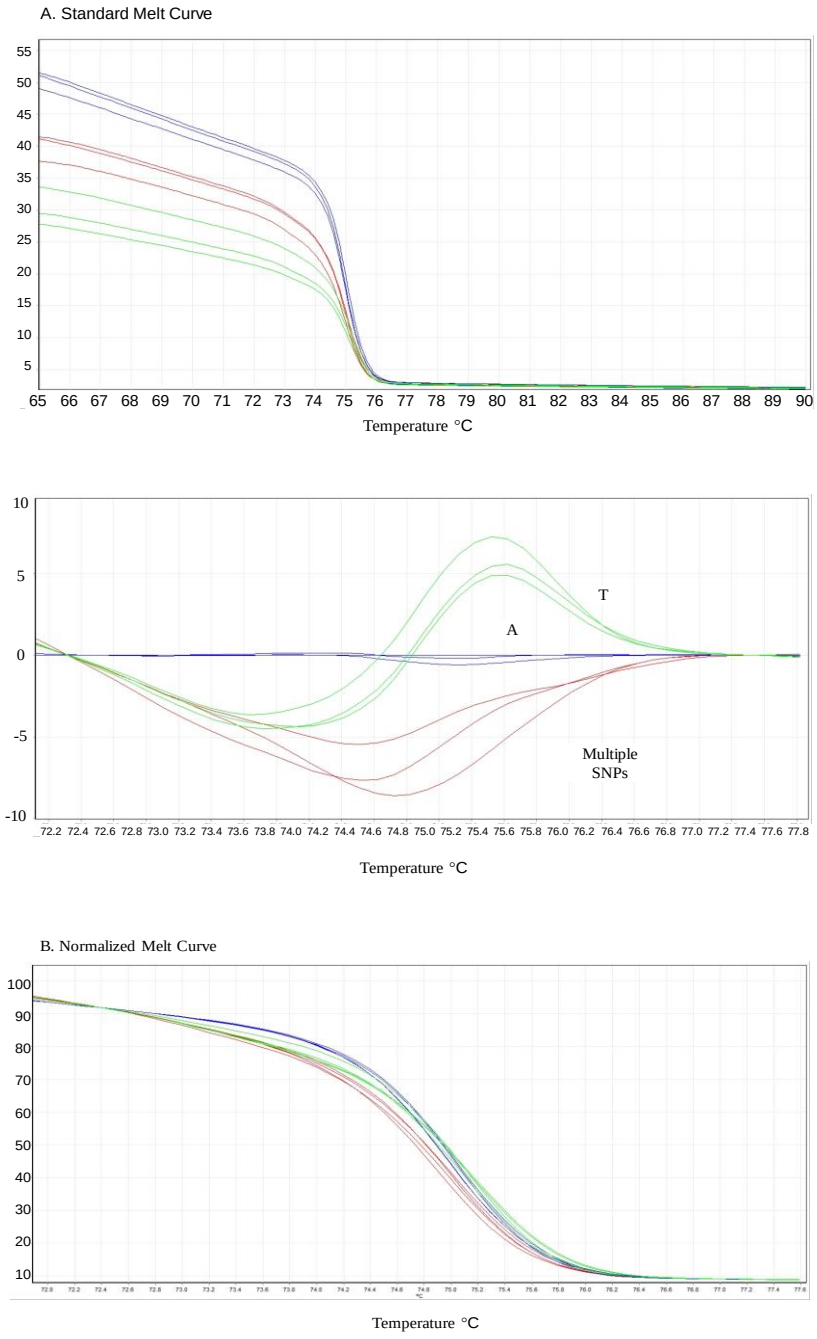

Figure 4. Discriminations of SNPs among three *Culex* complex species by using High Resolution Melting analysis. **(A)** Raw melting profile of real-time quantitative PCR showing dissociation of double stranded DNA into single strands as temperature is incrementally increased. **(B)** Normalization of fluorescence from the raw melting data for further discrimination in a difference plot. **(C)** The difference plot amplifies minute differences between the melting curves, thus genotype calling is dependent on similarities in melting behavior at different temperatures rather than overlapping melt profiles. *C. quinquefasciatus* are represented by green and *C. pipiens* f. molestus and *Culex pipiens* f. pipiens by red and blue, respectively.

effected by rearing temperature, emphasizes the need for molecular methods of mosquito identification (Barr 1957; Smith & Fonseca 2004; Sundararaman 1949; Wilton & Jakob 1985). With the belowground C. pipiens f. molestus rarely hybridizing with aboveground C. pipiens f. pipiens, it seems unlikely that the inconclusive results from examination of the male phallosomes reported in literature is due to cross breeding between the two populations, yet warrants future morphometric investigation of the hybrid forms (Kothera et al. 2010).

Our HRM assay was designed to include C. pipiens f. molestus because it can easily be mistaken for other forms of Culex, particularly the above ground Culex pipiens f. pipiens, despite having the unique properties of being stenogamous, autogenous, and homodynamic (Kent et al. 2007; Vinogradova 2000). HRM analysis will help in the proper identification of these vectors which allows us to accurately conduct population studies. In turn, by examining mosquito gene flow, migration and temporal variations, population genetics may be used to predict and prevent epidemics in human populations. Indeed, many scientists have already attempted to address the need for a quick, reliable mosquito identification assay. Unfortunately, each of the previously proposed methods was hindered by limiting parameters and inconsistent results. Methods such as microsatellite analysis, while being highly accurate, involve a difficult, time consuming method requiring access to a sequencer (Smith & Fonseca 2004). Past attempts to create a diagnostic restriction fragment length polymorphism digest or barcoding of the mtDNA loci have proven inaccurate in the field (Kent et al. 2007; Kothera et al. 2010).

Recent molecular studies allowed us to focus on the introns of the sex-linked ace-2 gene (Bourguet et al. 1998; Diaz-Badillo et al. 2011; Kasai et al. 2008; McAbee et al.

2008; Smith & Fonseca 2004). At 23,166 bp the ace-2 gene spans 10 exons and has a transcript length of approximately 1,908 bp (Gene ID: CPIJ802216). The ability of the PCR primers to efficiently anneal with exon 2 and exon 3 of the ace-2 gene across the different members of the Culex pipiens complex indicates that this region is strongly conserved, and warrants that the nested intron is a good candidate locus for SNP discovery. Further analysis via sequencing revealed that the lab strain C. pipiens f. pipiens and the wild C. quinquefasciatus seemed to have very little variability despite the differences in their species, geography, and distribution. Interestingly, the sequence data from the Chicago strain of C. pipiens f. molestus seemed to diverge greatly from both C. pipiens f. pipiens and C. quinquefasciatus. This is rather odd considering the prevailing hypothesis that below ground C. pipiens f. molestus originated from local above ground populations of C. pipiens f. pipiens of Chicago, Illinois which is geographically proximal to Columbus, Ohio in comparison to Waco, Texas (Kent et al. 2007; Kothera et al. 2010).

There have been many PCR assays developed that focused on the same ace-2 gene as our proposed HRM assay. The original ace-2 assays were able to identify many different species of the Culex pipiens complex ranging around the world, but yielded varied results depending on slight changes in reaction conditions and presented vague results for hybrids (McAbee et al. 2008; Smith & Fonseca 2004). Notably, one multiplex PCR assay utilized single nucleotide differences located in intron 2 of the ace-2 gene and was designed to identify C. pipiens f. molestus, C. quinquefasciatus and C. pipiens f. pipiens, but have been reported to incorrectly classify 14.3-21.3% of aboveground mosquitoes as f. molestus (Bahnck & Fonseca 2006; Kothera et al. 2010).

While sequencing is still considered the most accurate method for genotyping, high-resolution melting curve analysis offers a quick, cost effective method for identifying mosquitoes that are difficult to distinguish using classic morphological markers (Ewing et al. 1998). Competitively priced at about a third the cost of other similar SNP genoptying methods, the possibility of human error is mitigated in HRM by the availability of run ready pre-mixes and the convenience of having both the amplification and the melting analysis preformed in one instrument without manipulations between stages, which reduces the possibility of contamination. In addition, the products from HRM may be sent directly for sequencing for verification of results, and has the benefit of amplifying the amount of sample available rather than destroying it as techniques such as restriction fragment analysis would (Reed et al. 2007; Reed & Wittwer 2004). It is our hope that after initial optimization that our HRM assay can modified for other qPCR platforms.

With the advent of new saturating dyes, such as Eva Green®, dye concentrations can be increased without concerns of interfering with polymerases, which in turn prevents reincorporation of dyes released during the melting process into the un-melted regions (Mao et al. 2007). This has led to higher levels of accuracy when genotyping with HRM which allowed us to identify, target and differentiate class IV (A/T) SNPs. The rarest of the SNPs, class IVs were previously all but impossible to distinguish via HRM due to the minute differences in melting curves and the unreliability of non-saturating dyes at such high resolution (Venter et al. 2001). The rarity of class IV SNPs led us to select one to differentiate between C. pipiens f. pipiens and C. quinquefasciatus,

which minimizes the chance of observed polymorphisms being the result of new mutations.

We suggest that HRM is uniquely suited to analyze SNPs at a high resolution as an alternative to traditional PCR The assay that we have developed for correctly identifying members of the Culex pipiens complex will improve the understanding of distribution and interactions of these mosquitoes, which in turn will assist in the control and prevention of disease transmission.

References

Bahnck CM, Fonseca DM (2006) Rapid assay to identify the two genetic forms of Culex (Culex) pipiens L. (Diptera: Culicidae) and hybrid populations. Am J Trop Med Hyg 75, 251-255.

Barr AR (1957) The distribution of Culex p. pipiens and C.P. quinquefasciatus in North America. Am J Trop Med Hyg 6, 153-165.

Bourguet D, Fonseca D, Vourch G, et al. (1998) The acetylcholinesterase gene Ace: a diagnostic marker for the Pipiens and Quinquefasciatus forms of the Culex pipiens complex. J Am Mosq Control Assoc 14, 390-396.

Clements AN (1992) The biology of mosquitoes, 1st edn. Chapman & Hall, London ; New York.

Darsie RF, Ward RA, Chang CC (1981) Identification and geographical distribution of the mosquitoes of North America, north of Mexico American Mosquito Control Association, Fresno, Calif.

Diamond MS (2009) West Nile encephalitis virus infection : viral pathogenesis and the host immune response Springer, New York, NY.

Diaz-Badillo A, Bolling BG, Perez-Ramirez G, et al. (2011) The distribution of potential West Nile virus vectors, Culex pipiens pipiens and Culex pipiens quinquefasciatus (Diptera: Culicidae), in Mexico City. Parasit Vectors 4, 70.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175-185.

Harbach RE, Harrison BA, Gad AM (1984) Culex (Culex) Molestus Forskal (Diptera, Culicidae) - Neotype Designation, Description, Variation, and Taxonomic Status. Proceedings of the Entomological Society of Washington 86, 521-542.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B-Biological Sciences 270, 313-321.

Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. Genome Research 6, 986-994.

Hubalek Z, Halouzka J (1999) West Nile fever--a reemerging mosquito-borne viral disease in Europe. Emerg Infect Dis 5, 643-650.

Kasai S, Komagata O, Tomita T, et al. (2008) PCR-based identification of Culex pipiens complex collected in Japan. Jpn J Infect Dis 61, 184-191.

Kent RJ, Harrington LC, Norris DE (2007) Genetic differences between Culex pipiens f. molestus and Culex pipiens pipiens (Diptera: Culicidae) in New York. J Med Entomol 44, 50-59.

Kothera L, Godsey M, Mutebi JP, Savage HM (2010) A comparison of aboveground and belowground populations of Culex pipiens (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. J Med Entomol 47, 805-813.

Lai CH, Tung KC, Ooi HK, Wang JS (2000) Competence of Aedes albopictus and Culex quinquefasciatus as vector of Dirofilaria immitis after blood meal with different microfilarial density. Vet Parasitol 90, 231-237.

Lanciotti RS, Kerst AJ, Nasci RS, et al. (2000) Rapid detection of west nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. J Clin Microbiol 38, 4066-4071.

Malcolm CA, Bourguet D, Ascolillo A, et al. (1998) A sex-linked Ace gene, not linked to insensitive acetylcholinesterase-mediated insecticide resistance in Culex pipiens. Insect Mol Biol 7, 107-120.

Mao F, Leung WY, Xin X (2007) Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. BMC Biotechnol 7, 76.

Margulies P (2004) West Nile virus, 1st edn. Rosen Pub. Group, New York.

McAbee RD, Green EN, Holeman J, et al. (2008) Identification of Culex pipiens complex mosquitoes in a hybrid zone of West Nile virus transmission in Fresno County, California. Am J Trop Med Hyg 78, 303-310.

Molaei G, Andreadis TG, Armstrong PM, et al. (2007) Host feeding pattern of Culex quinquefasciatus (Diptera: Culicidae) and its role in transmission of West Nile virus in Harris County, Texas. Am J Trop Med Hyg 77, 73-81.

Monath TP (1988) The Arboviruses : epidemiology and ecology CRC Press, Boca Raton, Fla.

Mutebi JP, Savage HM (2009) Discovery of Culex pipiens pipiens form molestus in Chicago. J Am Mosq Control Assoc 25, 500-503.

National Communicable Disease Center (U.S.), Communicable Disease Center (U.S.). (1969) Pictorial keys to arthropods, reptiles, birds, and mammals of public health significance For sale by the Supt. of Docs., U.S. Govt. Print. Off., Washington, Atlanta, Ga.,.

Reed GH, Kent JO, Wittwer CT (2007) High-resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics 8, 597-608.

Reed GH, Wittwer CT (2004) Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clin Chem 50, 1748-1754.

Roth LM (1948) A Study of Mosquito Behavior. An Experimental Laboratory Study of the Sexual Behavior of Aedes aegypti (Linnaeus). American Midland Naturalist 40, 265-352.

Sim C, Denlinger DL (2008) Insulin signaling and FOXO regulate the overwintering diapause of the mosquito Culex pipiens. Proc Natl Acad Sci U S A 105, 6777-6781.

Smith JL, Fonseca DM (2004) Rapid assays for identification of members of the Culex (Culex) pipiens complex, their hybrids, and other sibling species (Diptera: culicidae). Am J Trop Med Hyg 70, 339-345.

Spielman A (1967) Population structure in the Culex pipiens complex of mosquitos. Bull World Health Organ 37, 271-276.

Sundararaman S (1949) Biometrical studies on intergradation in the genitalia of certain populations of Culex pipiens and Culex quinquefasciatus in the United States. Am J Hyg 50, 307-314.

Venter JC, Adams MD, Myers EW, et al. (2001) The sequence of the human genome. Science 291, 1304-+.

Vinogradova EB (2000) Culex pipiens pipiens mosquitoes : taxonomy, distribution, ecology, physiology, genetic, applied importance and control Pensoft, Sofia.

Wilton DP, Jakob WL (1985) Temperature-induced morphological change in Culex pipiens. J Am Mosq Control Assoc 1, 174-177.

Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. Clin Chem 49, 853-860.

Data Accessibility

NCBI SNP Database (dbSNP accession no. ss539004799)

CHAPTER THREE

Identification and assessment of single nucleotide polymorphisms (SNPs) between *Culex* complex mosquitoes.

This chapter published as Kang DS, Sim C (2015) Identification and assessment of single nucleotide polymorphisms (SNPs) between *Culex* complex mosquitoes. *Molecular Ecology Resources.* 15 (1); 228, 13 pages.

Abstract

Culex pipiens complex mosquitoes are important vectors for many human pathogens including West Nile encephalitis, Rift Valley fever and Lymphatic filariasis. In this study we characterize a set of SNP markers between two biotypes of the *Culex pipiens* complex, *Culex pipiens* form molestus and *Culex pipiens* form pipiens, for use in a high-resolution genetic mapping and a population genetics study.

DNA pooled from 10 specimens of each biotype were sequenced and analyzed for variation in 28 genes. The total of 4714 bp across orthologs of *C. pipiens* form pipiens and *C. pipiens* form molestus revealed 44 SNPs in the 3279 bp coding regions, 48 SNPs in the 1435 bp non-coding regions, and 10 indels. The ratio of transitions to transversions approached 2:1, with transitions constituting the majority of synonymous coding substitutions.

The informative SNP markers were successfully identified and assessed from both *C. pipiens* biotypes. We expect that novel SNPs characterized in this study would be useful for genetic studies to elucidate the genetic basis of diverged eco-physiological traits between the two biotypes of the *C. pipiens* complex.

Introduction

Culex pipiens complex mosquitoes are vectors for many human pathogens such as West Nile encephalitis, Rift Valley Fever, and Lymphatic filariasis (Diamond 2009; Lai *et al.* 2000; Meegan *et al.* 1980; Monath 1988). Including the *Culex pipiens* form pipiens and *Culex pipiens* form molestus biotypes, the *Culex pipiens* complex mosquitoes are an urban vector, with global distribution (Reusken *et al.* 2010b; Shaikevich & Vinogradova 2004; Vinogradova 2000a). Despite sharing similar physiology, *Culex pipiens* complex mosquitoes each possess unique eco-physiological adaptations including, but not limited to survival of environmental stresses such as cold winters and divergent mating behaviors tailored to under or above ground habitats (Table 1). Isolation and disruption of the genetic bases of these differences in vector competence, geographical distribution, and behavioral/reproductive traits is crucial to the control these disease vectors. (Barr 1957; Clements 1992; Harbach *et al.* 1984; Spielman 1967).

As the most common form of genetic variation, single nucleotide polymorphisms (SNPs) are preeminent molecular markers in genetic high-resolution mapping and population genetics studies (Berger *et al.* 2001; Black *et al.* 2001; Venter *et al.* 2001; Wang *et al.* 1998). The abundance of SNPs allows for a higher number of evenly spaced informative markers, a potential advantage over previously employed microsatellite and RFLP markers (Bourguet *et al.* 1998; Ewing *et al.* 1998; Kothera *et al.* 2010). Previous investigations have revealed the genome of *Anopheles gambiae* has approximately 1 SNP

Table 1. Divergent eco-physiological traits between two biotypes of *Culex pipiens* complex mosquitoes.

Trait / Biotype	Breeding Site	Mating Pattern	Host-feeding Preference	Vitellogenesis	Overwintering
Pipiens	Epigeous (Above ground)	Eurygamous (Open spaces)	Orninthophilic (Birds)	Anautogenous (Requires bloodmeal)	Winter diapausing
Molestus	Hypogeous (Underground)	Stenogamous (Enclosed spaces)	Mammalophillic (Mammals)	Autogenous (First oviposition does not require bloodmeal)	No diapause

every 125 bp, *Anopheles funestus* has about 1 SNP every 138 bp, *Aedes aegypti* has 1 SNP about 83 bp, and that *Drosophila melanogaster* exhibits 1 SNP every 38 bp (Chan *et al.* 2012; Morlais *et al.* 2004; Morlais & Severson 2003; Wondji *et al.* 2007). Phenotypic differences between two biotypes are often the result of quantitative conserved genetic variation between populations. Based on the divergent life strategies of *Culex pipiens* complex mosquitoes, and the high SNP frequencies found in other members of Culicidae we hypothesized the two *Culex pipiens* complex biotypes will yield an abundance number of SNP markers for use in genetic study.

According to the neutral theory, drift (random sampling) and mutation (diversifying) work as opposing forces to create equilibrium to maintain SNP allele frequencies in natural populations (Gibson & Muse 2009). Synonymous coding region SNP substitutions and noncoding SNPs offer useful candidate markers as they are protected from selection pressures, while nonsynoymous substitutions resulting in beneficial amino acid changes are conserved by positive selection and are useful for characterization of specific phenotypic differences (Cohuet *et al.* 2008; Wondji *et al.* 2007).

