Mosquito Odorant Receptors: C-terminal Motifs, Subfamily Expansion, and Function

By

Raymond R. Miller

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biochemistry

Committee Members:

Dr. Zhijian (Jake) Tu

Dr. Glenda E. Gillaspy

Dr. Tim J. Larson

Dr. Igor V. Sharakhov

Dr. Liqing Zhang

July 9th, 2008

Blacksburg, VA

Keywords: mosquito, odorant receptor, chemosensory, olfaction, *Or83b*, *Anopheles*, *Aedes*

Mosquito Odorant Receptors: C-terminal Motifs, Subfamily Expansion, and Function

By
Raymond R. Miller

Abstract:

Many insects rely on olfaction as their primary method of interaction with their environment. One of the best examples of this is the olfactory driven host-seeking behavior displayed by female mosquitoes. Although mosquitoes are capable of extracting blood from a variety of hosts many mosquito species show marked preferences for particular host species. Mosquitoes displaying preference for humans above bovines are more likely to be disease vectors. Therefore understanding the molecular basis of this preference is important for public health. These differences may be the result of genetic variations in olfactory signaling components such as mosquito odorant receptors. This hypothesis is supported by several lines of evidence including the highly divergent and lineage-specific nature of this receptor family. Likely these differences are subtle and will be identified in highly focused studies. Even closely related sibling species of mosquitoes can display large behavioral differences. In our current study I have studied several aspects of both *Anopheles* and *Aedes* genus odorant receptors with emphasis on comparing receptors in species that are part of the *Anopheles* genus.

The first goal of this project was to study the insect odorant receptor family for potential sites of heterodimer formation. Numerous studies have shown that insect odorant receptors are involved in detection of odorants. More recent studies have demonstrated that odorant receptors are also involved in protein trafficking and in forming cation channels. Both of these activities involve heterodimer formation between odorant receptors that bind odorants and those that are part of the *Or83b* subfamily. There is little informaiton on how heterodimers are formed and where within the protein heterodimer sites exist. The C-terminal region has been implicated as sites for such heterodimer formation. A hidden markov model based program, Multiple em for motif elicitation (MEME), was used to uncover three motifs in the C-terminus of the odorant receptor peptides from *Anopheles gambiae*, *D. melanogaster*, and *Apis mellifera*. Previous studies have shown that insect odorant receptors are highly divergent between different insect lineages suggesting conservation of these motifs is functionally important. I propose that these motifs are involved in receptor-receptor protein interactions, contributing to the heterodimer formation between *Or83b* subfamily members and other odorant receptors.

The next goal was to identify odorant receptors in closely related mosquito species and compare and contrast them. This was accomplished by using public sequence data of *An. gambiae* and BAC library screening to identify orthologous gene clusters in *An. stephensi* and *An. quadriannulatus*. Although I have identified many

different odorant receptor genes the chapter in this dissertation discusses my work with the *Or2* gene cluster. Multi-species comparison of these orthologous regions in *An. gambiae*, *An. quadriannulatus*, and *An. stephensi* revealed highly conserved gene structure among the OR genes and the discovery of the *An. stephensi Or10x* gene (*AsOr10x*), which is present only in *An. stephensi*. *AsOr10x* showed a different expression pattern than *AsOr2* and *AsOr10*, the other members of this gene subfamily in *An. stephensi*. Therefore *AsOr10x* might be adapting or has adapted a new function. Analysis of the phylogeny and physical location of all known members of the *Or2*/*Or10* gene subfamily in *Anopheles*, *Aedes*, and *Culex* mosquitoes suggest that a few events of gene duplication and loss resulted in the current gene distribution.

The final focus of this project was to develop a method to study the function of mosquito odorant receptors. There is currently no *in vivo* system to study mosquito odorant receptors, and experimental systems pioneered in *D. melanogaster* are not transferable to mosquitoes. I decided to employ a reverse genetics strategy involving the silencing of three *Aedes aegypti* odorant and gustatory receptors of known or suspected function. These gustatory receptors are members of a small subfamily that encode olfactory and not taste receptors. As a preliminary step the expression profiles of these three genes and an additional gustatory receptor were determined using non-quantitative and quantitative RT-PCR. We found that the putative CO_2-detecting

gustatory receptors are expressed in *Ae. aegypti* larvae, and hence these larvae may respond to CO_2, an observation that has not been reported previously.