The availability of the *Culex pipiens quinquefasciatus* (Johannesburg strain) physical map (genome size 579 Mb, 18,883 predicted genes), a sister species of *Culex pipiens*, as a reference genome offers a wealth of opportunity for genetic investigation (http://cquinquefasciatus.vectorbase.org/) (Arensburger *et al.* 2010; Bartholomay *et al.* 2010; Hickner *et al.* 2013). Combined with the characterization of SNPs across the *Culex pipiens* complex genomes, the physical map will allow quantitative trait loci (QTL) analysis: the identification of which loci contribute to polygenic phenotypic variation (Severson *et al.* 2001).

Here, we report a set of 28 genes with informative SNP markers characterized from two biotypes: *Culex pipiens* form pipiens and *Culex pipiens* form molestus. Genes were selected to be evenly spaced across the reference *Culex pipiens quinquefasciatus* physical map, and were adapted from previous genetic mapping studies (Arensburger *et al.* 2010; Mori *et al.* 1999b). Genetic polymorphisms in both coding and noncoding regions are documented in detail with regard to nucleotide diversity, coding bias and gene function. The SNP markers reported in this study can increase the capacity of genetic maps and QTL studies investigating the genetic basis of key traits of disease vectors.

Methods

Mosquitoes Used for SNP Discovery

Culex pipiens form pipiens and *Culex pipiens* form molestus colonies are maintained in the vector biology laboratory at Baylor University. Specimens were raised at 25 ^{o}C, 75% relative humidity, with a 15 hour light: 9 hour dark (L:D) daily light cycle. Larvae are reared in de-chlorinated tap water and fed on Tetramin fish food (Tetra holding inc., Blacksburg, VA). Adults are maintained on honey-soaked sponges and kept

in large screened cages. *Culex pipiens* form pipiens was originally collected in Columbus, Ohio fall 2000 (Robich & Denlinger 2005b). In addition, the *Culex pipiens* form molestus was collected from the Calumet Water Reclamation Plant in Chicago, and was provided by Dr. Linda Kothera at the Centers for Disease Control and Prevention Division of Vector Borne Infectious Diseases at Fort Collins, Colorado (Mutebi & Savage 2009).

SNP Loci Selection

DNA sequences of microsatellite or RFLP Loci from the previous genetic studies were used to identify the contigs in the genome database of *Culex quinquefasciatus* (http://cquinquefasciatus.vectorbase.org/) (Arensburger *et al.* 2010; Mori *et al.* 1999a). The contigs containing the markers were identified, and then thirty-six candidate genes were selected near the microsatellite or RFLP locations. Primers, ranging between 300 and 600 bp, were developed using Primer3 (http://frodo.wi.mit.edu/primer3/) spanning the nested introns of the candidate genes.

Gene Extraction, Amplification and Sequencing

Genomic DNA for each form of mosquito was extracted from 10 pooled females with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer protocol. After being tested for purity on a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE), gDNA pooled from 10 female *Culex pipiens* form molestus or 10 female *Culex pipiens* form pipiens was then amplified by PCR. PCR was performed on a T100 thermal cycler (Bio-Rad, Hercules) with 100 ng of genomic DNA in a final volume of 50 µl containing 0.5 µl *Taq* polymerase (Qiagen), 5 µl 10X

buffer, 1.5 μl of 10 mM dNTPs, and 5 pmoles of primer. Amplification cycles consisted of 94°C for 2 minutes; 40 cycles of 94°C for 30 seconds, 47-53°C for 30 seconds, 72°C for 45 seconds; and a final extension of 72 for 5 minutes. Annealing temperatures were optimized for each primer pair, and PCR products were visualized on a 1.2% agarose gel with a 100 bp molecular weight ladder (Invitrogen, Carlsbad, CA). Among thirty-six primer pairs, single PCR band was amplified from 28 genes. The PCR products were cleaned by ExoSAP-IT (Affymetrix, Santa Clara, CA) treatment for DNA sequencing.

Sequencing and SNP Identification

Samples were next submitted to Macrogen USA for DNA sequencing (http://www.macrogenusa.com/). DNA sequences from PCR products were then inspected and aligned using CLC Main Workbench v6 (CLC bio). In both coding and non-coding regions SNPs were classified as Type I, II, III or IV. Codon positions of SNPs in coding regions were further identified, and then the ratios of transversions to transitions and those of synonymous and nonsynonymous substitutions were tested for coding bias. DNAsp v5.10.01 was then utilized to determine nucleotide diversity, Ks, and Ka (Librado & Rozas 2009).

SNP Validation

SNPs discovered in this study were validated by resequencing with reverse primers and Amplifluor SNPs Genotyping System (Millipore, Billerica, Massachusetts). Briefly, primers were designed with a unique hairpin loop at the 5' end with a quencher preventing the fluorophore reporter from fluorescing. Upon specific SNP annealing,

polymerases on the complementary strand open the hairpin allowing for either FAM or JOE fluorescence. Analyses were performed on a Rotor-Gene Q real-time thermal cycler (Qiagen), with reactions containing 10 ng genomic DNA, 0.5 µl 20X Amplifluor SNP FAM Primer, 0.5 µl 20X Amplifluor SNP Joe Primer, 0.5 µl 20X specific primer mix (containing 0.5 µM Green Forward Primer, 0.5 µM Red Forward Primer, 7.5 µM Common Reverse Primer), 1.0 µl 10X Reaction Mix S-Plus buffer, 0.8 µl 2.5 mM dNTPs and 0.1 µl Titanium *Taq* DNA Polymerase (Clontech). Amplification cycles consisted of 96°C for 4 minutes; 18 cycles of 96°C for 10 seconds, 53-58°C for 5 seconds, 72°C for 10 seconds; 22 cycles of 96°C for 10 seconds, 53-58°C for 20 seconds, 72°C for 40 seconds and a final extension at 72°C for 3 minutes. Annealing temperatures were optimized for each primer pair, and fluorescence was monitored as per manufacturer protocol.

Results

Amplification and Sequencing

Primer sets were designed to cross exon-intron boundaries which are potential polymorphic sites before PCR amplification from gDNA of each *Culex* pipiens complex mosquito. Out of 36 primer sets 28 were successfully amplified. The total of 4714 bp (base pairs) across orthologs of *C. pipiens* form pipiens and *C. pipiens* form molestus were then sequenced, revealing 44 SNPs in the 3279 bp coding regions, 48 SNPs in the 1435 bp non-coding regions, and 10 indels (Table 2). The length of each amplicon after stringent quality control varied between 24 and 306 bps. SNP data was then submitted to NCBI SNP database (dbSNP accession nos. 947844444 - 947844519). Distributions and diversities of SNPs in the *Culex pipiens* complex were reported in Table 2.

Polymorphism

Analysis revealed a higher frequency of transitions (60 of 94; 63.8%) than transversions (34 of 94; 36.2%) with an R ratio approaching 2:1 (1.8:1). 95.5% (44 of 46) of all coding region substitutions were synonymous with the third codon position yielding the highest frequency of synonymous substitutions (39 of 46; 84.8%), with every third codon replacement being synonymous. In the coding regions 28 of 46 (60.9%) of substitutions were synonymous transitions, while 16 of 46 (34.8%) were synonymous transversions. Furthermore, 1 of 46 (2.2%) of substitutions in the coding regions were non-synonymous transitions and 1 of 46 (2.2%) were non-synonymous transversions, resulting in 2 of 46 (4.4%) of all coding region substitutions resulting in missense amino acid replacements (Table 3). For all loci analyzed class I, transition substitutions (C/T and G/A) were the most common (59 of 92; 64.1%) and the frequency of transitions was not found to be significantly different (χ^2=0.048, P=0.8258) between the coding (28 of 44; 63.6%) and noncoding regions (31 of 48; 64.4%). The remaining coding

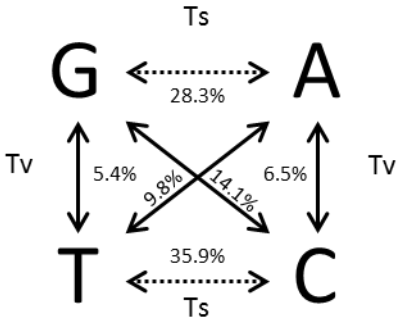

Figure 1. Distribution of single nucleotide polymorphisms for all regions examined between molestus and pipiens biotypes. Purine to purine substitutions and pyrimidine to pyrimidine substitutions are defined as transitions, while purines to pyrimidine mutations or vice versa are classified as transversions.

Table 2. Distributions and diversities of single nucleotide polymorphisms in the *Culex pipiens* complex.

	Coding																	Non-Coding				
	Codon Polymorphic Position													Nucleotide Diversity				Polymorphism				
	Transition					Transversion				# Polymorphism Types												Nucleotide Diversity
Gene	L (bp)	1st	2nd	3rd	Total	1st	2nd	3rd	Total	Syn	Nonsyn	Indel	Total	π	πn	Ks	Ka	L (bp)	Ts	Tv	Indel	Total	π
CPIJ006671	89	0	0	1	1	0	0	0	0	1	0	0	1	0.0112	0.0000	0.0492	0.0000	0	0	0	0	0	0.0000
CPIJ003890	125	0	0	1	1	0	0	0	0	1	0	0	1	0.0080	0.0000	0.0319	0.0000	0	0	0	0	0	0.0000
CPIJ009089	150	1	0	4	5	0	0	1	1	6	0	0	6	0.0400	0.0000	0.1865	0.0000	0	0	0	0	0	0.0000
CPIJ002431	110	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	65	3	2	0	5	0.0769
CPIJ004343	161	0	0	4	4	0	0	0	0	4	0	0	4	0.0248	0.0000	0.1019	0.0000	0	0	0	0	0	0.0000
CPIJ004272	62	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	64	8	2	2	12	0.1563
CPIJ013307	150	1	0	0	1	0	0	1	1	2	0	0	2	0.0133	0.0000	0.0551	0.0000	0	0	0	0	0	0.0000
CPIJ003470	104	0	0	0	0	0	1	0	1	0	1	0	1	0.0096	0.0136	0.0000	0.0137	35	0	0	0	0	0.0000
CPIJ000470	167	0	0	1	1	0	0	2	2	3	0	0	3	0.0180	0.0000	0.0706	0.0000	20	0	0	0	0	0.0000
CPIJ007696	198	0	0	3	3	0	0	3	3	6	0	0	6	0.0303	0.0000	0.1191	0.0000	18	3	0	0	3	0.1667
CPIJ005613	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	63	3	0	0	3	0.0476
CPIJ006471	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	172	2	3	0	5	0.0291
CPIJ004396	132	1	0	2	3	0	0	0	0	2	1	0	3	0.0227	0.0101	0.0711	0.0101	0	0	0	0	0	0.0000
CPIJ005878	68	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	149	1	1	1	3	0.0134
CPIJ008264	54	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	64	0	0	1	1	0.0000
CPIJ008265	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	106	6	4	1	11	0.0943
CPIJ018569	123	0	0	0	0	0	0	1	1	1	0	0	1	0.0081	0.0000	0.0364	0.0000	0	0	0	0	0	0.0000
CPIJ000207	229	1	0	2	3	0	0	1	1	4	0	0	4	0.0175	0.0000	0.0772	0.0000	0	0	0	0	0	0.0000
CPIJ005652	136	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	63	2	1	0	3	0.0476
CPIJ008758	111	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	76	0	1	0	1	0.0132
CPIJ010827	98	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	181	1	1	2	4	0.0110
CPIJ008915	229	0	0	1	1	0	0	0	0	1	0	0	1	0.0044	0.0000	0.0248	0.0000	0	0	0	0	0	0.0000
CPIJ007044	306	0	0	3	3	0	0	1	1	4	0	0	4	0.0131	0.0000	0.0574	0.0000	0	0	0	0	0	0.0000
CPIJ004516	128	0	0	0	0	0	0	1	1	1	0	0	1	0.0078	0.0000	0.0332	0.0000	0	0	0	0	0	0.0000
CPIJ013966	170	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	12	1	0	0	1	0.0833
CPIJ008915	155	0	0	2	2	0	0	4	4	6	0	0	6	0.0387	0.0000	0.1837	0.0000	0	0	0	0	0	0.0000
CPIJ013141	24	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	82	0	1	0	1	0.0122
CPIJ008369	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	152	1	1	3	5	0.0132

L is length of amplicon; Syn, synonymous substitutions; Nonsyn, replacement substitutions; Ts, transitions; Tv, Transversions; π, nucleotide diversity; πn nonsynonymous nucleotide diversity; Ks, average nucleotide substitutions per synonymous site; Ka, average nucleotide substitutions per non-synonymous site.

polymorphisms were transversions (16 of 44; 36.4%) with 4.5% (2 of 44) of all SNPs representing class II (C/A and G/T), 27.3% (12 of 44) class III (C/G) and 4.5% (2 of 44) class IV (A/T) polymorphisms. Noncoding regions exhibited 35.4% (17 of 48) transversions with 18.8% (9 of 48) representing class II, 2.1% (1 of 48) class III, and 14.6% (7 of 48) class IV polymorphisms (Fig. 1 and Table 4). No indels were found in the coding regions and a total of 10 indels were found in the noncoding region of 6 of 29 (20.6%) genes (Table 4). Table 4 also shows that the substitutions were synonymous transitions, while 16 of 46 (34.8%) were synonymous transversions. Furthermore, 1 of 46 (2.2%) of substitutions in the coding regions were non-synonymous transitions and 1 of 46 (2.2%) were non-synonymous transversions, resulting in 2 of 46 (4.4%) of all coding region substitutions resulting in missense amino acid replacements (Table 3). For all loci analyzed class I, transition substitutions (C/T and G/A) were the most common (59 of 92; 64.1%) and the frequency of transitions was not found to be significantly different ($\chi 2=0.048$, P=0.8258) between the coding (28 of 44; 63.6%) and noncoding regions (31 of 48; 64.4%). The remaining coding polymorphisms were transversions

Table 3. Distribution of synonymous and non-synonymous substitutions between coding regions, non-coding regions in respect to transitions and transversions.

Coding Region Replacements	**Total**	**%**
Synonymous		
Transitions	28	60.9
Transversions	16	34.8
Nonsynonymous		
Transitions	1	2.2
Transversions	1	2.2
Nonsense	0	0.0
Missense	2	4.4
Total	16	

Noncoding Region Replacements	**Total**	**%**
Transitions	31	64.6
Transversions	17	35.4
Total	48	

Total Replacements	**Total**	**%**
Transitions	60	63.8
Transversions	34	36.2
Total	94	

(16 of 44; 36.4%) with 4.5% (2 of 44) of all SNPs representing class II (C/A and G/T), 27.3% (12 of 44) class III (C/G) and 4.5% (2 of 44) class IV (A/T) polymorphisms. Noncoding regions exhibited 35.4% (17 of 48) transversions with 18.8% (9 of 48) representing class II, 2.1% (1 of 48) class III, and 14.6% (7 of 48) class IV polymorphisms (Fig. 1 and Table 4). No indels were found in the coding regions and a total of 10 indels were found in the noncoding region of 6 of 29 (20.6%) genes (Table 4). Table 4 also shows that the frequency of transitions at four-fold degeneracy sites (4 of 19; 21.1%), was significantly lower (χ^2=14.892, P=0.0001) than that of coding regions (28 of 44; 63.6%).

Table 4. Polymorphism class and the degeneracy of the genetic code.

Polymorphism	Coding	Wobble Position	Fourfold degenerate	Non-coding
Transitions				
Class I (C/T or G/A)	28	24	4	31
Transversions				
Class II (C/A or G/T)	2	1	1	9
Class III (C/G)	12	12	12	1
Class IV (A/T)	2	2	2	7
Indels	0	0	0	10
Total	44	39	19	58

Diversity

On average, coding regions exhibited a SNP per 74.5 bp with an average nucleotide diversity (π) of 0.0096 ± 0.012, while noncoding regions exhibited a SNP per 29.9 bp with an average nucleotide diversity of 0.027 ± 0.047. In general, due to the rarity of non-synonymous substitutions the rate of non-synonymous substitutions at non-synonymous sites (Ka) was much lower (0.001 ± 0.003) than the rate of synonymous substitutions at synonymous sites (Ks) (0.0499 ± 0.077) leading to a Ka/Ks ratio of 0.02.

In coding regions, the highest nucleotide diversities (π) were the result of synonymous substitutions (Ks) and were found in the genes encoding Rab-7 (CPIJ009089), n-acetylgalactosaminyltransferase (CPIJ007696) and heat shock 70 kDa protein cognate 4 (CPIJ008915**).** In comparison, all nonsynonymous substitutions were found in predicted hypothetical proteins. The GC content of the coding regions was approximately 60.2% in the pipiens biotype and 59.8% in the molestus biotype, while the noncoding regions exhibited 40.2% GC content in the pipiens biotype and 40.1% GC content in the molestus biotype (Table 5). While the frequency of transitions to transversions was not significantly different between the coding and non-coding regions (χ^2=0.048, P=0.8258), among the transversion classes (Type II, III and IV SNPs) there was significant mutation bias favoring GC mutations in the coding regions (χ^2=21.5, P=0.0001), that was not observed in the noncoding regions (χ^2=4.294, P=0.1168) (Table 4).

Discussion

In this study we investigated the single nucleotide polymorphisms between *Culex pipiens* f. pipiens, and *Culex pipiens* f. molestus. While we discovered polymorphisms between the two biotypes, we found that amino acids were strongly conserved in functional regions. Our results were similar to those found in *Aedes aegypti* and *Anopheles funestus* with the vast majority of coding region SNPs being synonymous substitutions with the third codon yielding the highest percentage of coding substitutions (Morlais & Severson 2003; Wondji *et al.* 2007).

The ratios of transitions to transversions approaches 2:1 for both biotypes, similar to those found in mosquitoes such as *Anopheles funestus, Aedes aegypti* and *Culex theileri* (Demirci *et al.* 2012; Morlais & Severson 2003; Wondji *et al.* 2007). Ka is the

number of nonsynonymous substitutions at possible nonsynonymous sites and Ks is the number of synonymous substitutions at synonymous sites. While it is lower than those of *Anopheles funestus, Anopheles gambiae* or *Aedes aegypti*, the low average Ka/Ks ratio (0.022) found between the biotypes of *Culex pipiens* is consistent with expectations that coding regions would have pressures against nonsynonymous substitution due to purifying selection (Morlais *et al.* 2004; Morlais & Severson 2003; Wondji *et al.* 2007).

The bias of substitutions towards transitions and third codon position replacements is likely due to the degeneracy of the genetic code. Substitutions at the third codon position, or wobble position, are least likely to code for missense/nonsense substitutions with transition at this position predominantly resulting in a silent substitution (93.8%). Additionally, the third position is the only site which provides for four-fold degeneracy, with 50% of all third position substitutions resulting in the same amino acid regardless of which nucleic acid is present. Furthermore, the four-fold degenerate sites exhibit a marked shift away from the transition bias found in the coding regions. As substitutions at four-fold degenerate sites are neutral, this shift towards transversions is unsurprising, as statistically there are twice as many possible transversion (C/A, G/T, C/G, A/T) mutations as transitions (C/T, G/A).

The frequency of indels between the two *Culex pipiens* biotypes (20.6%) was higher than the intraspecific diversity found in *Anopheles funestus* (8%), but slightly lower than those found in *Aedes aegypti* (24%) and *Anopheles gambiae* (25%) (Morlais *et al.* 2004; Morlais & Severson 2003; Wondji *et al.* 2007). An easy explanation for the differences in indel frequencies may be distinct selection pressures such as insecticide resistance, filarial resistance, and various mosquito behaviors (Morlais *et al.* 2004;

Morlais & Severson 2003; Wondji *et al.* 2007). The lack of indels in the coding regions of *Culex pipiens* further supports the presence of strong purifying selection against deleterious frame shift mutations. However, further studies required to validate this speculation.

While coding preference varied between genes, the average GC content of the *Culex* biotypes were found to be similar to *Culex theileri* (52.5%), as well previous studies of *Culex pipiens* (63.2%), and was higher in coding regions than noncoding regions (Demirci *et al.* 2012; Samantha 2007). Our results support previous studies suggesting a bias towards higher GC content in Diptera (Morlais & Severson 2003). This bias for higher GC content amongst transversions in regions coding for functional transcripts could be due to selection against stop codons (TAA, TAG, TGA), which have relatively low GC content (Wuitschick & Karrer 1999).

Noncoding regions exhibited higher nucleotide diversity (0.027) than coding regions (0.0096), but lower nucleotide diversity than synonymous coding positions (0.0449). This lower nucleotide diversity if noncoding regions relative to synonymous coding positions is similar to results found in *Anopheles funestus, Anopheles gambiae* or *Aedes aegypti* (Morlais *et al.* 2004; Morlais & Severson 2003; Wondji *et al.* 2007). Variability of nucleotide diversity between each locus is likely a result of each gene having a unique function, and consequently unique selection pressure. While each gene was variable, the generally higher nucleotide diversity of noncoding regions than coding regions implies that noncoding regions are under lower purifying selection than coding regions, but that the redundancy of the third codon position uniquely relieves it of selection pressure. The highest coding region mutation frequencies were found in genes

involved in immune defense and were associated with lysosome biosynthesis (CPIJ009089), pathogen recognition (CPIJ007696), and endocytosis (CPIJ008915). These findings support previous hypotheses that mosquito genes associated with immune responses are under diversifying selection to adapt in response to constantly evolving pathogens (Morlais & Severson 2003). Unfortunately, because the nonsynonymous substitutions occurred in hypothetical proteins we are unable to presume whether these genes are subject to diversifying selection, or if they might be pseudogenes. The genes with low levels of polymorphism located in highly conserved domains may suggest functional adaptation of each *Culex* biotype. However, further investigation into the functional analysis of these alleles is needed.

The increasing risk of vector-borne diseases necessitates new control measures targeting genes that control critical phenotypes in the vectors. We expect that the SNPs and indels characterized from the 28 genes in this study will provide informative genetic markers for future genetic studies. Thus we can eventually identify the genes linked to crucial phenotypic differences relevant to disease transmission in the *Culex pipiens* complex.

References

1. Monath TP: The Arboviruses : epidemiology and ecology. Boca Raton, Fla.: CRC Press; 1988.

2. Diamond MS: West Nile encephalitis virus infection : viral pathogenesis and the host immune response. New York, NY: Springer; 2009.

3. Lai CH, Tung KC, Ooi HK, Wang JS: Competence of Aedes albopictus and Culex quinquefasciatus as vector of Dirofilaria immitis after blood meal with different microfilarial density. Vet Parasitol 2000, 90(3):231-237.

4. Meegan JM, Khalil GM, Hoogstraal H, Adham FK: Experimental Transmission and Field Isolation Studies Implicating Culex-Pipiens as a Vector of Rift-Valley Fever Virus in Egypt. American Journal of Tropical Medicine and Hygiene 1980, 29(6):1405-1410.

5. Vinogradova AB: Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control Pensoft; 2000.

6. Shaikevich EV, Vinogradova EB: [Molecular genetic methods for the identification of the urban mosquito Culex pipiens pipiens F. molestus (Diptera, Culicidae)]. Parazitologiia 2004, 38(5):406-412.

7. Reusken CBEM, de Vries A, Buijs J, Braks MAH, den Hartog W, Scholte EJ: First evidence for presence of Culex pipiens biotype molestus in the Netherlands, and of hybrid biotype pipiens and molestus in northern Europe. Journal of Vector Ecology 2010, 35(1):210-212.

8. Barr AR: The distribution of Culex p. pipiens and C.P. quinquefasciatus in North America. Am J Trop Med Hyg 1957, 6(1):153-165.

9. Spielman A: Population structure in the Culex pipiens complex of mosquitos. Bull World Health Organ 1967, 37(2):271-276.

10. Harbach RE, Harrison BA, Gad AM: Culex (Culex) Molestus Forskal (Diptera, Culicidae) - Neotype Designation, Description, Variation, and Taxonomic Status. P Entomol Soc Wash 1984, 86(3):521-542.

11. Clements AN: The biology of mosquitoes, 1st edn. London ; New York: Chapman & Hall; 1992.

12. Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G, Dickson BJ: Genetic mapping with SNP markers in Drosophila. Nat Genet 2001, 29(4):475-481.

13. Black WC, Baer CF, Antolin MF, DuTeau NM: Population genomics: Genome-wide sampling of insect populations. Annu Rev Entomol 2001, 46:441-469.

14. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J et al: Large-scale identification, mapping, and genotyping of

single-nucleotide polymorphisms in the human genome. Science 1998, 280(5366):1077-1082.

15. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al: The sequence of the human genome. Science 2001, 291(5507):1304-+.

16. Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998, 8(3):175-185.

17. Bourguet D, Fonseca D, Vourch G, Dubois MP, Chandre F, Severini C, Raymond M: The acetylcholinesterase gene Ace: a diagnostic marker for the Pipiens and Quinquefasciatus forms of the Culex pipiens complex. J Am Mosq Control Assoc 1998, 14(4):390-396.

18. Kothera L, Godsey M, Mutebi JP, Savage HM: A comparison of aboveground and belowground populations of Culex pipiens (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. J Med Entomol 2010, 47(5):805-813.

19. Morlais I, Severson DW: Intraspecific DNA variation in nuclear genes of the mosquito Aedes aegypti. Insect Mol Biol 2003, 12(6):631-639.

20. Wondji CS, Hemingway J, Ranson H: Identification and analysis of single nucleotide polymorphisms (SNPs) in the mosquito Anopheles funestus, malaria vector. BMC Genomics 2007, 8:5.

21. Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D: Intraspecific nucleotide variation in Anopheles gambiae: new insights into the biology of malaria vectors. Am J Trop Med Hyg 2004, 71(6):795-802.

22. Chan AH, Jenkins PA, Song YS: Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. Plos Genet 2012, 8(12).

23. Demirci B, Lee Y, Lanzaro GC, Alten B: Altitudinal genetic and morphometric variation among populations of Culex theileri Theobald (Diptera: Culicidae) from northeastern Turkey. J Vector Ecol 2012, 37(1):197-209.

24. Lee Y, Seifert SN, Nieman CC, McAbee RD, Goodell P, Fryxell RT, Lanzaro GC, Cornel AJ: High Degree of Single Nucleotide Polymorphisms in California Culex pipiens (Diptera: Culicidae) sensu lato. Journal of Medical Entomology 2012, 49(2):299-306.

25. Gibson G, Muse SV: A primer of genome science, 3rd edn. Sunderland, Mass.: Sinauer Associates; 2009.

26. Cohuet A, Krishnakumar S, Simard F, Morlais I, Koutsos A, Fontenille D, Mindrinos M, Kafatos FC: SNP discovery and molecular evolution in Anopheles gambiae, with special emphasis on innate immune system. BMC Genomics 2008, 9.

27. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F et al: Sequencing of Culex quinquefasciatus Establishes a Platform for Mosquito Comparative Genomics. Science 2010, 330(6000):86-88.

28. Bartholomay LC, Waterhouse RM, Mayhew GF, Campbell CL, Michel K, Zou Z, Ramirez JL, Das S, Alvarez K, Arensburger P et al: Pathogenomics of Culex quinquefasciatus and Meta-Analysis of Infection Responses to Diverse Pathogens. Science 2010, 330(6000):88-90.

29. Hickner PV, Mori A, Chadee DD, Severson DW: Composite Linkage Map and Enhanced Genome Map for Culex pipiens Complex Mosquitoes. Journal of Heredity 2013, 104(5):649-655.

30. Severson DW, Brown SE, Knudson DL: Genetic and physical mapping in mosquitoes: molecular approaches. Annu Rev Entomol 2001, 46:183-219.

31. Mori A, Severson DW, Christenson BM: Comparative linkage maps for the mosquitoes (Culex pipiens and Aedes aegypti) based on common RFLP loci. Journal of Heredity 1999, 90(1):160-164.

32. Demirci B, Lee Y, Lanzaro GC, Alten B: Identification and characterization of single nucleotide polymorphisms (SNPs) in Culex theileri (Diptera: Culicidae). J Med Entomol 2012, 49(3):581-588.

33. Samantha MP: Nucleotide distribution patterns in insect genomes. Systemix Reports 2007, 1.

34. Wuitschick JD, Karrer KM: Analysis of genomic G + C content, codon usage, initiator codon context and translation termination sites in Tetrahymena thermophila. J Eukaryot Microbiol 1999, 46(3):239-247.

35. Robich RM, Denlinger DL: Diapause in the mosquito Culex pipiens evokes a metabolic switch from blood feeding to sugar gluttony. Proc Natl Acad Sci U S A 2005, 102(44):15912-15917.

36. Mutebi JP, Savage HM: Discovery of Culex pipiens pipiens form molestus in Chicago. J Am Mosq Control Assoc 2009, 25(4):500-503.

37. Mori A, Severson DW, Christensen BM: Comparative linkage maps for the mosquitoes (Culex pipiens and Aedes aegypti) based on common RFLP loci. J Hered 1999, 90(1):160-164.

38. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009, 25(11):1451-1452.

CHAPTER FOUR

A comparative examination of key expression differences between the *Culex pipiens* form pipiens and *Culex pipiens* form molestus biotypes.

Abstract

The taxonomy *Culex pipiens* complex of mosquitoes is still debated, but in North America it is generally regarded to include *Culex pipiens* form pipiens, *Culex pipiens* form molestus and *Culex quinquefasciatus*. While these mosquitoes have very similar morphometry, they each have unique life strategies specifically adapted to their unique ecological niche. Differences include the capability for overwintering diapause, blood meal preference, mating habitat preference, and reliance on blood meals to create eggs. Here we use RNA-seq expression profiling to investigate the transcription differences that result in the divergent phenotypes of these closely related biotypes, and then select five candidate genes for further functional analysis as targets for vector control.

Introduction

Culex pipiens complex mosquitoes transmit numerous pathogens responsible for human diseases including West Nile encephalitis, Rift Valley fever and Lymphatic filariasis (Atkinson *et al.* 1995; Fonseca *et al.* 2004; Hubalek & Halouzka 1999; Kimura *et al.* 2010; Lai *et al.* 2000; Lanciotti *et al.* 2000; Meegan 1979; Michalski *et al.* 2010; Tsai & Mitchell 1989). *C. pipiens* form pipiens and *Culex pipiens* form molestus are two biotypes included in the *C. pipiens* complex of mosquitoes (Reusken *et al.* 2010b; Shaikevich & Vinogradova 2004; Vinogradova 2000a). While both mosquitoes are

found in urban areas, form pipiens mosquitoes reside above ground and established worldwide distribution long before the global spread of form molestus. In contrast, form molestus mosquitoes were not established until the 20th century and inhabit subterranean, man-made infrastructure such as sewers and tunnels (Byrne & Nichols 1999; Kothera *et al.* 2010; Reusken *et al.* 2010a; Shaikevich & Vinogradova 2004; Vinogradova 2000c). As such, while the biotypes share morphological characteristics, pipiens and molestus mosquitoes have adapted unique suites of eco-physiological traits in order to survive two very distinct ecological niches. Key differences between the biotypes include from breeding site preference, blood meal preference, mating patterns, egg production, pollution resistance and diapause (Barr 1957; Byrne & Nichols 1999; Clements 1992; Harbach *et al.* 1984; Spielman 1967).

While the epigenous (above ground) form pipiens thrive in the open (eurygamy), the hypogeous (below ground) form molestus have adapted to reproduce in restricted spaces (stenogamy). Consequently, the feeding and oviposition (egg laying) preferences of molestus have appropriately shifted from the ornithophilic (bird biting) and anautogeneous (eggs laying requires a blood meal) patterns of their pipiens cousins to mammalophilic (mammal biting) and autogenous (egg laying without blood meal) behavior more appropriate for the type and availability of subterranean blood meals (Harbach *et al.* 1985; Mattingly 1967; Vinogradova 2000a). Finally, diapause is an anticipated, alternative developmental program resulting in dormancy that form pipiens mosquitoes use to survive harsh winters, but is absent in the below ground molestus (Denlinger D. L. *et al.* 2005; Tauber *et al.* 1986).

While nucleotide divergence between the two biotypes have been studied in detail, very few studies, with the notable exception of a recent transcriptome study by Price and Fonseca, have investigated the epigenetic control of their unique traits (Arthofer *et al.* 2015; Kang & Sim 2013; Price & Fonseca 2015). Fortunately, the *Culex quinquefasciatus* genome is available as a reference for ontological analysis (Arensburger *et al.* 2010). In this study, we used Illumina RNA-Seq to simultaneously identify and quantify differences in transcript abundance between adult females of the two biotypes one week after adult eclosion. Here, we further validate candidate genes with qRT-PCR and subsequently found 5 genes with differences in transcript abundance between the two biotypes. A deeper understanding of the genetic basis of divergent traits may provide a novel method of control for these disease vectors. Five genes with different transcript abundances showed a diverse biological process including cuticle formation, juvenile hormone synthesis / regulation, olfaction, insecticide resistance, and cellular division/reproduction.

Materials and Methods

Mosquito Rearing

Culex pipiens form pipiens and *Culex pipiens* form molestus specimens are raised under 75% relative humidity, 25 ^{o}C, and a 15 hour light: 9 hour dark (L:D) daily light cycle. Tetramin fish food is fed to larvae (Tetra holding inc., Blacksburg, VA), while adults are maintained on honey-soaked sponges. *Culex pipiens* form pipiens are the Buckeye Strain from Columbus, Ohio fall 2000 (Robich & Denlinger 2005b). *Culex pipiens* form molestus were provided by Dr. Linda Kothera at the Centers for Disease Control and Prevention Division of Vector Borne Infectious Diseases at Fort Collins,

Colorado and originated from the Calumet Water Reclamation Plant in Chicago, Illinois (Mutebi & Savage 2009). Lab mosquitoes are kept in large screens in separate incubation rooms.

Sequencing, Reference Mapping and Clustering

Total RNA extraction was performed seven days after adult eclosion using a standard TRIzol method (Life Technologies, Carlsbad, CA), before being tested on a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) for purity. Samples were obtained and pooled from three females of *Culex pipiens* form pipiens and from three females of *Culex pipiens* form molestus.

After TruSeq mRNA library construction the samples were sequenced on an Illumina HiSeq platform (Illumina Inc., San Diego, CA). Reads were first splice junction mapped via TopHat (v1.3.3), and then transcripts were assembled and tested for differences in abundance using Cufflinks (v2.0.2). (Langmead *et al.* 2009) (Trapnell *et al.* 2010) Cuffdiff was then used to examine differences in expression between the two samples. Throughout this process the genome of *Culex quinquefasciatus* (Johannesburg strain version 2.2) served as a reference genome. Differences of transcript abundance between two biotypes were then visualized using a volcano plot based on FPKM (Fragments per Kilobase of exon per Million fragments mapped) and statistical significance threshold.

The DAVID (Database for Annotation, Visualization and Integrated Discovery) bioinformatics resource v 6.7 was then used to identify and categorize significantly divergent transcripts based on ontology (Huang da *et al.* 2009a, b). Genes were further

Table 1. Gene identifications and qRT-PCR primers.

GeneID	Primer ID		Primer
CPIJ003488 (Cut-1)	adult cuticle protein 1 precursor, putative	Forward Reverse	AGAAGGAGGCTCGCTATCTG GGAGCCAGATTCAGCGACT
CPIJ001822 (Hex-1)	hexamerin 1.1 precursor, putative	Forward Reverse	GTTCTTCTTCATGGTCGCCC GAATGGCAGCGTGTCGAC
CPIJ019485 (Jhest)	juvenile hormone esterase, putative	Forward Reverse	ATCAAACCGAACACTGCACG ATGACTCTGCTGGTGGTGTT
CPIJ000566 (Cath)	cathepsin C, putative	Forward Reverse	CTCGGCCACTTCTCTCTCAG GCAGAGTTCTCACGGTCTCT
CPIJ004167 (Odor)	odorant receptor, putative	Forward Reverse	TCAGCTCCGAACCCAAGTAG TTCAACCTGTTCAACGTGGC

classified by biological process, cellular component and molecular function via the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases.

qRT-PCR Validation

Five transcripts were then selected based on fold change, statistical significance, and relevant ontologies as targets for vector control. An iQ5 real-time PCR detection system (Bio-Rad, Hercules, CA) was then used for qRT-PCR validation with ribosomal protein L19 (RpL19) serving as an internal control. The results were then subjected to a Student's t-test to examine statistical significance. Primer information can be found in Table 1.

Results

Data Analysis

After sequencing by HiSeq 2000 *Culex pipiens* f. pipiens samples, exhibited 32,916,095 total read pairs with an average read length of 101 base pairs and 6,649,051,190 total bases. *Culex pipiens* f. molestus yielded 29,628,110 total read pairs

with an average read length of 101 base pairs and 5,984,878,220 total bases read. The number of reads between the two samples sets were not statistically significant (p=0.4).

56.76-59.05% of the form pipiens transcript reads and 37.47-39.25% of the form molestus uniquely aligned to the reference *Culex quinquefasciatus* genome (genone version?, vectorbase.org). Due to read quality and multiple mapping, TopHat surpressed 0.61-0.62% of the form pipiens reads, and 0.22-0.23% form molestus reads.

Differential Expression

Culex quinquefasciatus, Culex pipiens form pipiens, and *Culex pipiens* form molestus females shared a high level of homology. When compared to the *Cx. quinquefasciatus* reference genome form pipiens only presented 3,789 novel reads out of 32,916,095 total reads and form molestus expressed 4,784 unique reads out of 29,628,110 total reads. Form pipiens yielded 20,041 alternative splices and form molestus yielded 19,367 alternative splices. Finally, the pipiens biotype presented 9,192 potentially novel isoforms while molestus yielded 8,148 novel isoforms. Cufflinks analysis of mapped reads to the reference genome was used to calculate difference in transcript abundance, which was expressed as FPKM (Table 2).

DAVID analysis revealed 16 annotation clusters with significantly higher transcript abundance in the pipiens biotype, with clusters enzymes involving serine

Table 2. Transcript abundance by fragments per kilobase of transcript per million mapped reads (FPKM).

	FPKM<10	10 ≤ FPKM < 100	100 ≤ FPKM <1000	1000 ≤FPKM
Culex pipiens form pipiens	12,156	5,943	844	159
Culex pipiens form molestus	11,070	6,729	931	132

having the highest increase in activity, followed by terms involved in nucleoside binding, ATP binding, and transcriptional regulators. In comparison, the molestus biotype showed 22 clusters, with the highest enrichment in nucleoside binding terms, followed by nucleosome and chromatin assembly, then metabolic processes and polysaccharide binding.

In order to narrow down the pool of candidate genes, we arbitrarily chose the $\log_2$-fold-change threshold of 5.5, at which 73 transcripts had significantly different transcript abundance between adult, female *Culex pipiens* form pipiens and *Culex pipiens* form molestus. Interestingly, *Culex pipiens* form molestus presented a much higher number of up regulated transcripts at this threshold compared to *Culex pipiens* form pipiens. Ontologies of those genes were classified as under structure, olfaction, insecticide resistance, cellular division/reproduction, metabolism, immunity, diapause or hypothetical. Figure 1 reveals the categories of gene function of the divergent transcripts. While the majority of upregulated transcripts were classified as hypothetical proteins, *Culex pipiens* form pipiens exhibited an increase in genes related to metabolism and immunity, and *Culex pipiens* form molestus upregulated genes involved with structure, and cell growth/division. Transcripts were assessed based on fold change differences, statistical significance and relevant ontologies for validation via volcano plot (Fig. 2) for further analysis. Five transcripts: adult cuticle protein 1 precursor, hexamerin 1.1 precursor, juvenile hormone esterase, cathepsin C, and an odorant receptor were selected for phenotypic relevance (Table 3), and then validated with qRT-PCR (Fig. 3).