The purpose of silencing these receptors is to generate a loss-of-function behavior phenotype that will allow for inference of receptor function. Recombinant Sindbis viruses were used to knockdown mRNA levels of these receptors. GFP-expressing recombinant Sindbis viruses were shown to infect chemosensory tissue. Additional viruses containing fragments of receptor genes were found capable of lowering odorant and gustatory receptor mRNA levels. Infected mosquitoes displayed varying levels of gene knockdown with one virus generating supression of mRNA levels to 15.0% of normal. These mRNA levels may not be low enough to generate an unambiguous phenotype. Future experimentation is focused on developing more effective recombinant viruses and identifying characteristics of viruses more effective in receptor gene knockdown. A safe and effective behavior assay setup is needed to test the behavioral responses of these infected mosquitoes. In this study I outline a preliminary behavior assay that is being developed and optimized. When established it will provide a powerful tool in the study of both basic mosquito behavior and phenotype screening of recombinant Sindbis virus-infected mosquitoes.

Acknowledgements:

There are many people I want to thank for their help and guidance in working on this project. First I'd like to thank my advisor Dr. Zhijian (Jake) Tu. He's encouraged and supported me through the entire project. I'm truly grateful for all his help (especially) during times when everything didn't go as well as I'd like. I'd like to thank all the other members of my committee: Dr. Glenda Gillaspy, Dr. Tim Larson, Dr. Igor Sharakhov, and Dr. Liqing Zhang for all the input and time they've contributed by serving on my committee and in their contributing to my professional development. I want to thank all of my lab mates Monica Alvarez, Jim Biedler, Monique Coy, Chevon Dunnings, Song Li, Andy Mead, Yumin Qi, Randy Saunders, and Senay Sengul for all their help, encouragement, ideas, and just good cheer. I want to thank Jim, Monique, and Song for their help in getting me started in the lab, teaching me many of the techniques, and serving as sounding boards for many discussions. I'm grateful to Qi for his help in developing several of the SINVs. Randy has been extremely helpful with providing ALL the mosquitoes I've used for my experiments, his help developing the behavior assays, and last but not least ordering many of the reagents we've used. Dr. Zach Adelman and Dr. Kevin Myles have been immensely helpful in work related to the Sindbis virus for which I'm very grateful.

I also want to thank my family and friends for their long support during this time. I especially want to thank my mom and dad for all the help they've given me, for

pushing me to learn as much as I can, and supporting my drive to seek a PhD. My brother Scott who has always supported me and has always been both a brother and friend. My close friends Jim Hutchinson, Paul Weinwurm, and Phil Sharp for all the support they've given and all the fun we've had. I want to thank anybody who has impacted me that I might have not have mentioned specifically by name.

Table of Contents:

List of Figures:

List of Tables:

Chapter 1. Introduction

1.1: Mosquitoes - General Information

Mosquitoes are part of the family Culicidae, which is part of the order Diptera along with other two wing or "true flies." Mosquitoes have a large impact on global health due to the ability of some mosquito species to spread diseases such as malaria, dengue, yellow fever, and West Nile encephalitis. Mosquitoes eggs are laid near or on a body of water, and then hatch into first instar larvae. Mosquito larvae are multi-segmented aquatic insects covered by pairs of palmate hairs that allow the larvae to orient themselves in water. They occasionally surface in order to breath through an air tube placed on their posterior. Larvae spend all their time searching for food, and as the larvae grow they go through several molting stages or instars. Mosquitoes undergo metamorphosis after the fourth and final instar, pupate, and emerge as adults [1].