Discussion

Our results offer a number of potential gene expression targets behind the divergent phenotypes of the *Culex pipiens* form molestus and form pipiens biotypes. We expect that these candidate genes can be used in the disruption or overexpression of key traits relevant to disease transmission. DAVID analysis showed largest categories of significantly upregulated gene expression that involved in nucleotide production, gene expression control and unknown function. Furthermore, an N50 = 476 kb and only 13% of the genome assembled, the low quality of the *Culex quinquefasciatus* genome in conjunction with the wide array of phenotypic differences between the molestus and pipiens biotypes severely reduced the potential use of hypothetical proteins as targets of vector control (Arensburger *et al.* 2010; Naumenko *et al.* 2015).

At the 5.5 $\log_2$-fold-change threshold the increase of metabolic transcripts in the above ground pipiens may be an anticipation of energetically demanding task of finding a blood meal host, and the increased in transcripts related to structure, growth and division in the underground molestus may be in preparation for autogenous oogenesis. While these categories may contain critical transcripts to these phenotypes, again the categories were too broadly ubiquitous to successfully pinpoint targets for vector control.

Fortunately, sorting our data by statistical significance and fold change narrowed down our data enough set to manually examine transcript ontologies for biological relevance. From this narrowed data set five expression targets were selected for maximum potential of key ecophysiological traits: adult cuticle protein 1 precursor, hexamerin 1.1 precursor, juvenile hormone esterase, cathepsin C, and an odorant receptor.

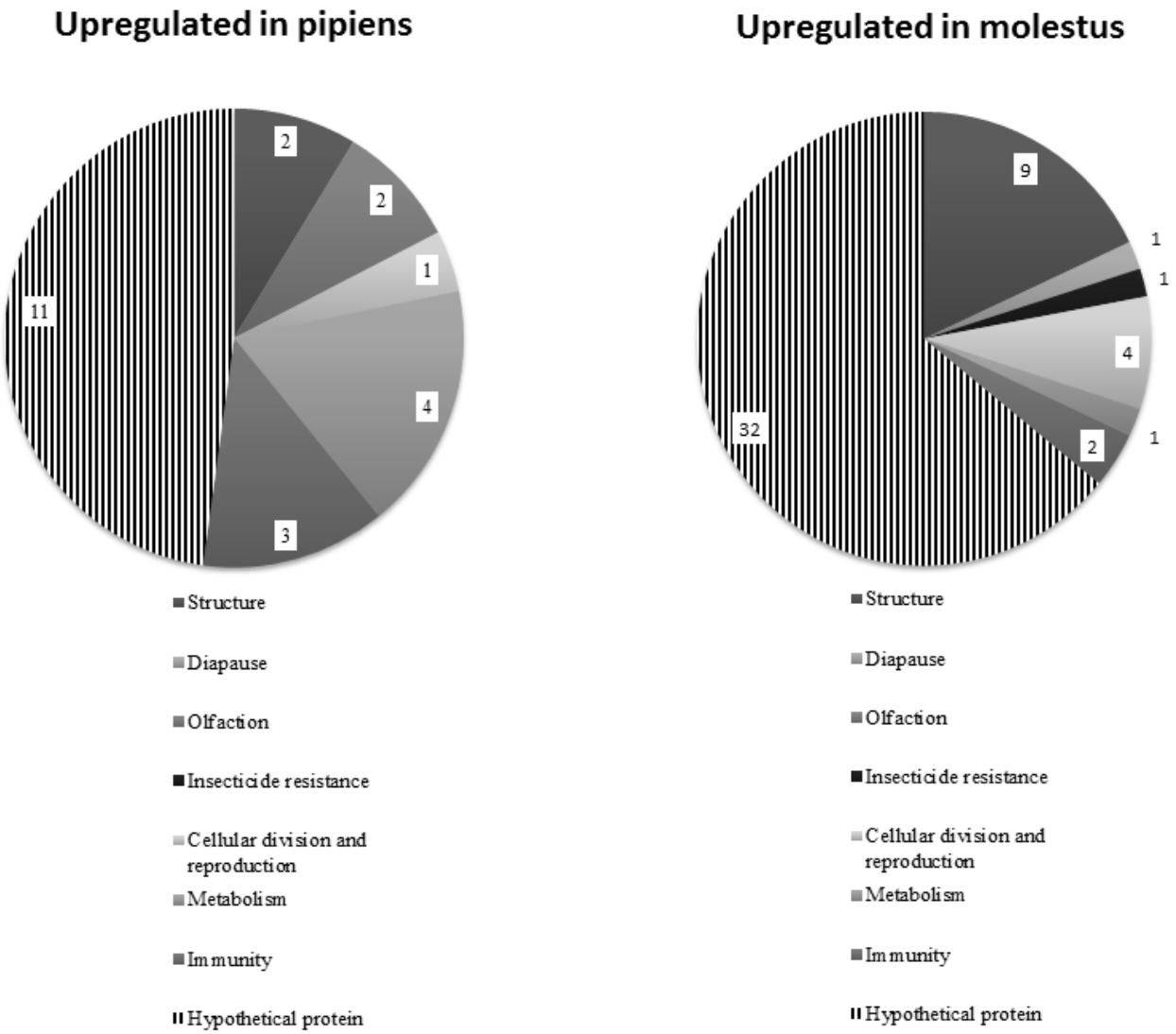

Fig. 1. Significantly upregulated transcripts classified by ontology.

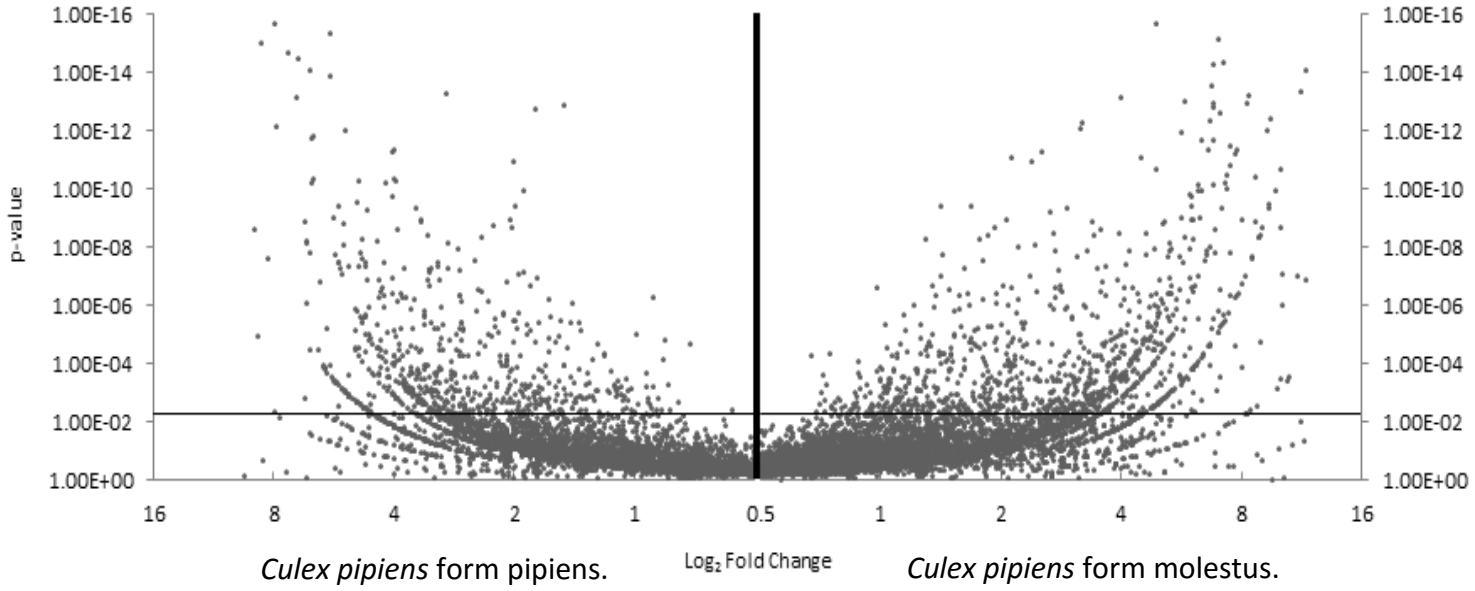

Fig. 2. Volcano plot of $\log_2$ fold change vs statistical significance.

Table 3. Illumina HiSeq 2000 differential expression profiles of *Culex pipiens* form pipiens (P) and *Culex pipiens* form molestus (M) 7 days after female eclosion.

GeneID	Putative Function	FPKM P	FPKM M	Log2 Fold Change
CPIJ003488 (Cut-1)	adult cuticle protein 1 precursor, putative	525.24	0.75	-9.46
CPIJ001822 (Hex-1)	hexamerin 1.1 precursor, putative	25.47	0.1	-7.96
CPIJ019485 (Jhest)	juvenile hormone esterase precursor, putative	4.73	0.08	-5.96
CPIJ000566 (Cath)	cathepsin C, putative	0.09	35.91	8.66
CPIJ004167 (Odor)	odorant receptor, putative	0.05	33.69	9.41

*Fragments per kilobase of exon per million fragments mapped (FPKM).

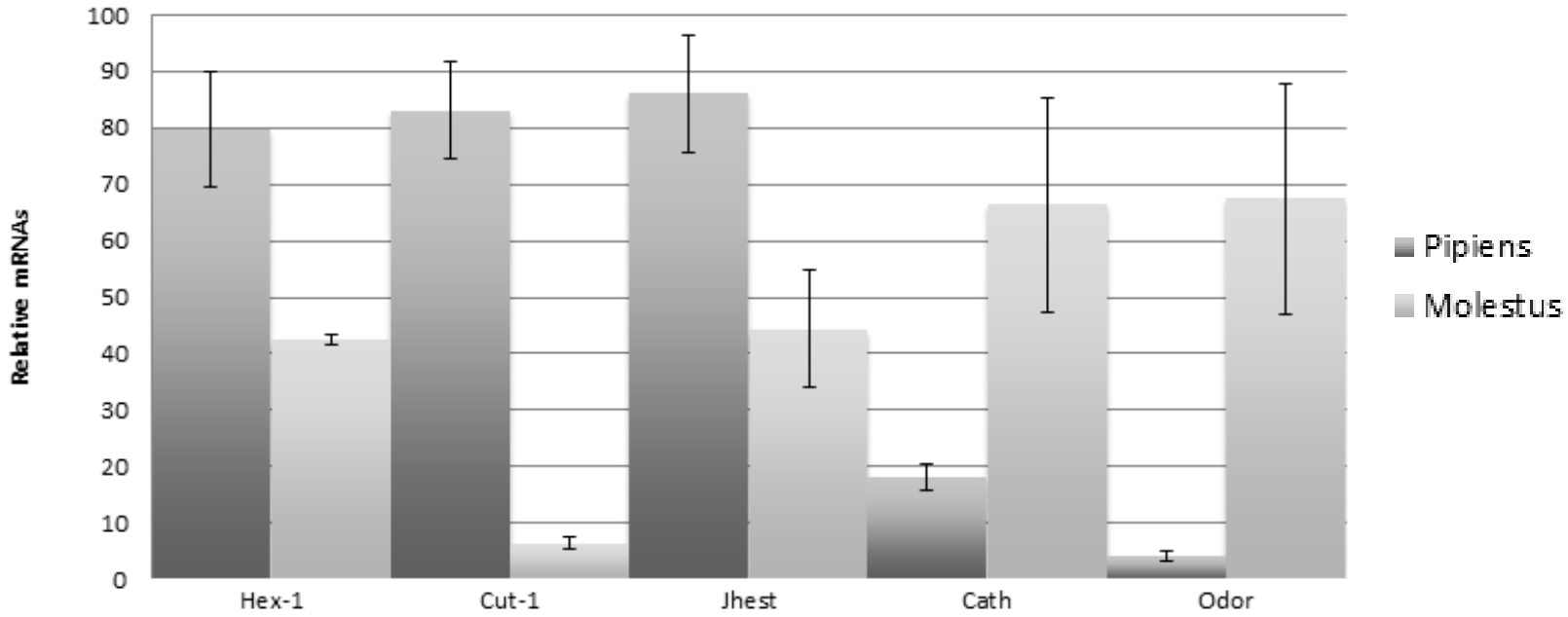

Fig. 3. Expression abundance of female *Cx. pipiens* form molestus and *Cx. pipiens* form pipiens at 7 days after adult eclosion via quantitative real-time PCR. Ribosomal protein large subunit 19 (RpL19) as a loading control. Error bars represent standard error.

An adult cuticle protein was selected due to high upregulation in the form pipiens mosquitoes. While it is not unusual that cuticle protein would be actively expressed in adult mosquitoes, the marked difference in which transcripts are upregulated between the different biotypes warrant further examination. It is known that in *Drosophila melanogaster* and *Culex pipiens* form pipiens the cuticle thickens as a stress tolerance mechanism resulting harder, less permeable integuments (Baker & Russell 2009a; Li & Denlinger 2009). As mechanical properties of insect cuticles are highly optimized for biological properties, it seems likely the expression profile differences are related to a critical phenotype (Andersen *et al.* 1995). Here, we select an adult cuticle protein with highly divergent expression between the two biotypes for further functional analysis.

A hexamerin precursor and a juvenile hormone esterase precursor were selected due to their roles in reproductive programming. Synthesized from fat bodies, hexamerins are a class of protein around 500,000 daltons associated with adult reproductive development in mosquitoes, such as *Anopheles gambiae* (Telfer & Kunkel 1991; Zakharkin *et al.* 1997). In *Drosophila* the absence of hexamarin reduced fly fertility and are known to modulate juvenile hormone mediated reproductive differentiation in the termite genus *Reticulitermes* (Roberts *et al.* 1991; Zhou *et al.* 2007a). Utilizing RNAi and phenotypic rescue Zhou et al. have already demonstrated that presence or absence of hexamarin may influence the developmental fate of insects, and provide excellent candidate targets for vector control (Zhou *et al.* 2007a).

In insects, juvenile hormone esterase serves as a major demethylating regulator of juvenile hormone titer in preparation for critical developmental periods (Lassiter *et al.* 1995; Schomburg & Stephan 1998). In *Aedes aegypti* juvenile hormone esterase and

juvenile hormone III are inversely correlated in response to blood meals. Furthermore, in *Aedes aegypti* blood fed 3 days after adult eclosion, early egg development correlated with an increase in juvenile hormone esterase titer after 36 hours, before necessary a decline 42 hours after blood feeding and pre-blood meal levels restored by 66 hours (Shapiro *et al.* 1986). As form molestus mosquitoes are autogenous, it stands to reason that by 7 days after adult eclosion, reduction of juvenile hormone esterase in specimens may be correlated with restoration of juvenile hormone to pre-blood meal levels.

While the role of cathepsin C or dipeptidyl peptidase I (DPP-I) is relatively unknown in insects, in general it is an activator of serine proteases in resulting in inflammation (Turk *et al.* 2001). Molestus biotype females exhibited very significant upregulation of cathepsin C, which unsurprisingly coincided with an upregulation of serine proteases. In addition to their role as activators of the innate immune system, serine proteases are also associated with diapause, juvenile hormone, hemolymph coagulation, synthesis of antimicrobial peptides and melanin (Gorman & Paskewitz 2001; Robich & Denlinger 2005b; Valenzuela *et al.* 2002). As another closely related serine protease activator, cathepsin B has been implicated in embryonic degredation of vitellin and is key to protein catabolism of lysozymes, a cathepsin C was selected for further functional analysis (Cho *et al.* 1999).

Isolating, understanding and interfering with the mechanisms underlying mosquito olfaction has been a successful approach to vector control (Carey *et al.* 2010; Hallem *et al.* 2004; Syed & Leal 2009; Takken 1991). In mosquitoes, olfactory receptor neurons (ORNs) receptive to semiochemicals are used to breeding sites and nutrients, including blood meals (Dethier 1957; Gibson & Torr 1999). DEET (*N*,*N*-diethyl-meta-

toluamide) has proven to be an effective, topical insect repellent and is known to effect odorant receptors in *Culex quinquefasciatus* and *Anopheles gambiae* (Ditzen *et al.* 2008; Syed & Leal 2008). *Culex quinquefasciatus,* in particular, has been shown to possess around 1,300 olfactory sensilla and several *Culex* odorant binding proteins have already been cloned or mapped to specific sensilla (Ishida *et al.* 2002; Leal *et al.* 2008; McIver 1970; Pelletier & Leal 2009). DEET (*N*,*N*-diethyl-meta-toluamide) has proven to be an effective, topical insect repellent and is known to effect odorant receptors in *Culex quinquefasciatus* and *Anopheles gambiae* (Ditzen *et al.* 2008; Syed & Leal 2008). Here, we present an odorant receptor with significantly increased expression in the molestus biotype as a candidate olfactory receptor for mammal semiochemicals.

The *Culex pipiens* complex includes form pipiens and f. molestus, which in spite of displaying a great deal of variation in ecophysiological traits, are nearly morphologically indistinguishable. The genetic basis of these trait divergences between the species is important steps to understand speciation process, population genetics, and epidemiology of vector-borne diseases. Furthermore, our transcriptome study is consistent with previous studies, and that an abundance of potentially critical targets are available for genetic approaches to vector control.

References

Andersen SO, Hojrup P, Roepstorff P (1995) Insect cuticular proteins. Insect Biochemistry and Molecular Biology 25, 153-176.

Arensburger P, Megy K, Waterhouse RM, et al. (2010) Sequencing of Culex quinquefasciatus Establishes a Platform for Mosquito Comparative Genomics. Science 330, 86-88.

Arthofer W, Banbury BL, Carneiro M, et al. (2015) Genomic resources notes accepted 1 August 2014-30 September 2014. Mol Ecol Resour 15, 228-229.

Atkinson CT, Woods KL, Dusek RJ, Sileo LS, Iko WM (1995) Wildlife disease and conservation in Hawaii: pathogenicity of avian malaria (Plasmodium relictum) in experimentally infected iiwi (Vestiaria coccinea). Parasitology 111 Suppl, S59-69.

Baker D, Russell S (2009) Gene expression during Drosophila melanogaster egg development before and after reproductive diapause. BMC Genomics 10, 242.

Barr AR (1957) The distribution of Culex p. pipiens and C.P. quinquefasciatus in North America. Am J Trop Med Hyg 6, 153-165.

Byrne K, Nichols RA (1999) Culex pipiens in London Underground tunnels: differentiation between surface and subterranean populations. Heredity (Edinb) 82 (Pt 1), 7-15.

Carey AF, Wang G, Su C-Y, Zwiebel LJ, Carlson JR (2010) Odorant reception in the malaria mosquito Anopheles gambiae. Nature 464, 66-71.

Cho WL, Tsao SM, Hays AR, et al. (1999) Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. J Biol Chem 274, 13311-13321.

Clements AN (1992) The biology of mosquitoes, 1st edn. Chapman & Hall, London ; New York.

Denlinger D. L., Yocum G. D., L. RJ (2005) Hormonal control of diapause. In: Comprehensive Molecular Insect Science (eds. Gilbert L. I., Iatrou K., S. GS), pp. 615-650. Elsevier, Amsterdam.

Dethier VG (1957) The sensory physiology of blood-sucking arthropods. Experimental Parasitology 6, 68-122.

Ditzen M, Pellegrino M, Vosshall LB (2008) Insect Odorant Receptors Are Molecular Targets of the Insect Repellent DEET. Science 319, 1838-1842.

Fonseca DM, Keyghobadi N, Malcolm CA, et al. (2004) Emerging vectors in the Culex pipiens complex. Science 303, 1535-1538.

Gibson G, Torr SJ (1999) Visual and olfactory responses of haematophagous Diptera to host stimuli. Medical and Veterinary Entomology 13, 2-23.

Gorman MJ, Paskewitz SM (2001) Serine proteases as mediators of mosquito immune responses. Insect Biochemistry and Molecular Biology 31, 257-262.

Hallem EA, Nicole Fox A, Zwiebel LJ, Carlson JR (2004) Olfaction: Mosquito receptor for human-sweat odorant. Nature 427, 212-213.

Harbach RE, Dahl C, White GB (1985) Culex (Culex) Pipiens-Linnaeus (Diptera, Culicidae) - Concepts, Type Designations, and Description. Proceedings of the Entomological Society of Washington 87, 1-24.

Harbach RE, Harrison BA, Gad AM (1984) Culex (Culex) Molestus Forskal (Diptera, Culicidae) - Neotype Designation, Description, Variation, and Taxonomic Status. Proceedings of the Entomological Society of Washington 86, 521-542.

Huang da W, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13.

Huang da W, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.

Hubalek Z, Halouzka J (1999) West Nile fever--a reemerging mosquito-borne viral disease in Europe. Emerg Infect Dis 5, 643-650.

Ishida Y, Cornel A, Leal W (2002) Identification and Cloning of a Female Antenna-Specific Odorant-Binding Protein in the Mosquito Culex quinquefasciatus. Journal of Chemical Ecology 28, 867-871.

Kang D, Sim C (2013) Identification of Culex complex species using SNP markers based on high-resolution melting analysis. Mol Ecol Resour.

Kimura M, Darbro JM, Harrington LC (2010) Avian malaria parasites share congeneric mosquito vectors. J Parasitol 96, 144-151.

Kothera L, Godsey M, Mutebi JP, Savage HM (2010) A comparison of aboveground and belowground populations of Culex pipiens (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. J Med Entomol 47, 805-813.

Lai CH, Tung KC, Ooi HK, Wang JS (2000) Competence of Aedes albopictus and Culex quinquefasciatus as vector of Dirofilaria immitis after blood meal with different microfilarial density. Vet Parasitol 90, 231-237.

Lanciotti RS, Kerst AJ, Nasci RS, et al. (2000) Rapid detection of west nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. J Clin Microbiol 38, 4066-4071.

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

Lassiter MT, Apperson CS, Roe RM (1995) Juvenile hormone metabolism during the fourth stadium and pupal stage of the southern house mosquito, Culex quinquefasciatus Say. Journal of insect physiology 41, 869-876.

Leal WS, Barbosa RMR, Xu W, et al. (2008) Reverse and Conventional Chemical Ecology Approaches for the Development of Oviposition Attractants for <italic>Culex</italic> Mosquitoes. PLoS One 3, e3045.

Li A, Denlinger DL (2009) Pupal cuticle protein is abundant during early adult diapause in the mosquito Culex pipiens. J Med Entomol 46, 1382-1386.

Mattingly PF (1967) The systematics of the Culex pipiens complex. Bulletin of the World Health Organization 37, 257-261.

McIver SB (1970) Comparative Study of Antennal Sense Organs of Female Culcine Mosquitoes. The Canadian Entomologist 102, 1258-1267.

Meegan JM (1979) The Rift Valley fever epizootic in Egypt 1977-78. 1. Description of the epizzotic and virological studies. Trans R Soc Trop Med Hyg 73, 618-623.

Michalski ML, Erickson SM, Bartholomay LC, Christensen BM (2010) Midgut barrier imparts selective resistance to filarial worm infection in Culex pipiens pipiens. PLoS Negl Trop Dis 4, e875.

Mutebi JP, Savage HM (2009) Discovery of Culex pipiens pipiens form molestus in Chicago. J Am Mosq Control Assoc 25, 500-503.

Naumenko AN, Timoshevskiy VA, Kinney NA, et al. (2015) Mitotic-Chromosome-Based Physical Mapping of the Culex quinquefasciatus Genome. PLoS One 10.

Pelletier J, Leal WS (2009) Genome Analysis and Expression Patterns of Odorant-Binding Proteins from the Southern House Mosquito <italic>Culex pipiens quinquefasciatus</italic>. PLoS One 4, e6237.

Price DC, Fonseca DM (2015) Genetic divergence between populations of feral and domestic forms of a mosquito disease vector assessed by transcriptomics. Peerj 3.

Reusken CB, de Vries A, Buijs J, et al. (2010a) First evidence for presence of Culex pipiens biotype molestus in the Netherlands, and of hybrid biotype pipiens and molestus in northern Europe. J Vector Ecol 35, 210-212.

Reusken CBEM, de Vries A, Buijs J, et al. (2010b) First evidence for presence of Culex pipiens biotype molestus in the Netherlands, and of hybrid biotype pipiens and molestus in northern Europe. Journal of Vector Ecology 35, 210-212.

Roberts DB, Jowett T, Hughes J, Smith DF, Glover DM (1991) The major serum protein of Drosophila larvae, larval serum protein 1, is dispensable. Eur J Biochem 195, 195-201.

Robich RM, Denlinger DL (2005) Diapause in the mosquito Culex pipiens evokes a metabolic switch from blood feeding to sugar gluttony. Proc Natl Acad Sci U S A 102, 15912-15917.

Schomburg D, Stephan D (1998) Juvenile-hormone esterase. In: Enzyme Handbook 15 (eds. Schomburg D, Stephan D), pp. 21-26. Springer Berlin Heidelberg.

Shaikevich EV, Vinogradova EB (2004) [Molecular genetic methods for the identification of the urban mosquito Culex pipiens pipiens F. molestus (Diptera, Culicidae)]. Parazitologiia 38, 406-412.

Shapiro AB, Wheelock GD, Hagedorn HH, et al. (1986) Juvenile hormone and juvenile hormone esterase in adult females of the mosquito Aedes aegypti. Journal of insect physiology 32, 867-877.

Spielman A (1967) Population structure in the Culex pipiens complex of mosquitos. Bull World Health Organ 37, 271-276.

Syed Z, Leal WS (2008) Mosquitoes smell and avoid the insect repellent DEET. Proceedings of the National Academy of Sciences 105, 13598-13603.

Syed Z, Leal WS (2009) Acute olfactory response of Culex mosquitoes to a human- and bird-derived attractant. Proceedings of the National Academy of Sciences 106, 18803-18808.

Takken W (1991) The Role of Olfaction in Host-Seeking of Mosquitoes: A Review. International Journal of Tropical Insect Science 12, 287-295.

Tauber MJ, Tauber CA, Masaki S (1986) Seasonal adaptations of insects Oxford University Press, New York.

Telfer WH, Kunkel JG (1991) The function and evolution of insect storage hexamers. Annual Review of Entomology 36, 205-228.

Trapnell C, Williams BA, Pertea G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511-515.

Tsai TF, Mitchell CJ (1989) St. Louis encephalitis. In: The arboviruses: epidemiology and ecology (ed. Monath TP), pp. 113-143. CRC, Boca Raton, FL.

Turk D, Janjić V, Štern I, et al. (2001) Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. The EMBO Journal 20, 6570-6582.

Valenzuela JG, Pham VM, Garfield MK, Francischetti IMB, Ribeiro JMC (2002) Toward a description of the sialome of the adult female mosquito Aedes aegypti. Insect Biochemistry and Molecular Biology 32, 1101-1122.

Vinogradova AB (2000a) Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control Pensoft.

Vinogradova EB (2000b) Culex pipiens pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control Pensoft, Moscow.

Zakharkin SO, Gordadze AV, Korochkina SE, et al. (1997) Molecular cloning and expression of a hexamerin cDNA from the malaria mosquito, Anopheles gambiae. Eur J Biochem 246, 719-726.

Zhou X, Tarver MR, Scharf ME (2007) Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. Development 134, 601-610.

CHAPTER FIVE

A transcriptomic approach to identification of candidate genes for vector control of *Culex pipiens* mosquitoes.

Abstract

Diapause is a key eco-physiological adaptation in the West Nile Virus vector, *Culex pipiens.* Unfortunately, the mechanisms behind the molecular regulation of diapause are not yet well known. We utilize Illumina RNA-seq to simultaneously identify and quantify relative transcript levels between diapausing and non-diapausing *Culex pipiens* females. After examining 65,623,095 read pairs, we identified 41 genes with significantly different transcript abundances between these two groups. Further we discussed transcriptome divergences between them, which are related to the juvenile hormone synthesis, anaerobic metabolism, innate immunity and cold tolerance.

Author Summary

Diapause is one of the key eco-physiological traits for *Culex pipiens* to overwinter. This alternative developmental program activates the unique set of genes by which the mosquito species is able to cope cold and food shortage. As diapause is a quantitative trait, controlled by a complex network of gene interactions, we examined differences at a transcript levels between diapausing and non-diapausing females. Interestingly, many of these genes have similar structure as those found in other diapausing insects and offer some insight into how these transcripts result in diapause phenotypes. Furthermore, the genes without similar insect orthologs may serve as new

targets of ecological control, which may suppress these disease vectors and agricultural pests.

Introduction

As a vector of West Nile Virus, Eastern Equine Encephalitis virus, and pathogens of lymphatic filariasis the *Culex pipiens* mosquitoes have been of growing concern in the US (Diamond 2009; Lai *et al.* 2000; Meegan *et al.* 1980; Monath 1988). These ubiquitous mosquitoes were the driving force behind the historic 2012 West Nile outbreak that resulted in 5,674 reported cases of human disease in the United States (Control & Prevention 2013). Additionally, evidence of the rise of insecticide resistance in these mosquitoes emphasizes the necessity of alternative vector control strategies (Weill *et al.* 2004).

The invasive *Culex pipiens* complex mosquitoes are virtually indistinguishable by simple morphometrics yet exhibit a robust range of life strategies driven by genetic architecture that may provide a crucial framework for vector control (Harbach *et al.* 1985; Vinogradova 2000a). In the Northern house mosquito, *Culex pipiens*, diapause is an alternative developmental program in which animals sense impending changes in their environment and adapt accordingly (Eldridge 1968). Diapause is an anticipated response triggered by shortened day lengths and low temperature, which in turn restrict the release of juvenile hormone via endocrine signaling (Sim & Denlinger 2008). The diapause program induces diverse phenotypes on physiology and development including delayed reproductive development, stress tolerance, sugar gluttony and nutrient rationing. The physiological and developmental changes allow the mosquitoes to effectively utilize favorable conditions and avoid adverse conditions. In periods of harsh conditions, such

as winter, diapause is a critical adaptation to the survival of these vectors of human disease. Despite the crucial role of diapause to mosquito survival, it remains largely unknown how these organisms are able to predict environmental change (Bowen 1992; Denlinger D. L. *et al.* 2005; Meuti & Denlinger 2013; Mitchell & Briegel 1989; Robich & Denlinger 2005a; Sanburg & Larsen 1973; Sim & Denlinger 2013; Spielman & Wong 1973; Tauber *et al.* 1986).

In general, diapause is a quantitative trait in which a multiple genes share complex interactions that result with the phenotype. Several minor QTLs and a major QTL have already been identified, but mapping studies have impeded by a lack of markers (Mori *et al.* 2007). Here we hypothesize when comparing diapausing and non-diapausing *Culex pipiens* form pipiens differences in transcriptional profiles would elucidate the molecular basis of diapause program in this species. RNA-seq technology was used to simultaneously quantify and identify the transcriptional profiles of diapausing and non-diapausing adult, female *Culex pipiens* mosquitoes.

Materials and Methods

Mosquito Rearing

Culex pipiens colonies are laboratory colonies maintained at Baylor University. As previously described, non-diapausing specimens are raised with a 15 hour light: 9 hour dark (L:D) daily light cycles at 25 ^{o}C, and 75% relative humidity. Diapausing specimens were reared with 9 hour light:15 hour dark (L:D) daily light cycles at 18 ^{o}C, and 75% relative humidity (Robich & Denlinger 2005b). Diapausing incidence was confirmed by measurement of the primary ovarian follicles and germariums as previously described (Christophers 1911). Larvae are reared in de-chlorinated tap water and fed on

Tetramin fish food (Tetra Holding Inc., Blacksburg, VA). Adults are maintained on honey-soaked sponges and kept in large screened cages. *Culex pipiens* mosquitoes were originally collected in Columbus, Ohio fall 2000 (Robich & Denlinger 2005b).

Total RNA Extraction

Total RNA was extracted from three adult diapausing *Culex pipiens* and three non-diapausing adult *Culex pipiens* females 7-10 days after eclosion using TRIzol isolation (Life Technologies, Carlsbad, CA). Total RNA purity was then tested on a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Library Preparation and Sequencing

Samples were then subjected to TruSeq mRNA library construction. Samples were purified twice using poly-T oligo-attached magnetic beads, before fragmentation and priming for cDNA synthesis. cDNA was synthesized using reverse transcriptase and random primers adapted into double stranded (ds) cDNA which were then removed with Ampure XP beads (Beckman Coulter, Pasadena, CA). ds cDNA was then end repaired, converting any resulting overhangs into blunt ends, before adapter adenylation of the 3' for pair ended ligation. Next, adapters were ligated to ds cDNA and then selectively amplified by PCR.

After quality control, bridge amplification was performed on a flow cell, which was then loaded into a HiSeq 2000 Illumina platform (Illumina Inc., San Diego, CA). A single molecular array was synthesized with reverse termination, resulting in unique clusters of nucleotides strands which were then loaded for extension and imagining.

Then resulting clusters were then extended one base at a time with nucleotides containing reversible fluorophores with the cluster giving a single, unified signal for each base.

Data Analysis

Reads were then aligned using TopHat v1.3.3 against the *Culex quinquefasciatus* Johannesburg strain reference genome as found on VectorBase (https://www.vectorbase.org). TopHat employs the short read aligner Bowtie to identify exon splice junctions (Langmead *et al.* 2009). Next, Cufflinks (v2.0.2) was used to assemble transcripts, estimate abundance and test for differences in RNA expression. Additionally, Cufflinks identifies alternative isoforms of target genes, as it does not use existing genetic annotations (Trapnell *et al.* 2010). Cufflinks then extrapolates relative transcript abundance from normalized reads and expressed in Fragments per Kilobase of exon per Million fragments mapped (FPKM), where FPKM is calculated as 10^9 X number of mappable exon reads / (number of total mappable reads X number of base pairs in the exon). Cuffdiff was used to highlight significant differences in transcript expression, splicing, and promoter usage. FPKM values of diapausing and non-diapausing *Culex pipiens* were comparatively examined and expressed in $\log_2$, where gene targets up regulated in non-diapausing specimens exhibited positive fold changes, while targets up regulated in diapausing specimens expressed negative fold changes. Transcripts were then visualized to compare FPKM to significance on a volcano plot (reference about volcano plot?). The transcripts with significant fold changes were then screened for relevant ontologies.

Gene Ontology

DAVID (Database for Annotation, Visualization and Integrated Discovery) bioinformatics resource v 6.7 was utilized to cluster significant changes in gene expression (Huang da *et al.* 2009a). Genes were then classified by biological process, cellular component and molecular function via the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases.

qRT-PCR Validation

Transcripts abundance of genes with known ontologies was screened to identify candidates for vector control. Next, qRT-PCR validation was performed on five candidate genes associated with different functions using an iQ5 real-time PCR detection system (Bio-Rad, Hercules, CA). 50 ng of DNA was reverse transcribed and amplified via superscript III RNase H-reverse transcriptase (Invitrogen, Carlsabad, CA), per manufacture protocol, and compared to ribosomal protein L19 (RpL19), an endogenous housekeeping gene, as an internal control. Transcript divergence from the qRT-PCR results were then evaluated for statistical significance via the Student's t-test. Candidate genes and primer information can be found in Table 1.

Table 1. qRT-PCR primers and associated genes.

GeneID	Primer IDs		Primers
CPIJ007618	alcohol dehydrogenase	Forward	CTGTTGGAAGCTGGAGGAGA
	(OH-deh)	Reverse	CTCTCACGTACACCATTGCG
CPIJ020026	glyocogen debranching enzyme	Forward	CATGTACAAGGACACGCTCG
	(glyd1)	Reverse	GGAGTTGTCGTAGTTTCCGC
CPIJ012251	troponin C	Forward	GACAAGACCGGCCACATTC
	(trop)	Reverse	CATCATGAAGACCTCGCGC
CPIJ012704	pyrroline-5-carboxylate reductase	Forward	AGGCAAGCTGTTCATTTCGG
	(pyr)	Reverse	TTCCAACCGATTCGAACAGC
CPIJ011998	z-carboxypeptidase A1 precursor	Forward	CTGGAGAGCACACACCAAAC
	(z-carb)	Reverse	CATCCCAACTGTCATCGCTG

Results

Data Analysis

HiSeq 2000 sequencing yielded 42,175,155 total read pairs for *Culex pipiens* diapausing samples, with an average read length of 101 base pairs and 8,519,373,230 total bases read. Comparatively, *Culex pipiens* non-diapausing samples yielded 33,447,940 total read pairs with an average read length of 101 base pairs and 6,756,483,880 total bases read. A student t-test reveals that the numbers of reads between the diapausing and non-diapausing samples were significantly different ($p=0.0132$).

Around 55.36-56.44% of the diapausing transcript reads uniquely aligned to the reference *Culex quinquefasciatus* genome from vectorbase.org. Comparatively, 54.52-54.40% of the non-diapausing reads aligned uniquely to the reference genome. TopHat revealed that 0.74-0.75% of the diapause reads, and 0.39-0.40% non-diapause reads had multiple mapping sites or were of low quality. Due to the non-specific nature of these transcripts they were suppressed.

Differential Expression

Examination of whole, female adults revealed a transcriptional high homology between *Culex quinquefasciatus* and *Culex pipiens* in both diapausing and non-diapausing samples. The transcriptome of diapausing females only presented 4,303 unique reads out of 42,175,115 total reads when compared to the *Cx. quinquefasciatus* reference genome, and the non-diapausing sample expressed 4,007 unique reads out of 33,447,940 total reads. Additionally, transcripts from diapausing females of *Cx. pipiens* revealed 21,146 alternative splices compared to the reference *Cx. quinquefasciatus*

genome, yielding 9,388 potentially novel isoforms. Non-diapausing females similarly revealed 20,468 alternative splices, and yielded 9,142 novel isoforms.

Cufflinks analysis of mapped reads to the reference genome was used to calculate difference in transcript abundance, which was expressed as FPKM. An examination of our data reveals the diapausing females of *Cx. pipiens* exhibited 11,193 transcripts with FPKMs below 10 (FPKM<10), 6,174 transcripts with FPKMs greater than or equal to 10 and less than 100 ($10 \leq FPKM < 100$), 838 transcripts with FPKM greater than or equal to 100 and less than 1000 ($100 \leq$ FPKM <1000), and 129 transcripts with a FPKM greater than or equal to 1,000 (FPKM ≤1000). In comparison, the non-diapausing females of *Cx. pipiens* exhibited 11,061 transcripts with FPKMs below 10 (FPKM<10), 6,198 transcripts with FPKMs greater than or equal to 10 and less than 100 ($10 \leq FPKM < 100$), 708 transcripts with FPKM greater than or equal to 100 and less than 1000 ($100 \leq FPKM <1000$), and 259 transcripts with a FPKM greater than or equal to 1,000 (FPKM ≤1000).

Distribution of transcripts can be seen in a volume plot (Fig. 1). Further examination of fold change differences (log_2) reveals 241 transcripts upregulated in diapausing females and 207 transcripts upregulated in non-diapausing females (Fig. 2). qRT-PCR validation of alcohol dehydrogenase, a glycogen debranching enzyme, troponin C, pyrroline-5-carboxylate reductase, and z-carboxypeptidase A1 precursor support the accuracy of our RNA-seq results (Fig. 3).