Adult mosquitoes have a physiology and life style that involves searching for food, mating, raising eggs, and then depositing them to pass on their genetic heritage. Adults are multi-segmented insects with an abdomen, thorax, and head. Three pairs of legs and a pair of wings are connected to the thorax. The head has a pair of multifaceted eyes, a pair of antenna, a pair of maxillary palp, and a proboscis. The proboscis is an elongated mouthpiece that is used in extracting nectar from plants. Only females have a highly serrated proboscis that allows them to pierce the skin of

vertebrate hosts to extract blood. After emerging from their pupae adults begin to search for food, mates, and blood sources. They do so in response to internal and external cues that stimulate preprogramed behavioral responses. Internal cues include age, circadian rhythms, genetic factors, etc; while external cues are environmental, heat, visual, and olfactory. The most important external cues are olfactory cues that are important in mating [2, 3], feeding [4], and host-seeking [5]. The purpose of blood-feeding is to obtain a source of nutrition for the development of oocytes. After breaking down the blood the females seek out a site for oviposition. It is the process of blood-feeding that makes mosquitoes such a dangerous disease vector. As female mosquitoes extract blood from a host they also inject a mixture of proteins into the host including anti-coagulates. If a female mosquito is infected they also inject disease causative agents into the host. In the following sections I will outline the current knowledge of mosquito host-seeking, mosquito species complexes, and mosquito diseases.

1.1.1: Mosquito Host-seeking Behavior

There are many different genetic, environmental, and behavioral factors that lead to a mosquito species being considered a human disease vector. Vector species tend to be located near population centers while non-vectors reside in more rural areas. Significant factors contributing to vector capacity are behavioral factors such as a

preference for close association with humans. The African malaria mosquito *Anopheles gambiae* prefers to live near, and sometimes in the same dwelling as it's human hosts [6]. The most significant behavioral factor for a vector is the attraction of mosquitoes to human olfactory cues.

It has been established for some time that mosquitoes seek out a host for a blood-meal based on their response to host olfactory cues [5, 7]. Mosquito species are attracted primarily to mammals and birds [5], but they also take blood from reptiles [8]. Mosquito species are able to take blood-meals from a wide range of vertebrate hosts, but many mosquitoes species can show a particular preference for a particular host species. Specialist-feeders seek out one host species as in the case of *An. gambiae*, which feeds almost exclusively on humans [9]. *An. gambiae*'s closely related sibling species *An. quadriannulatus* is also a specialist feeder, but is preferentially zoophilic (animal-loving) [10]. In contrast there are many more opportunistic-feeders that show little or no preference for a host species [5, 11]. *An. stephensi*, an Asian malaria vector and a close relative of the above mentioned *Anopheles* mosquitoes bites humans, but also responds to bovine olfactory cues [12]. *Aedes aegypt*i is another opportunistic feeder although it is a major vector of dengue and yellow fever. Several species of *Culex* prefer birds, but are able to transmit West Nile virus and *Wuchereria bancrofti* to humans perhaps through bridge vectors. A wide range of literature covers the host-preference of these and many more vector and non-vector

species of mosquitoes (see [5] for review). These and future studies on behavior will help determine the basis of mosquito species preferential attraction to humans. With this information new strategies of disease control can be pursued. The most promising is to use such information to develop more effective repellents than the current generation of DEET-based repellents [13].

1.1.2: Mosquito Species Complexes

The resurgence of malaria has shown the pitfalls of traditional methods of disease control. This has driven interest in a new strategy of releasing genetically modified mosquitoes into the wild. These modified mosquitoes would contain transgenes making them refractory to some aspect of disease transmission. By mating with local populations of mosquitoes the transgene will spread through the population. There are many practical and social pitfalls with this strategy [14, 15]. One of the most pressing problems is our lack of knowledge about the population structure of the mosquito species transmitting these diseases. If a vector species is composed of several populations that have limited gene flow then transgenes will have to be driven through multiple populations. This scenario is also relevant where mosquito vector species are in so called species complexes. What follows is a brief introduction into mosquito taxonomy and species complexes in *Anopheles* mosquitoes.