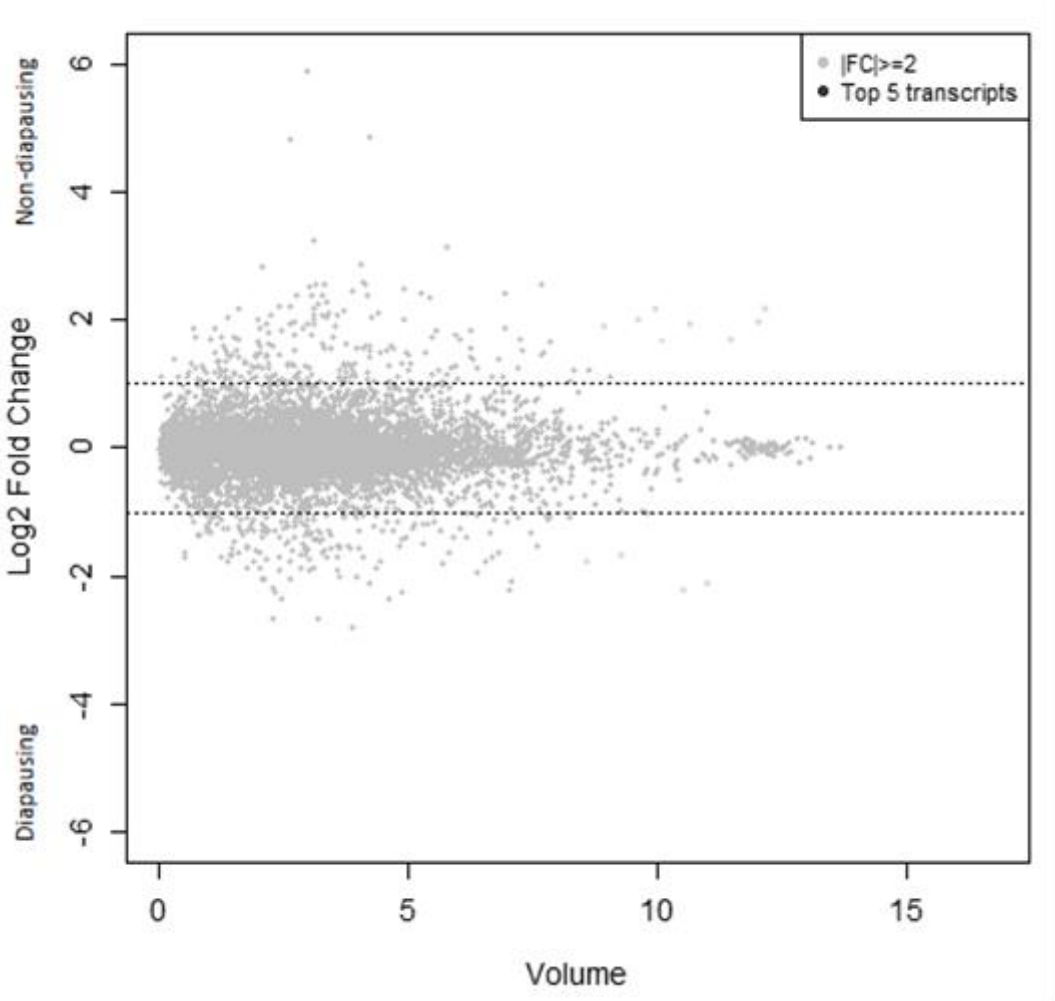

Figure 1. Volume plot distribution of transcripts.

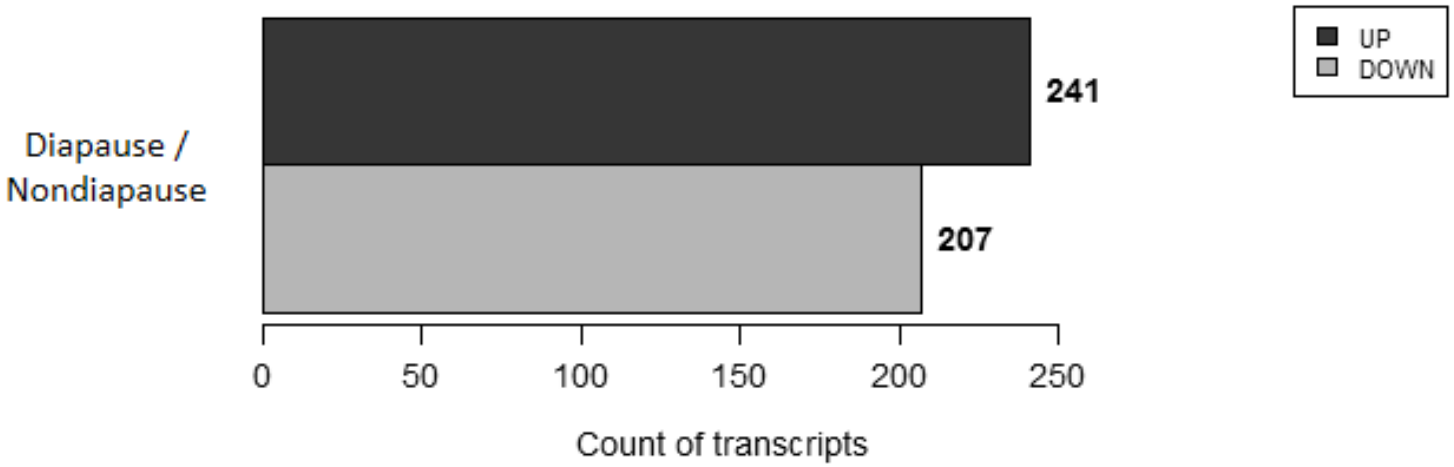

Figure 2. Total genes upregulated in diapausing and non-diapausing females of *Cx. pipiens.* UP, DOWN regulated count by |FC| >=2.

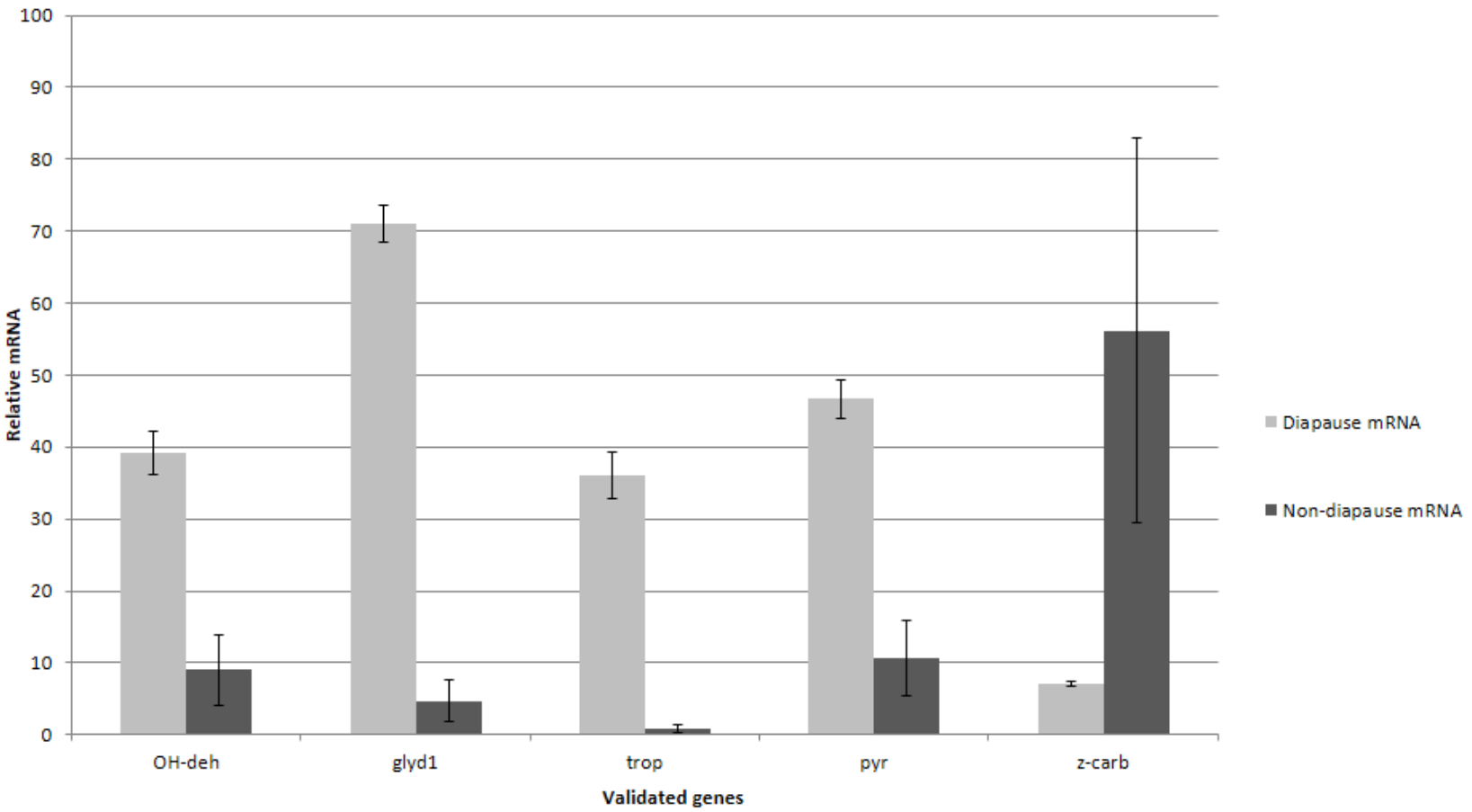

Figure 3. Expression abundance of diapausing vs. non-diapausing females of *Cx. pipiens* at 7 days after adult eclosion via quantitative real-time PCR. Ribosomal protein large subunit 19 (RpL19) as a loading control. Error bars represent standard error.

An examination of P values reveals 41 transcripts with significantly different transcript abundance between diapausing and non-diapausing adult females of *Cx. pipiens* (Table 2). DAVID Analysis of the significantly differently expressed transcripts reveals 12 genes categorized under biological process, 6 genes under cellular components and 17 genes under molecular function, with many transcripts related to glycolysis. Of these significantly different transcripts, three transcripts were mapped to known metabolic/signaling KEGG pathways involved with "starch and sucrose metabolism." However, fourteen transcripts did not map to the reference genome, and 9 were conserved hypothetical proteins. Interestingly, non-diapausing females of *Cx. pipiens* presented a lower number (10) of up regulated transcripts at this threshold compared to diapausing counterparts. Fig. 4 reveals the gene function categories of the divergent transcripts.

As the ultimate numbers of significantly different transcripts were relatively low the ontology of each individual transcript was manually investigated beyond DAVID analysis, which is a broad genome wide technique that focuses on categorization rather than the significance of individual transcripts. Ontologies of those genes were classified as relating to the juvenile hormone pathway, anaerobic metabolism, innate immunity and cold tolerance or as hypothetical proteins. Fig. 4 reveals the categories of gene function of the divergent transcripts. While the majority of upregulated transcripts were classified as hypothetical proteins, diapausing females exhibited an increase in genes related to metabolism, juvenile hormone, and cold resistance while non-diapausing females upregulated genes involved with metabolism and structural fortification. Transcripts were assessed based on fold change differences, significance and ontologies for validation via volcano plot (Fig 5).

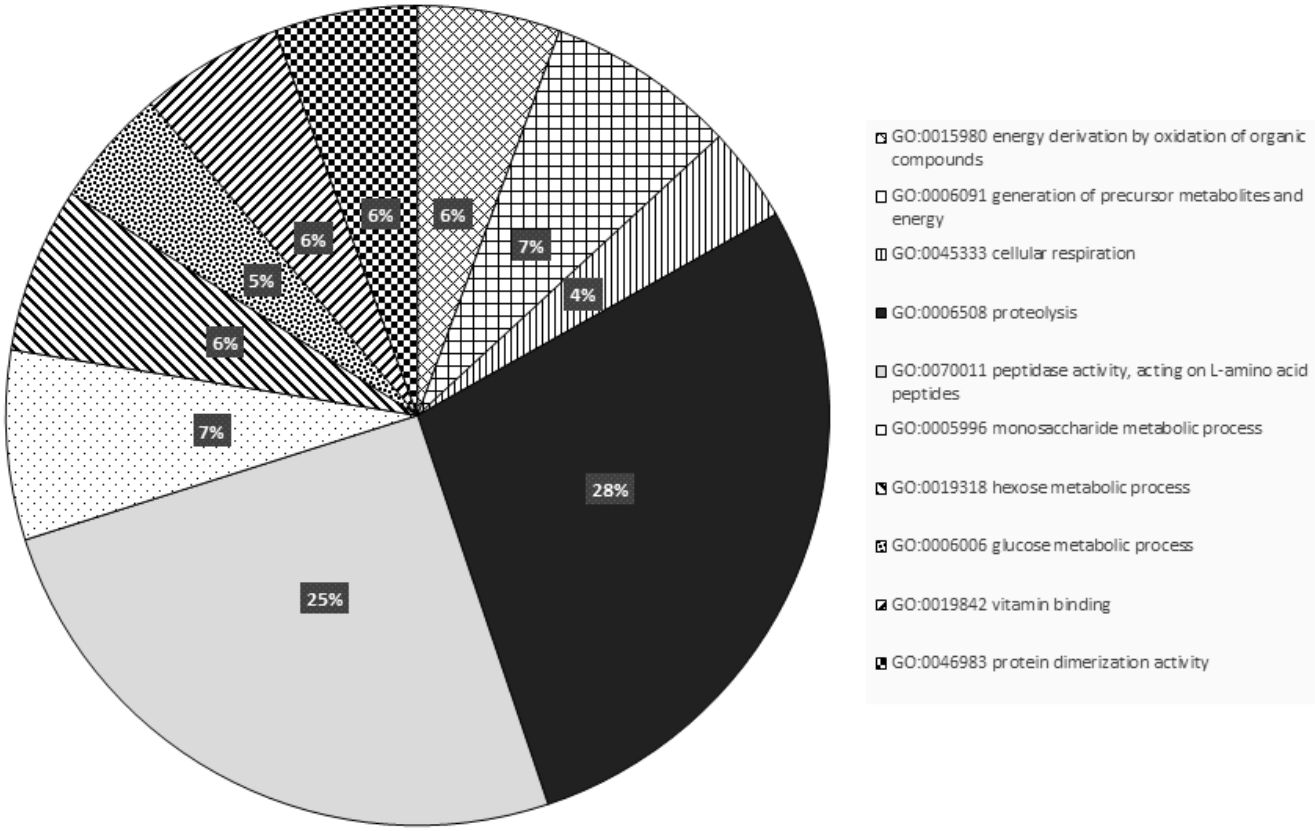

Figure 4. Significantly upregulated transcripts classified by ontology.

Table 2. Differential gene expression profile between diapausing and non-diapausing females of *Cx. pipiens* 7 days after adult eclosion by using Illumina HiSeq 2000 platform.

Gene	Locus	Description	Diapause FPKM*	Non-diapause FPKM	Log_2 Fold Change
CPIJ005941	supercont3.104:135706-144687	ADP,ATP carrier protein 2	1297.29	384.94	-1.75
CPIJ005208	supercont3.108:113365-120827	alpha-amylase	10.17	1.80	-2.50
CPIJ018863	supercont3.1358:78853-81474	hyaluronon glucosaminidase precursor	1.06	0	N/A
CPIJ007618	supercont3.148:25822-56252	alcohol dehydrogenase	4.72	1.16	-2.02
CPIJ019028	supercont3.1589:59519-69546	ran	28.75	86.02	1.58
CPIJ007201	supercont3.163:705539-707969	serine protease	13.35	0	N/A
CPIJ002089	supercont3.21:536022-537672	salivary protein	113.21	45.57	-1.31
CPIJ020177	supercont3.2736:532-9439	nascent polypeptide associated complex alpha subunit	17.98	81.19	2.17
CPIJ020026	supercont3.2812:2034-6343	glycogen debranching enzyme	80.24	27.86	-1.53
CPIJ011172	supercont3.299:36590-73789	dynein beta chain	1.24	4.24	1.77
CPIJ011998	supercont3.346:205841-207234	zinc carboxypeptidase A 1 precursor	73.26	226.73	1.63
CPIJ012251	supercont3.398:300000-334450	troponin C	769.78	275.20	-1.48
CPIJ012704	supercont3.404:144065-145118	pyrroline-5-carboxylate reductase	75.81	19.86	-1.93
CPIJ013040	supercont3.447:1311-4950	glycogen debranching enzyme	64.44	22.82	-1.50
CPIJ000449	supercont3.5:167316-168492	galactose-specific C-type lectin, putative	61.53	11.57	-2.41
CPIJ014348	supercont3.599:237687-239756	sodium/hydrogen exchanger 8	1.325	0	N/A
CPIJ015401	supercont3.675:30064-44889	galactose-specific C-type lectin, putative	101.15	18.29	-2.47

*RNA-seq results expressed in terms of fragments per kilobase of exon per million fragments mapped (FPKM).

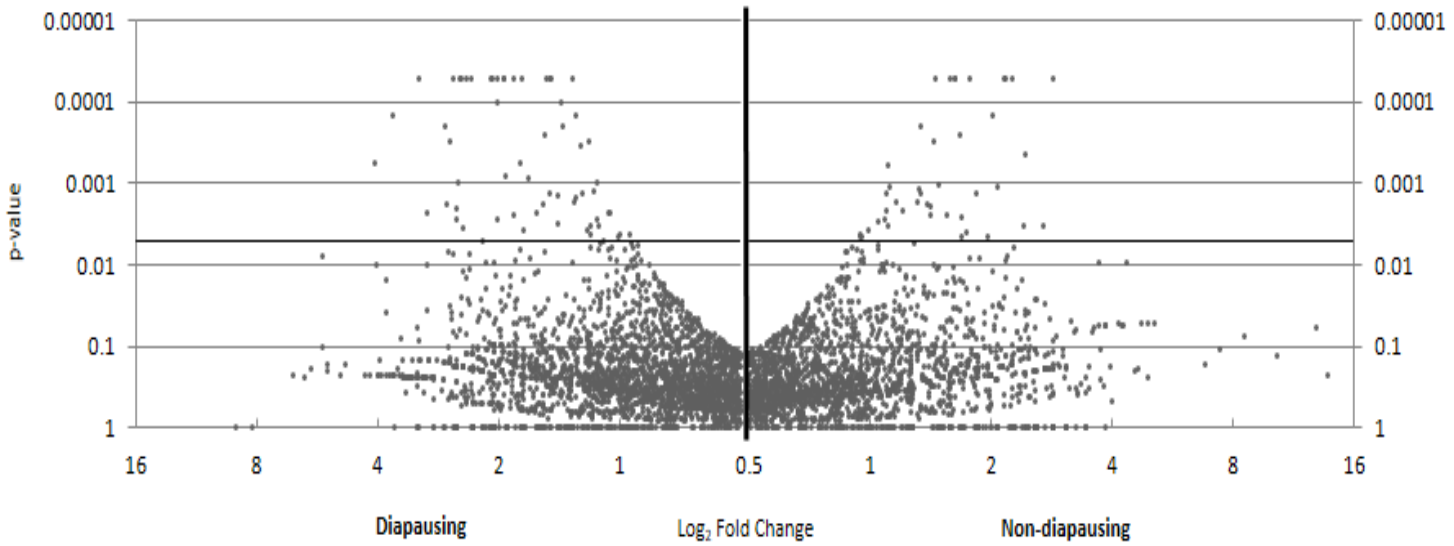

Figure 5. Volcano plot of $\log_2$ fold change vs statistical significance.

Discussion

The increasing prevalence of West Nile Virus in the US emphasizes the need to study the molecular regulation responsible for key adaptations, such as diapause, in *Culex* mosquitoes. Furthermore, the availability of the *Culex quinquefasciatus* reference genome and RNA-seq technology offers a an easy, cost effective method to simultaneously identify and quantify differences in gene expression arising from divergent traits such as diapause in *Cx. pipiens*. We utilized transcriptional profiling to identify potential gene targets for further functional analysis into control of the diapausing female mosquitoes.