There are at least 3,500 mosquito species distributed throughout the world. Mosquitoes are part of the family of Culicidae and are subdivided into three subfamilies Anophelinae, Toxorhynchitinae, and Culicinae [16, 17]. The Culicinae subfamily has the largest number of species. Toxorhynchitinae mosquitoes do not blood-feed, and hence are not disease vectors. Both Anophelinae and Culicinae subfamilies contain many significant disease vector species. All major malaria vectors are members of the Anophelinae subfamily making studies of these species (both vector and non-vector) of significant impact in combating the global malaria epidemic.

The Anophelinae subfamily is the most basal subfamily of the Culicidae, and is currently subdivided into the *Anopheles* and *Chagasia* genera. The exact details of phylogenetic relationships between mosquitoes within this subfamily remains controversial. For example originally Anophelinae was subdivided into three genera *Anopheles*, *Bironella*, and *Chagasia* with both *Anopheles* and *Bironella* being sister genera [18]. Now it is theorized that *Anopheles* and *Bironella* are both part of the same clade [17], but are paraphyletic [18, 19]. Not surprisingly within each genus the resolution of phylogenetic relationships is even more confusing. Many Anophelinae mosquito species including most of the significant disease vectors have been found to be part of species complexes of several sibling and incipient species [20]. Species within these complexes are morphologically indistinguishable, confusing field identification. To add even more confusion, sibling species are not necessarily

completely reproductively isolated, and have some potential to crossbreed as only male hybrids are sterile. Incipient mosquito species are populations that may be diverging from the main species population and forming genetic barriers against crossbreeding. This population structure may lead to scenarios where either mosquito population control or transgenic efforts are frustrated due to a lack of information.

The *An. gambiae* species complex is the most well studied mosquito species complex (for reviews see [9, 21, 22]). It is composed of at least seven species: *An. gambiae sensu stricto*, *An. quadriannulatus* (A), *An. quadriannulatus* (B), *An. arabiensis*, *An. bwambae*, *An. merus*, and *An. melas* [9]. *An. quadriannulatus* (A) is a specialist feeder displaying a zoophilic host preference. Recently it has been shown there exists a reproductively isolated *An. quadriannulatus* population in Ethiopia named *An. quadriannulatus* B [23]. The most prominent member of the species complex *Anopheles gambiae sensu stricto* (abbreviated henceforth as *An. gambiae*) is the major vector of human malaria in sub-Sahara Africa. Highly anthropophilic it associates near and within human settlements. *An. gambiae* is composed of two molecular forms, which are believed to be undergoing incipient speciation. Field studies have shown strong non-random mating [24, 25], and only a small number of hybrids [26] suggesting that these two populations have or are becoming two new species. Chromosomal [27, 28] and molecular studies support this theory [29]. Of the four species remaining species, *An. arabiensis* is the most significant malaria vector. *An. arabiensis* shows

preference for outdoors host-seeking (exophilicity) in contrast to *An. gambiae*'s endophilicity [30]. In other studies *An. arabiensis* is found to respond as well as *An. gambiae* to human odors but unlike *An. gambia*e also responds to CO_2 [6, 11].

From the above section several themes about mosquito species complexes are apparent. 1) Species complexes are composed of several morphologically indistinguishable species that have the potential to cross-breed. 2) A species within this complex may be two incipient species that have limited gene flow. 3) Even though species complex members are all closely related they display behavioral differences. There are three major reasons to study mosquito species complexes. These complexes have recently formed and there are possibly several new species arising from genetically isolated populations. This provides an excellent case study of evolution where we see adaptations of closely related mosquito species (changes in host-seeking) in response to recent environmental changes (the rise of human population centers). The many disease vectors which are part of these complexes supply the second reason to study species complexes. The last reason is disease vectors display genetic and behavioral traits that allow them to excel as vectors. In comparing them with sibling species that don't display the same vectoral capacity we might identify the molecular basis of these disease vector traits. New methods of disease control may be developed from this information.

1.1.3: Mosquito-borne Diseases

Throughout history mosquito-borne diseases have had a large effect on human civilization. Each year mosquito-borne diseases such as malaria and dengue are responsible for countless deaths and unfathomable misery. Malaria is the worst of these diseases infecting 300-500 million people a year and killing at least a million [31]. There is a rough correlation between poverty and incidence of malaria. Most cases of malaria occur in the poorest countries of sub-Saharan Africa, while malaria has been eliminated from first-world countries in the past century [32]. These diseases are caused by pathogens transmitted by mosquitoes. Whenever infected female mosquitoes bite humans to extract a blood-meal they inject saliva that contains anticoagulants and disease causative agents. Transmission of a pathogenic protist of the *Plasmodium* family results in malaria [32], while viruses of the *Flavivirus* genus cause dengue fever [33].

Long thought to be third-world diseases, an increasing number of cases of mosquito-borne diseases are being reported in places they were formally eradicated from. This may be due to many factors such as the increase of travel/trade between epidemic and non-epidemic areas, global warming increasing the habitable range of mosquitoes, the increased resistance of mosquitoes to insecticides allowing a resurgence of native populations. Dengue fever statistics show a startling increase in the number of cases of clinical infection worldwide with no end in sight [34]. Many

traditional methods of disease control have revolved around control mosquito populations, reduced human-mosquito contact (ex: bed-nets and repellents), and treating diseases with a variety of drugs. These methods have sadly proved inadequate in the current world. Many insecticides such as DDT have become less effective or ineffective due to the increasing resistance of mosquito populations. Newer insecticides are being developed but all current candidates are more expensive to produce and not as effective as DDT once was. A similar situation is present as many disease causative agents are becoming resistant to drugs. Chloroquine resistance is present in most mosquito populations leading to a dramatic increase in malaria mortality [35]. The next generation of drugs has numerous disadvantages similar to those of insecticides. There is little doubt that adaptation of both mosquitoes and parasites will render these solutions transitory. Newer strategies must be developed and one of the first steps will be to continue basic research of mosquitoes to understand more about their biology as it relates to disease transmission.

1.2: Insect Olfaction

1.2.1: Importance of Insect Olfaction

To humans the sense of smell seems a disposable or ethereal sense. We are more likely to interact with our environment at range through sight and sound. Recent studies suggest that we have lost much of our ancestral sense of smell when

compared with mice and even other primates [36, 37]. In contrast olfaction is the most important sense that insects have to interact with their environment. Responses to olfactory cues allow insects to find food sources [38], find mates [39, 40], avoid danger [41], and in the case of blood-feeding mosquitoes find hosts [5]. Olfaction is important in the learning and foraging behavior of honeybees [42, 43]. Insect olfaction may also be involved in speciation. The fly species *Rhagoletis pomonella* feeds, mates, and ovipositions on the native hawthorne or the more recently introduced domesticated apple fruit [44]. It has been found that this fly species is now composed of two races, one that is attracted to hawthorne and one to apples. These two races are estimated to have very limited gene flow [45], and are accumulating fixed differences at several loci [46]. It has been shown that there are significant differences in the olfactory preference of these races to hawthorne or apple odors [47]. Olfaction might be responsible for or contribute to incipient speciation in *R. pomonella*.

The study of insect olfaction has lagged behind that of vertebrate olfaction [48], but recent advances are closing this gap [49, 50]. In subsequent sections I will outline the current knowledge about the organization and signaling in the olfactory system of insects. I will first describe the overall organization of the tissues involved in olfaction. Then I will focus on the molecular components and signaling occurring in the peripheral olfactory system. One of these signaling protein families is the large and diverse family of insect odorant receptors (ORs). In the post-genomics era a great deal

is being discovered about the distribution, regulation, and function of these fascinating receptors.