The production of juvenile hormone by the *corpora allata* is central to the diapausing program of *Cx. pipiens* (Spielman 1974). The surgical removal of the *corpora allata* from non-diapausing females results in a simulated diapause which is rescued by the application of juvenile hormone (Weaver *et al.* 1998). While this seems to illustrate a simple model of endocrine control, the diapause program has multifaceted downstream effects, which complicate the understanding of this dynamic suite of phenotypes. Here we have examined the differential transcript profiles of *Culex pipiens* for potential links to the diapause syndrome.

Allatostatin is known to halt juvenile hormone production in *Culex pipiens*, and is well documented as a regulator of diapause in other insects as well (Hoffmann *et al.* 1999; Sim & Denlinger 2008; Stay *et al.* 1994). Studies of the cockroach, *Diploptera punctata*, have further demonstrated that allatostatin is associated with a dose-dependent upregulation of digestive enzymes in the insect midgut with allatostatin reactive cells transversing the midgut and basal lamina (Fuse *et al.* 1999; Yu *et al.* 1995). As a

principal carbohydrate metabolizing enzyme in insects, alpha-amylase is specifically responsible for converting starch into maltose in *Diploptera punctata* and is upregulated in the presence of allatostatin (Fuse *et al.* 1999; Khan 1963). Similarly, alpha-amylase is upregulated in diapausing females of *Cx. pipiens*, which points to a potential link between alpha-amylase and allatostation, and further suggests that the digestive enzyme is associated with an increased efficiency of nutrient uptake in diapausing females.

Differential gene expression analysis further revealed an increase of transcripts involved with anaerobic metabolism in diapausing females. In particular, two separately upregulated glyocogen debranching enzymes indicate an increase of glycolysis and gluconeogenesis in diapausing specimens. As diapausing females of *Cx. pipiens* are subjected to periods of fasting, starvation and low energy diets, these anticipatory preparations are unsurprising. Furthermore, increased anaerobic metabolism is a trait found during dormancy of several animals including *S. crassipalpis*, *D. melanogaster*, *C. elegans* and *Wyeomyia smithii* (Baker & Russell 2009b; Emerson *et al.* 2010; Jeong *et al.* 2009; Ragland *et al.* 2010; Wang & Kim 2003) .

The depression of the adaptive immune system and the importance of the innate immune system during diapause is well documented (Nakamura *et al.* 2011). A transcript associated with innate immunity, serine protease, was also upregulated in diapausing females. In addition to being known to be upregulated by juvenile hormone, serine proteases are known to be involved in hemolymph coagulation, synthesis of antimicrobial peptides and melanin, and are responsible for the activation of the immune system in the presence of pathogens (Gorman & Paskewitz 2001; Robich & Denlinger 2005b; Valenzuela *et al.* 2002).

Two galactose-specific C-type lectins, were also upregulated in diapausing females. These results correspond with those observed in the cotton bollworm, *Helicoverpa armigera,* in that the innate immune system is fortified against bacterial and fungal infections in response to a decline in the adaptive immune system during the diapause (Nakamura *et al.* 2011; Zhang *et al.* 2013). It has been previously shown that these cold-tolerant calcium-dependent carbohydrate-binding pattern recognition proteins are able to recognize pathogens and serve as a crucial initiators of the innate immune responses such as phagocytosis, prophenoloxidase activation and hemocyte nodule formation. (Yu *et al.* 2002; Zelensky & Gready 2005). While serine proteases and C-type lectins are also associated with blood feeding in insects, it is unlikely that the upregulation of these enzymes and a third salivary protein in diapausing females is related to anticoagulation as neither trial was offered blood (Charlab *et al.* 1999; Price & Fonseca 2015; Valenzuela *et al.* 2002).

Cold tolerance is a hallmark of the diapausing phenotype, as many of the physiological changes associated with the trait are in direct preparation of winter. In addition to the previously mentioned lectins another cold tolerance gene, pyrroline-5-carboxylate reductase, has been linked with increased cold-shock tolerance in *D. melanogaster* and is responsible for the final step in the biosynthesis of proline (Misener *et al.* 2001; Misener & Walker 2000). Increased cold tolerance is commonly correlated with an upregulation of proline in overwintering insects, and proline may be an important source of metabolic fuel for overwintering (Fields *et al.* 1998; Shimada & Riihimaa 1990; Storey & Storey 1988). The upregulation of pyrroline-5-carboxylate reductase and a hyalurononglucosaminase precursor indicate an anticipatory preparation for overwintering

in diapausing *Culex pipiens*. Furthermore, the upregulation of troponin C suggests a fortification of structural components in diapausing individuals. In soldier termites, the presence of troponin C is associated with the thickening of the cuticle and musculature, which may in turn lead to increase cold in diapausing mosquitoes (Zhou *et al.* 2007b). In contrast, non-diapausing females exhibited an upregulation of zinc carboxypeptidase A 1 precursor. While this transcript may be tied to molting and is upregulated by 20-hydroxyecdysone, which in turn is downstream of juvenile hormone, as both diapausing and non-diapausing females molt it is unknown how this result might be significant (Ote *et al.* 2005).

While many of these ontologies offer insight into the control of diapause, the ubiquitous nature and broad categorizations of several ontological categorizations preclude speculation into the gene function of several candidates. As such, expression differences of ran, sodium/hydrogen exchanger 8, and dynein beta chain, and a nascent polypeptide associated complex subunit have not been addressed in this manuscript, but warrant future investigation. Furthermore, the upregulation of alcohol dehydrogenase in the diapausing females was unexpected this class of enzyme has been tied to juvenile hormone production, but the broad variety of alcohol dehydrogenases prevent us from speculating as to the nature of this specific transcript (Mayoral *et al.* 2009).

The fragmentation of the *Culex quinquefasciatus* genome (N50 = 476 kb), and low chromosomal assignment of the total genome assembly (13%) emphasize the need for further investigation of the *Culex pipiens* complex mosquitoes (Arensburger *et al.* 2010; Naumenko *et al.* 2015). Furthermore, the lack of full physical map limits the prospect of genome wide association study, and many transcripts are hypothetical

proteins of unknown ontology. Fortunately, the ontologies of many of our transcripts correspond with previous studies of diapause and validate the accuracy of our results, but the most exciting aspect of our results are the nine hypothetical proteins and the unknown transcripts associated with the syndrome and their potential as candidate genes of this important suite of adaptations (Table 2).

Table 3. Hypothetical proteins and their positions on the *Culex quinquefasciatus* (Johannesburg strain) reference genome.

Gene ID	**Locus**
CPIJ006495	supercont3.125:61017-61995
CPIJ011623	supercont3.329:415108-415963
CPIJ012164	supercont3.368:82667-83532
CPIJ012185, CPIJ012186	supercont3.374:183520-184852
CPIJ013706	supercont3.514:131855-159297
CPIJ014352	supercont3.599:260569-262212
CPIJ015860	supercont3.679:55208-56318
CPIJ016534	supercont3.772:2934-4658
CPIJ016689	supercont3.792:100918-188468

References

Arensburger P, Megy K, Waterhouse RM, et al. (2010) Sequencing of Culex quinquefasciatus Establishes a Platform for Mosquito Comparative Genomics. Science 330, 86-88.

Baker DA, Russell S (2009) Gene expression during Drosophila melanogaster egg development before and after reproductive diapause. BMC Genomics 10.

Bowen MF (1992) Patterns of Sugar Feeding in Diapausing and Nondiapausing Culex pipiens (Diptera: Culicidae) Females. Journal of Medical Entomology 29, 843-849.

Charlab R, Valenzuela JG, Rowton ED, Ribeiro JM (1999) Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly Lutzomyia longipalpis. Proceedings of the National Academy of Sciences 96, 15155-15160.

Christophers S (1911) The development of the egg follicle in Anophelines. Paludism 1, 73-88.

Control CfD, Prevention (2013) West Nile virus and other arboviral diseases--United States, 2012. MMWR. Morbidity and mortality weekly report 62, 513.

Denlinger D. L., Yocum G. D., L. RJ (2005) Hormonal control of diapause. In: Comprehensive Molecular Insect Science (eds. Gilbert L. I., Iatrou K., S. GS), pp. 615-650. Elsevier, Amsterdam.

Diamond MS (2009) West Nile encephalitis virus infection : viral pathogenesis and the host immune response Springer, New York, NY.

Eldridge BF (1968) The effect of temperature and photoperiod on blood-feeding and ovarian development in mosquitoes of the Culex pipiens complex. Am J Trop Med Hyg 17, 133-140.

Emerson KJ, Bradshaw WE, Holzapfel CM (2010) Microarrays Reveal Early Transcriptional Events during the Termination of Larval Diapause in Natural Populations of the Mosquito, Wyeomyia smithii. PLoS One 5.

Fields PG, Fleurat-Lessard F, Lavenseau L, et al. (1998) The effect of cold acclimation and deacclimation on cold tolerance, trehalose and free amino acid levels in Sitophilus granarius and Cryptolestes ferrugineus (Coleoptera). Journal of insect physiology 44, 955-965.

Fuse M, Zhang JR, Partridge E, et al. (1999) Effects of an allatostatin and a myosuppressin on midgut carbohydrate enzyme activity in the cockroach Diploptera punctata. Peptides 20, 1285-1293.

Gorman MJ, Paskewitz SM (2001) Serine proteases as mediators of mosquito immune responses. Insect Biochemistry and Molecular Biology 31, 257-262.

Harbach RE, Dahl C, White GB (1985) Culex (Culex) Pipiens-Linnaeus (Diptera, Culicidae) - Concepts, Type Designations, and Description. Proceedings of the Entomological Society of Washington 87, 1-24.

Hoffmann KH, Meyerinng-Vos M, Lorenz MW (1999) Allatostatins and allatotropins: Is the regulation of corpora allata activity their primary function? European Journal of Entomology 96, 255-266.

Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1-13.

Jeong PY, Kwon MS, Joo HJ, Paik YK (2009) Molecular Time-Course and the Metabolic Basis of Entry into Dauer in Caenorhabditis elegans. PLoS One 4.

Khan M (1963) The distribution of proteinase, invertase and amylase activity in various parts of alimentary canal of Locusta migratoria L. Indian J. Ent. 25, 200-203.

Lai CH, Tung KC, Ooi HK, Wang JS (2000) Competence of Aedes albopictus and Culex quinquefasciatus as vector of Dirofilaria immitis after blood meal with different microfilarial density. Vet Parasitol 90, 231-237.

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

Mayoral JG, Nouzova M, Navare A, Noriega FG (2009) NADP+-dependent farnesol dehydrogenase, a corpora allata enzyme involved in juvenile hormone synthesis. Proceedings of the National Academy of Sciences 106, 21091-21096.

Meegan JM, Khalil GM, Hoogstraal H, Adham FK (1980) Experimental Transmission and Field Isolation Studies Implicating Culex-Pipiens as a Vector of Rift-Valley Fever Virus in Egypt. American Journal of Tropical Medicine and Hygiene 29, 1405-1410.

Meuti ME, Denlinger DL (2013) The Role of Circadian Clock Genes in the Overwintering Diapause of the Northern House Mosquito, Culex pipiens. Integrative and Comparative Biology 53, E145-E145.

Misener SR, Chen CP, Walker VK (2001) Cold tolerance and proline metabolic gene expression in Drosophila melanogaster. Journal of insect physiology 47, 393-400.

Misener SR, Walker VK (2000) Extraordinarily high density of unrelated genes showing overlapping and intraintronic transcription units1. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1492, 269-270.

Mitchell CJ, Briegel H (1989) Inability of Diapausing Culex pipiens (Diptera: Culicidae) to Use Blood for Producing Lipid Reserves for Overwinter Survival. Journal of Medical Entomology 26, 318-326.

Monath TP (1988) The Arboviruses : epidemiology and ecology CRC Press, Boca Raton, Fla.

Mori A, Romero-Severson J, Severson DW (2007) Genetic basis for reproductive diapause is correlated with life history traits within the Culex pipiens complex. Insect Mol Biol 16, 515-524.

Nakamura A, Miyado K, Takezawa Y, et al. (2011) Innate immune system still works at diapause, a physiological state of dormancy in insects. Biochemical and biophysical research communications 410, 351-357.

Naumenko AN, Timoshevskiy VA, Kinney NA, et al. (2015) Mitotic-Chromosome-Based Physical Mapping of the Culex quinquefasciatus Genome. PLoS One 10.

Ote M, Mita K, Kawasaki H, et al. (2005) Identification of molting fluid carboxypeptidase A (MF-CPA) in Bombyx mori. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 141, 314-322.

Price DC, Fonseca DM (2015) Genetic divergence between populations of feral and domestic forms of a mosquito disease vector assessed by transcriptomics. Peerj 3.

Ragland GJ, Denlinger DL, Hahn DA (2010) Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. Proceedings of the National Academy of Sciences of the United States of America 107, 14909-14914.

Robich R, Denlinger D (2005a) Diapause in the mosquito Culex pipiens evokes a metabolic switch from blood feeding to sugar gluttony. Proc Natl Acad Sci U S A 102, 15912-15917.

Robich RM, Denlinger DL (2005b) Diapause in the mosquito Culex pipiens evokes a metabolic switch from blood feeding to sugar gluttony. Proc Natl Acad Sci U S A 102, 15912-15917.

Sanburg LL, Larsen JR (1973) Effect of photoperiod and temperature on ovarian development in Culex pipiens pipiens. J Insect Physiol 19, 1173-1190.

Shimada K, Riihimaa A (1990) Cold-induced freezing tolerance in diapausing and non-diapausing larvae of Chymomyza costata (Diptera: Drosophilidae) with accumulation of trehalose and proline. Cryo letters.

Sim C, Denlinger DL (2008) Insulin signaling and FOXO regulate the overwintering diapause of the mosquito Culex pipiens. Proc Natl Acad Sci U S A 105, 6777-6781.

Sim C, Denlinger DL (2013) Insulin signaling and the regulation of insect diapause. Front Physiol 4, 189.

Spielman A (1974) Effect of synthetic juvenile hormone on ovarian diapause of Culex pipiens mosquitoes. J Med Entomol 11, 223-225.

Spielman A, Wong J (1973) Environmental control of ovarian diapause in Culex pipiens. Annals of the Entomological Society of America 66, 905-907.

Stay B, Tobe SS, Bendena WG (1994) Allatostatins - Identification, Primary Structures, Functions and Distribution. Advances in Insect Physiology, Vol 25 25, 267-337.

Storey KB, Storey JM (1988) Freeze tolerance in animals. Physiol Rev 68, 27-84.

Tauber MJ, Tauber CA, Masaki S (1986) Seasonal adaptations of insects Oxford University Press, New York.

Trapnell C, Williams BA, Pertea G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511-515.

Valenzuela JG, Pham VM, Garfield MK, Francischetti IMB, Ribeiro JMC (2002) Toward a description of the sialome of the adult female mosquito Aedes aegypti. Insect Biochemistry and Molecular Biology 32, 1101-1122.

Vinogradova AB (2000) Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control Pensoft.

Wang J, Kim SK (2003) Global analysis of dauer gene expression in Caenorhabditis elegans. Development 130, 1621-1634.

Weaver RJ, Edwards JP, Bendena WG, Tobe SS (1998) Structures, functions and occurrences of insect allatostatic peptides 65, 3-32.

Weill M, Malcolm C, Chandre F, et al. (2004) The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. Insect Molecular Biology 13, 1-7.

Yu C, Stay B, Ding Q, Bendena W, Tobe S (1995) Immunochemical identification and expression of allatostatins in the gut of Diploptera punctata. Journal of insect physiology 41, 1035-1043.

Yu XQ, Zhu YF, Ma C, Fabrick JA, Kanost MR (2002) Pattern recognition proteins in Manduca sexta plasma. Insect Biochem Mol Biol 32, 1287-1293.

Zelensky AN, Gready JE (2005) The C-type lectin-like domain superfamily. FEBS J 272, 6179-6217.

Zhang Q, Lu YX, Xu WH (2013) Proteomic and metabolomic profiles of larval hemolymph associated with diapause in the cotton bollworm, Helicoverpa armigera. BMC Genomics 14.

Zhou XG, Tarver MR, Scharf ME (2007) Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. Development 134, 601-610.

COMPILED REFERENCES

Amraoui F, Tijane M, Sarih M, Failloux AB (2012) Molecular evidence of Culex pipiens form molestus and hybrids pipiens/molestus in Morocco, North Africa. *Parasit Vectors* **5**, 83.

Andersen SO, Hojrup P, Roepstorff P (1995) Insect cuticular proteins. *Insect Biochemistry and Molecular Biology* **25**, 153-176.

Arensburger P, Megy K, Waterhouse RM, *et al.* (2010) Sequencing of Culex quinquefasciatus Establishes a Platform for Mosquito Comparative Genomics. *Science* **330**, 86-88.

Arthofer W, Banbury BL, Carneiro M, *et al.* (2015) Genomic resources notes accepted 1 August 2014-30 September 2014. *Mol Ecol Resour* **15**, 228-229.

Atkinson CT, Woods KL, Dusek RJ, Sileo LS, Iko WM (1995) Wildlife disease and conservation in Hawaii: pathogenicity of avian malaria (Plasmodium relictum) in experimentally infected iiwi (Vestiaria coccinea). *Parasitology* **111 Suppl**, S59-69.

Baker D, Russell S (2009a) Gene expression during Drosophila melanogaster egg development before and after reproductive diapause. *BMC Genomics* **10**, 242.

Baker DA, Russell S (2009b) Gene expression during Drosophila melanogaster egg development before and after reproductive diapause. *BMC Genomics* **10**.

Barr AR (1957) The distribution of Culex p. pipiens and C.P. quinquefasciatus in North America. *Am J Trop Med Hyg* **6**, 153-165.

Bartholomay LC, Waterhouse RM, Mayhew GF, *et al.* (2010) Pathogenomics of Culex quinquefasciatus and Meta-Analysis of Infection Responses to Diverse Pathogens. *Science* **330**, 88-90.

Berger J, Suzuki T, Senti KA, *et al.* (2001) Genetic mapping with SNP markers in Drosophila. *Nature Genetics* **29**, 475-481.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology* **46**, 441-469.

Bourguet D, Fonseca D, Vourch G, *et al.* (1998) The acetylcholinesterase gene Ace: a diagnostic marker for the Pipiens and Quinquefasciatus forms of the Culex pipiens complex. *J Am Mosq Control Assoc* **14**, 390-396.

Bowen MF (1992) Patterns of Sugar Feeding in Diapausing and Nondiapausing *Culex pipiens* (Diptera: Culicidae) Females. *Journal of Medical Entomology* **29**, 843-849.

Byrne K, Nichols RA (1999) Culex pipiens in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity (Edinb)* **82 (Pt 1)**, 7-15.

Carey AF, Wang G, Su C-Y, Zwiebel LJ, Carlson JR (2010) Odorant reception in the malaria mosquito Anopheles gambiae. *Nature* **464**, 66-71.

Chan AH, Jenkins PA, Song YS (2012) Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. *Plos Genetics* **8**.

Charlab R, Valenzuela JG, Rowton ED, Ribeiro JM (1999) Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly Lutzomyia longipalpis. *Proceedings of the National Academy of Sciences* **96**, 15155-15160.

Cho WL, Tsao SM, Hays AR, *et al.* (1999) Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. *J Biol Chem* **274**, 13311-13321.

Christophers S (1911) The development of the egg follicle in Anophelines. *Paludism* **1**, 73-88.

Clements AN (1992) *The biology of mosquitoes*, 1st edn. Chapman & Hall, London ; New York.

Cohuet A, Krishnakumar S, Simard F, *et al.* (2008) SNP discovery and molecular evolution in Anopheles gambiae, with special emphasis on innate immune system. *BMC Genomics* **9**.

Control CfD, Prevention (2013) West Nile virus and other arboviral diseases--United States, 2012. *MMWR. Morbidity and mortality weekly report* **62**, 513.