(1) Olfactory Tissue: Surface and Sensilla

Most insects have two olfactory tissues: a pair of antenna and a pair of maxillary palp (or palpus). In at least one species [51] the proboscis also has a role in olfactory perception in addition to serving as a gustatory organ [52]. The morphology of both olfactory tissues varies between species, and is also sexually dimorphic within species. For example female *Anopheles* mosquitoes have very long antennas with a sparse amount of hairs, while males have shorter and more ornate antennas. Arrays of hair-like structures, known as sensilla, cover the surfaces of these organs. Sensilla house the chemosensory, thermosensory, and mechanosensory systems of insects [52]. *Drosophila melanogaster* sensilla can be broadly classified based on their morphology and wall structure as trichodea, basiconica, and coeloconica [53]. Some sensilla are classified as intermedia sensilla having characteristics of both trichodea and basiconica [53, 54]. These groupings can be further divided based on differences in length and thickness. Not all insect species have the exact same sensilla distribution, and even closely related species have variability in the number of each sensilla type present [55]. In subsequent discussion I will briefly mention these differences for the sake of clarity of presentation and discusses about “typical” sensilla.

(2) Below the Surface of Olfactory Tissue

Based on moth [56] and *D. melanogaster* [52, 57] structural studies we know a great deal about the organization of typical insect olfactory (and non-olfactory) sensilla. Sensilla are composed of a protective outer coat covering several olfactory specific cells. Between the coat and the cells is the sensilla lymph, a hydrophilic environment. The coat is an extension of the antenna (or palp or proboscis) cuticle, and along this coat are series of holes or pores that allow odorants to pass through the cuticle and enter the lymph. Beneath the cuticle are support (or auxiliary) cells and olfactory sensory neurons (OSNs). The three types of support cells are the tormogen, trichogen, thecogen, and a typical sensilla can have a total of three to four support cells. The thecogen forms a protective sheath around the OSN(s) cell body. All of these cells have very well developed golgi apparatus and rough endoplasmic reticulum along with many coated pits and vesicles indicating significant function in protein synthesis and trafficking. It is hypothesized that these cells excrete odorant-binding proteins (OBPs) [56, 58], and odorant degrading enzymes (ODEs) into the lymph in addition to maintaining the composition of the lymph [59]. The OSNs are the cells where actual olfactory signaling occurs, and each sensilla can have anywhere from one to four neurons depending on the sensilla type [57]. OSNs are known to express two olfactory protein families odorant receptors (ORs) [60-63], and sensory neuron membrane proteins (SNMPs) [64-66]. OSNs are elongated cells with dendrites extending into the

lymph and axons that project to the antennal lobe (AL). Odorants interact with ORs on the surface of a dendrite, and the subsequent signal is passed through the neuron axon to the AL.

The AL is an organ composed of spherical units of neuropil called glomeruli, and it is to these units that that OSN axons form synapses. The AL seems to be structurally and possibly functionally equivalent to the vertebrate olfactory bulb. The exact purpose of both organs remains unknown. In fruit flies the axons of all neurons expressing one (or coexpressing two) particular receptor(s) synapse to the same glomerulus [54, 67]. There is one case where this may not be true, and axons from different neurons synapse to the same glomerulus [67]. The axons of sensilla classes segregate to distinct regions in the AL. For example the *D. melanogaster* antennal basiconica sensilla neurons synapse to the medial region of the AL [54]. There is also some functional subdivision of the AL with neurons that are broadly tuned to many different odors synapsing to the dorsal and medial regions, while more fine tuned neurons synapse to the ventral and lateral regions [67]. There is a complex network of local neurons and projection neurons connecting the different glomeruli and the AL to the higher centers of insect brain including the mushroom bodies and the lateral horn. It is beyond the focus of this introduction to discuss these topics.

1.2.2: Insect Olfactory Signaling

The detection of odorants and the transmission of the subsequent signal to the higher brain centers is a process known as olfactory signaling [49]. The pathway can be broadly subdivided based on location into peripheral or central olfactory signaling. The sensilla make up the peripheral, while the AL and higher brain centers make up the central system. The purpose of this section is to review the current information about the signaling taking place on the surface of and in OSN(s). Much of what is hypothesized to take place in insects is based on the well studied mammalian olfactory signaling pathway [48].

When an insect encounters odorants these odorants first enter through cuticle pores that lead to the sensillum lymph. This a very hydrophilic environment which may hinder the ability of some odorants to traverse it. The diverse family of soluble proteins called odorant binding proteins (OBPs) are known to bind odorants [68]. It is hypothesized that these proteins transport odorants through the lymph, protect odorants from chemical or enzymatic damage, and add specificity to olfactory signaling [69]. Most odorants may not require OBPs. To date only one OBP has been shown to be required for detection of a specific group of chemicals [58].

Regardless of mechanism of delivery, these odorants arrive near the OSN membrane surface where they can be bound by cognant ORs. This binding is theorized to cause a conformational shift in the receptor structure leading to activation

of a hypothetical G-protein signaling cascade similar to that occurring in mammalian olfaction [48]. Gα subunits expressed in the head [70], along with arrestins [71], and other G-protein signaling related components have been identified in insects [49]. Additionally the second messenger created in these downstream events is unknown, but evidence strongly suggests inositol 1,4,5-trisphosphate (IP3) [72, 73]. The second messenger should open up ion channels leading to depolarization of the OSN, and this signal is transmitted to AL and then the higher brain centers where a response is coordinated. The G-protein complex reassembles leading to a release of the odorant, which is removed by an ODE leaving the system back at rest. This possible signaling cascade is compelling and supported indirectly by many studies to date, but two studies have recently cast doubt on this model [74, 75]. In both studies *D. melanogaster* ORs were shown to form odor activated cation channels, and generate changes in action potential independent of other ion channel proteins. One of the studies goes further and suggests that ORs do not require G-proteins [74], while the other claims they can utilize them [75]. Further studies may resolve previous data with this newer information. In the following sections I will discus the work being done on insect ORs.

(1) Insect Odorant Receptors: Discovery

Insect odorant receptors (ORs) were first discovered in *D. melanogaster* through bioinformatic approaches [60, 61]. In both investigations computer algorithms were used on existing *D. melanogaster* genome sequence to identify putative open reading frames encoding proteins ~200-900 amino acids long with seven-transmembrane domains. A group of these genes were found expressed specifically in the antenna and maxillary palp, but not in other parts of the head or body. A third group identified similar genes by screening an antenna/maxillary palp cDNA library [62]. The completion of the *D. melanogaster* genome has lead to the identification of a total of 60 OR genes encoding 62 OR proteins [76]. These ORs have been given a unified naming system [76, 77]. Subsequent studies have succeeded in determining the expression pattern of most of these genes in adult and young fruit flies through *in situ* hybridization [54, 63, 78, 79]. There has also been a concerted effort to determine the function of these proteins using electroantennograms (EAGs) [80] in conjunction with the "empty neuron" system [81] and heterologous cell expression systems [82-84]. The identification of all *D. melanogaster* ORs coupled with the burst of whole genome sequencing of insects has lead to the identification and preliminary characterization of non-*Drosophila* ORs.

(a) Nomenclature of Insect Odorant Receptors

As in many different fields there is an overabundance of confusion when comparing similar things (be they proteins, pathways, or genes) in different species due to the different nomenclature systems used by different groups. Insect OR naming systems are consistent within one species but not between species. For example members of the *Or83b* family are named *Or83b*, *Or2*, or *Or7* depending on the insect lineage or species. Within species the numbering system tends to reflect what order ORs were annotated, and not a common physical or phylogenetic relationship [85-87]. For example *An. gambiae Or3*, *Or4*, and *Or5* are closely related and physically clustered, while *Or1* and *Or2* are neither closely related or reside in the same physical location. In subsequent discussion we will use the "*SpOrX*" nomenclature in describing specific OR genes (protein product will not be italicized). This format is a two letter genus and species name abbreviation "Sp" followed by "Or" to designate an odorant receptor, and "X" which will stand for the specific number of the OR in that species as reported by the discoverer. This nomenclature system is the one used by several other groups describing insect ORs [86, 88, 89].