Darsie RF, Ward RA, Chang CC (1981) *Identification and geographical distribution of the mosquitoes of North America, north of Mexico* American Mosquito Control Association, Fresno, Calif.

Demirci B, Lee Y, Lanzaro GC, Alten B (2012) Identification and characterization of single nucleotide polymorphisms (SNPs) in Culex theileri (Diptera: Culicidae). *J Med Entomol* **49**, 581-588.

Denlinger D (2002) Regulation of diapause. *Annual Review of Entomology* **47**, 93 - 122.

Denlinger D. L., Yocum G. D., L. RJ (2005) Hormonal control of diapause. In: *Comprehensive Molecular Insect Science* (eds. Gilbert L. I., Iatrou K., S. GS), pp. 615-650. Elsevier, Amsterdam.

Dethier VG (1957) The sensory physiology of blood-sucking arthropods. *Experimental Parasitology* **6**, 68-122.

Diamond MS (2009) *West Nile encephalitis virus infection : viral pathogenesis and the host immune response* Springer, New York, NY.

Ditzen M, Pellegrino M, Vosshall LB (2008) Insect Odorant Receptors Are Molecular Targets of the Insect Repellent DEET. *Science* **319**, 1838-1842.

Eldridge BF (1966) Environmental Control of Ovarian Development in Mosquitoes of the Culex pipiens Complex. *Science* **151**, 826-828.

Eldridge BF (1968) The effect of temperature and photoperiod on blood-feeding and ovarian development in mosquitoes of the Culex pipiens complex. *Am J Trop Med Hyg* **17**, 133-140.

Emerson KJ, Bradshaw WE, Holzapfel CM (2010) Microarrays Reveal Early Transcriptional Events during the Termination of Larval Diapause in Natural Populations of the Mosquito, Wyeomyia smithii. *PLoS One* **5**.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-185.

Fields PG, Fleurat-Lessard F, Lavenseau L, *et al.* (1998) The effect of cold acclimation and deacclimation on cold tolerance, trehalose and free amino acid levels in Sitophilus granarius and Cryptolestes ferrugineus (Coleoptera). *Journal of insect physiology* **44**, 955-965.

Fonseca DM, Keyghobadi N, Malcolm CA, *et al.* (2004) Emerging vectors in the Culex pipiens complex. *Science* **303**, 1535-1538.

Fuse M, Zhang JR, Partridge E, *et al.* (1999) Effects of an allatostatin and a myosuppressin on midgut carbohydrate enzyme activity in the cockroach Diploptera punctata. *Peptides* **20**, 1285-1293.

Gibson G, Muse SV (2009) *A primer of genome science*, 3rd edn. Sinauer Associates, Sunderland, Mass.

Gibson G, Torr SJ (1999) Visual and olfactory responses of haematophagous Diptera to host stimuli. *Medical and Veterinary Entomology* **13**, 2-23.

Gorman MJ, Paskewitz SM (2001) Serine proteases as mediators of mosquito immune responses. *Insect Biochemistry and Molecular Biology* **31**, 257-262.

Hahn D, Denlinger D (2007) Meeting the energetic demands of insect diapause: nutrient storage and utilization. *J Insect Physiol* **53**, 760 - 773.

Hallem EA, Nicole Fox A, Zwiebel LJ, Carlson JR (2004) Olfaction: Mosquito receptor for human-sweat odorant. *Nature* **427**, 212-213.

Harbach RE, Dahl C, White GB (1985) Culex (Culex) Pipiens-Linnaeus (Diptera, Culicidae) - Concepts, Type Designations, and Description. *Proceedings of the Entomological Society of Washington* **87**, 1-24.

Harbach RE, Harrison BA, Gad AM (1984) Culex (Culex) Molestus Forskal (Diptera, Culicidae) - Neotype Designation, Description, Variation, and Taxonomic Status. *Proceedings of the Entomological Society of Washington* **86**, 521-542.

Hickner PV, Mori A, Chadee DD, Severson DW (2013) Composite Linkage Map and Enhanced Genome Map for Culex pipiens Complex Mosquitoes. *Journal of Heredity* **104**, 649-655.

Hoffmann KH, Meyerinng-Vos M, Lorenz MW (1999) Allatostatins and allatotropins: Is the regulation of corpora allata activity their primary function? *European Journal of Entomology* **96**, 255-266.

Huang da W, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13.

Huang da W, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.

Hubalek Z, Halouzka J (1999) West Nile fever--a reemerging mosquito-borne viral disease in Europe. *Emerg Infect Dis* **5**, 643-650.

Ishida Y, Cornel A, Leal W (2002) Identification and Cloning of a Female Antenna-Specific Odorant-Binding Protein in the Mosquito Culex quinquefasciatus. *Journal of Chemical Ecology* **28**, 867-871.

Jeong PY, Kwon MS, Joo HJ, Paik YK (2009) Molecular Time-Course and the Metabolic Basis of Entry into Dauer in Caenorhabditis elegans. *PLoS One* **4**.

Jobling B (1938) On two subspecies of Culex pipiens L (Diptera). *Transactions of the Royal Entomological Society of London* **87**, 193-216.

Kang D, Sim C (2013) Identification of Culex complex species using SNP markers based on high-resolution melting analysis. *Mol Ecol Resour*.

Kent RJ, Harrington LC, Norris DE (2007) Genetic differences between Culex pipiens f. molestus and Culex pipiens pipiens (Diptera: Culicidae) in New York. *J Med Entomol* **44**, 50-59.

Khan M (1963) The distribution of proteinase, invertase and amylase activity in various parts of alimentary canal of Locusta migratoria L. *Indian J. Ent.* **25**, 200-203.

Kimura M, Darbro JM, Harrington LC (2010) Avian malaria parasites share congeneric mosquito vectors. *J Parasitol* **96**, 144-151.

Knight KL, Stone A (1977) A catalog of the mosquitoes of the world (Diptera, Culicidae).

Kothera L, Godsey M, Mutebi JP, Savage HM (2010) A comparison of aboveground and belowground populations of Culex pipiens (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. *J Med Entomol* **47**, 805-813.

Lai CH, Tung KC, Ooi HK, Wang JS (2000) Competence of Aedes albopictus and Culex quinquefasciatus as vector of Dirofilaria immitis after blood meal with different microfilarial density. *Vet Parasitol* **90**, 231-237.

Lanciotti RS, Kerst AJ, Nasci RS, *et al.* (2000) Rapid detection of west nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. *J Clin Microbiol* **38**, 4066-4071.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**.

Lassiter MT, Apperson CS, Roe RM (1995) Juvenile hormone metabolism during the fourth stadium and pupal stage of the southern house mosquito, Culex quinquefasciatus Say. *Journal of insect physiology* **41**, 869-876.

Leal WS, Barbosa RMR, Xu W, *et al.* (2008) Reverse and Conventional Chemical Ecology Approaches for the Development of Oviposition Attractants for <italic>Culex</italic> Mosquitoes. *PLoS One* **3**, e3045.

Li A, Denlinger DL (2009) Pupal cuticle protein is abundant during early adult diapause in the mosquito Culex pipiens. *J Med Entomol* **46**, 1382-1386.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Mattingly PF (1967) The systematics of the Culex pipiens complex. *Bulletin of the World Health Organization* **37**, 257-261.

Mayoral JG, Nouzova M, Navare A, Noriega FG (2009) NADP+-dependent farnesol dehydrogenase, a corpora allata enzyme involved in juvenile hormone synthesis. *Proceedings of the National Academy of Sciences* **106**, 21091-21096.

McIver SB (1970) Comparative study of antennal sense organs of female culicine mosquitoes. *The Canadian Entomologist* **102**, 1258-1267.

Meegan JM (1979) The Rift Valley fever epizootic in Egypt 1977-78. 1. Description of the epizzotic and virological studies. *Trans R Soc Trop Med Hyg* **73**, 618-623.

Meegan JM, Khalil GM, Hoogstraal H, Adham FK (1980) Experimental Transmission and Field Isolation Studies Implicating Culex-Pipiens as a Vector of Rift-Valley Fever Virus in Egypt. *American Journal of Tropical Medicine and Hygiene* **29**, 1405-1410.

Meuti ME, Denlinger DL (2013) The Role of Circadian Clock Genes in the Overwintering Diapause of the Northern House Mosquito, Culex pipiens. *Integrative and Comparative Biology* **53**, E145-E145.

Michalski ML, Erickson SM, Bartholomay LC, Christensen BM (2010) Midgut barrier imparts selective resistance to filarial worm infection in Culex pipiens pipiens. *PLoS Negl Trop Dis* **4**, e875.

Misener SR, Chen CP, Walker VK (2001) Cold tolerance and proline metabolic gene expression in Drosophila melanogaster. *Journal of insect physiology* **47**, 393-400.

Misener SR, Walker VK (2000) Extraordinarily high density of unrelated genes showing overlapping and intraintronic transcription units1. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1492**, 269-270.

Mitchell CJ, Briegel H (1989) Inability of Diapausing *Culex pipiens* (Diptera: Culicidae) to Use Blood for Producing Lipid Reserves for Overwinter Survival. *Journal of Medical Entomology* **26**, 318-326.

Monath TP (1988) *The Arboviruses : epidemiology and ecology* CRC Press, Boca Raton, Fla.

Mori A, Romero-Severson J, Severson DW (2007) Genetic basis for reproductive diapause is correlated with life history traits within the Culex pipiens complex. *Insect Mol Biol* **16**, 515-524.

Mori A, Severson DW, Christensen BM (1999a) Comparative linkage maps for the mosquitoes (Culex pipiens and Aedes aegypti) based on common RFLP loci. *J Hered* **90**, 160-164.

Mori A, Severson DW, Christenson BM (1999b) Comparative linkage maps for the mosquitoes (Culex pipiens and Aedes aegypti) based on common RFLP loci. *Journal of Heredity* **90**, 160-164.

Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D (2004) Intraspecific nucleotide variation in Anopheles gambiae: new insights into the biology of malaria vectors. *Am J Trop Med Hyg* **71**, 795-802.

Morlais I, Severson DW (2003) Intraspecific DNA variation in nuclear genes of the mosquito Aedes aegypti. *Insect Mol Biol* **12**, 631-639.

Mutebi JP, Savage HM (2009) Discovery of Culex pipiens pipiens form molestus in Chicago. *J Am Mosq Control Assoc* **25**, 500-503.

Nakamura A, Miyado K, Takezawa Y, *et al.* (2011) Innate immune system still works at diapause, a physiological state of dormancy in insects. *Biochemical and biophysical research communications* **410**, 351-357.

Nasci RS, Savage HM, White DJ, *et al.* (2001) West Nile virus in overwintering Culex mosquitoes, New York City, 2000. *Emerging Infectious Diseases* **7**, 742-744.

Naumenko AN, Timoshevskiy VA, Kinney NA, *et al.* (2015) Mitotic-Chromosome-Based Physical Mapping of the Culex quinquefasciatus Genome. *PLoS One* **10**.

Ote M, Mita K, Kawasaki H, *et al.* (2005) Identification of molting fluid carboxypeptidase A (MF-CPA) in Bombyx mori. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **141**, 314-322.

Pelletier J, Leal WS (2009) Genome Analysis and Expression Patterns of Odorant-Binding Proteins from the Southern House Mosquito <italic>Culex pipiens quinquefasciatus</italic>. *PLoS One* **4**, e6237.

Price DC, Fonseca DM (2015) Genetic divergence between populations of feral and domestic forms of a mosquito disease vector assessed by transcriptomics. *Peerj* **3**.

Ragland GJ, Denlinger DL, Hahn DA (2010) Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14909-14914.

Reusken CB, de Vries A, Buijs J, *et al.* (2010a) First evidence for presence of Culex pipiens biotype molestus in the Netherlands, and of hybrid biotype pipiens and molestus in northern Europe. *J Vector Ecol* **35**, 210-212.

Reusken CBEM, de Vries A, Buijs J, *et al.* (2010b) First evidence for presence of Culex pipiens biotype molestus in the Netherlands, and of hybrid biotype pipiens and molestus in northern Europe. *Journal of Vector Ecology* **35**, 210-212.

Roberts DB, Jowett T, Hughes J, Smith DF, Glover DM (1991) The major serum protein of Drosophila larvae, larval serum protein 1, is dispensable. *Eur J Biochem* **195**, 195-201.

Robich R, Denlinger D (2005a) Diapause in the mosquito *Culex pipiens* evokes a metabolic switch from blood feeding to sugar gluttony. *Proc Natl Acad Sci U S A* **102**, 15912-15917.

Robich RM, Denlinger DL (2005b) Diapause in the mosquito Culex pipiens evokes a metabolic switch from blood feeding to sugar gluttony. *Proc Natl Acad Sci U S A* **102**, 15912-15917.

Samantha MP (2007) Nucleotide distribution patterns in insect genomes. *Systemix Reports* **1**.

Sanburg LL, Larsen JR (1973) Effect of photoperiod and temperature on ovarian development in Culex pipiens pipiens. *J Insect Physiol* **19**, 1173-1190.

Schomburg D, Stephan D (1998) Juvenile-hormone esterase. In: *Enzyme Handbook 15* (eds. Schomburg D, Stephan D), pp. 21-26. Springer Berlin Heidelberg.

Severson DW, Brown SE, Knudson DL (2001) Genetic and physical mapping in mosquitoes: molecular approaches. *Annual Review of Entomology* **46**, 183-219.

Shaikevich EV, Vinogradova EB (2004) Molecular genetic methods for the identification of the urban mosquito Culex pipiens pipiens F. molestus (Diptera, Culicidae). *Parazitologiia* **38**, 406-412.

Shapiro AB, Wheelock GD, Hagedorn HH, *et al.* (1986) Juvenile hormone and juvenile hormone esterase in adult females of the mosquito Aedes aegypti. *Journal of insect physiology* **32**, 867-877.

Shimada K, Riihimaa A (1990) Cold-induced freezing tolerance in diapausing and non-diapausing larvae of Chymomyza costata (Diptera: Drosophilidae) with accumulation of trehalose and proline. *Cryo letters*.

Sim C, Denlinger DL (2008) Insulin signaling and FOXO regulate the overwintering diapause of the mosquito Culex pipiens. *Proc Natl Acad Sci U S A* **105**, 6777-6781.

Sim C, Denlinger DL (2013) Insulin signaling and the regulation of insect diapause. *Front Physiol* **4**, 189.

Sirivanakarn S, White GB (1978) Neotype Designation of Culex-Quinquefasciatus Say (Diptera-Culicidae). *Proceedings of the Entomological Society of Washington* **80**, 360-372.

Spielman A (1967) Population structure in the Culex pipiens complex of mosquitos. *Bull World Health Organ* **37**, 271-276.

Spielman A (1974) Effect of synthetic juvenile hormone on ovarian diapause of Culex pipiens mosquitoes. *J Med Entomol* **11**, 223-225.

Spielman A, Wong J (1973) Environmental control of ovarian diapause in Culex pipiens. *Annals of the Entomological Society of America* **66**, 905-907.

Stay B, Tobe SS, Bendena WG (1994) Allatostatins - Identification, Primary Structures, Functions and Distribution. *Advances in Insect Physiology, Vol 25* **25**, 267-337.

Storey KB, Storey JM (1988) Freeze tolerance in animals. *Physiol Rev* **68**, 27-84.

Sundararaman S (1949) Biometrical studies on intergradation in the genitalia of certain populations of Culex pipiens and Culex quinquefasciatus in the United States. *Am J Hyg* **50**, 307-314.

Syed Z, Leal WS (2008) Mosquitoes smell and avoid the insect repellent DEET. *Proceedings of the National Academy of Sciences* **105**, 13598-13603.

Syed Z, Leal WS (2009) Acute olfactory response of Culex mosquitoes to a human- and bird-derived attractant. *Proceedings of the National Academy of Sciences* **106**, 18803-18808.

Takken W (1991) The Role of Olfaction in Host-Seeking of Mosquitoes: A Review. *International Journal of Tropical Insect Science* **12**, 287-295.

Tatar M, Kopelman A, Epstein D, *et al.* (2001) A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107-110.

Tauber MJ, Tauber CA, Masaki S (1986) *Seasonal adaptations of insects* Oxford University Press, New York.

Telfer WH, Kunkel JG (1991) The function and evolution of insect storage hexamers. *Annual Review of Entomology* **36**, 205-228.

Trapnell C, Williams BA, Pertea G, *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515.

Tsai TF, Mitchell CJ (1989) St. Louis encephalitis. In: *The arboviruses: epidemiology and ecology* (ed. Monath TP), pp. 113-143. CRC, Boca Raton, FL.

Turk D, Janjić V, Štern I, *et al.* (2001) Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *The EMBO Journal* **20**, 6570-6582.

Valenzuela JG, Pham VM, Garfield MK, Francischetti IMB, Ribeiro JMC (2002) Toward a description of the sialome of the adult female mosquito Aedes aegypti. *Insect Biochemistry and Molecular Biology* **32**, 1101-1122.

Venter JC, Adams MD, Myers EW, *et al.* (2001) The sequence of the human genome. *Science* **291**, 1304-+.

Vinogradova AB (2000a) *Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control* Pensoft.

Vinogradova EB (2000b) *Culex pipiens pipiens mosquitoes : taxonomy, distribution, ecology, physiology, genetic, applied importance and control* Pensoft, Sofia.

Vinogradova EB (2000c) *Culex pipiens pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control* Pensoft, Moscow.

Wang DG, Fan JB, Siao CJ, *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.

Wang J, Kim SK (2003) Global analysis of dauer gene expression in Caenorhabditis elegans. *Development* **130**, 1621-1634.

Weaver RJ, Edwards JP, Bendena WG, Tobe SS (1998) Structures, functions and occurrences of insect allatostatic peptides **65**, 3-32.

Weill M, Malcolm C, Chandre F, *et al.* (2004) The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. *Insect Molecular Biology* **13**, 1-7.

Williams KD, Schmidt PS, Sokolowski MB (2010) Photoperiodism in insects: molecular basis and consequences of diapause. *Photoperiodism: The Biological Calendar*, 287-317.

Wilton DP, Jakob WL (1985) Temperature-induced morphological change in Culex pipiens. *J Am Mosq Control Assoc* **1**, 174-177.

Wondji CS, Hemingway J, Ranson H (2007) Identification and analysis of single nucleotide polymorphisms (SNPs) in the mosquito Anopheles funestus, malaria vector. *BMC Genomics* **8**, 5.

Wuitschick JD, Karrer KM (1999) Analysis of genomic G + C content, codon usage, initiator codon context and translation termination sites in Tetrahymena thermophila. *J Eukaryot Microbiol* **46**, 239-247.

Yu C, Stay B, Ding Q, Bendena W, Tobe S (1995) Immunochemical identification and expression of allatostatins in the gut of Diploptera punctata. *Journal of insect physiology* **41**, 1035-1043.

Yu XQ, Zhu YF, Ma C, Fabrick JA, Kanost MR (2002) Pattern recognition proteins in Manduca sexta plasma. *Insect Biochem Mol Biol* **32**, 1287-1293.

Zakharkin SO, Gordadze AV, Korochkina SE, *et al.* (1997) Molecular cloning and expression of a hexamerin cDNA from the malaria mosquito, Anopheles gambiae. *Eur J Biochem* **246**, 719-726.

Zelensky AN, Gready JE (2005) The C-type lectin-like domain superfamily. *FEBS J* **272**, 6179-6217.

Zhang Q, Lu YX, Xu WH (2013) Proteomic and metabolomic profiles of larval hemolymph associated with diapause in the cotton bollworm, Helicoverpa armigera. *BMC Genomics* **14**.

Zhou X, Tarver MR, Scharf ME (2007a) Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. *Development* **134**, 601-610.

Zhou XG, Tarver MR, Scharf ME (2007b) Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. *Development* **134**, 601-610.