(b) Insect Odorant Receptor Family

Insect ORs are a large and diverse family of seven-transmembrane proteins with 300-450 amino acid residues. In the past they have been thought to be GPCRs given they have seven-transmembrane structures, and they serve the same role that GPCRs do in vertebrate species [85]. Subsequent work has identified some of the "usual suspects" of G-protein signaling expressed in olfactory tissues lending weight to this assertion [49, 70]. Recently this view has been disputed [90] as insect ORs lack homology to any other non-insect GPCR family [91], their membrane topography is inverted when compared to other GPCRs [92, 93], and with some exceptions [94, 95] GPCRs do not form heterodimers while insect ORs do and may require it for protein trafficking and olfactory signaling [82, 92, 93]. Heterodimers of ORs form cation channels that are opened upon stimulus with specific odors implying ORs may be both receptors and signaling components [74, 75]. Given these data insect ORs might be a remarkable case of convergent evolution with vertebrate ORs.

The members of this insect protein family are highly divergent both within [76, 96] and between species [85, 87]. There is very little sequence conservation between members of this family. The only conservation seems to be several amino acid residues in the extreme C-terminal end of these receptors [60-63, 76], and a highly conserved tryptophan residue [96] (see Chapter 2). As more non-*Drosophila* ORs have been identified it has become increasingly clear that there are very few OR genes

conserved between different species with the notable exception of members of the *Or83b* family. Five additional insect species *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera*, *Bombyx mori*, and *Tribolium castaneum* have had their genome sequenced leading to the identification of many (if not all) of their respective OR genes [85-87, 97, 98]. Currently there are 79 *An. gambiae*, 131 *Ae. aegypti*, 120 *Ap. mellifera*, 48 *B. mori*, and 341 *Tr. castaneum* OR genes annotated. In comparing and contrasting ORs it is apparent that there has been a significant independent expansion of certain OR gene subfamilies within all of these species. For example *B. mori* and the moth *Heliothis virescens* share a subfamily of pheromone receptors not present in other non-moth species [97]. Even comparing the two mosquito species *An. gambiae* and *Ae. aegypti* has revealed a great deal of differences in overall OR gene content [87]. A handful of receptors from other insects have also been identified and confirm the trends mentioned above [89, 99]. ORs are highly divergent both within a species and between species, and that ORs have undergone lineage specific gene expansion.

(c) Insect Odorant Receptor Expression and Gene Regulation

Insect ORs are expressed in the chemosensory tissue of adult insect antenna, maxillary palp, and in one species, the proboscis [51]. Some of these receptors are also expressed in the much simpler larvae olfactory tissue [79]. *D. melanogaster in situ* hybridization studies have shown that most ORs are expressed in a small subset of

OSNs in a pattern conserved between individual animals [60-63]. The exception is one *D. melanogaster* receptor: *DmOr83b*, which is expressed in about 70-80% of antennal and all maxillary palp OSNs [63, 100]. Presumably all OR protein is localized to the dendrites of OSNs [100]. *D. melanogaster* ORs seem to have no function in neuron development [100], and as shown by the use of the "empty neuron" system ORs can be placed into non-native OSNs with no observed ill effects on the neurons [81]. Each OSN subset synapses to only one specific glomerulus in the AL [54] although in one case two separate neuron populations expressing *DmOr67d* and *DmOr82a* converge on a single glomerulus (VA6) [54, 67]. Taken together these data suggest that insects follow (with the exception of *Or83b*) "the one receptor one neuron" principle hypothesized to exist in the mammalian olfactory system. This principle in insects could be further expanded to the "one receptor one glomerulus" rule [101]. There are some notable exceptions to this rule as some OSNs co-express two or more ORs in addition to *Or83b* [102]. It is important to note that all of the above data comes exclusively from *D. melanogaster* studies.

An increasing body of work has begun to illuminate mRNA expression patterns of ORs in moths, mosquitoes, and beetle [88, 89, 98, 99, 103, 104]. In these species ORs are also expressed in the antenna and maxillary palp in small subsets of OSNs. A recent study has shown expression of a few *An. gambiae* ORs in the proboscis, but it is not yet known if other species exhibit this phenomena [51]. It is established that at

least some ORs are expressed in larvae and pupae but most ORs are expressed in adults. [87, 98]. ORs can be up-regulated or down-regulated in response to major life cycle events such as blood-feeding in mosquitoes [88]. There are several sexually dimorphic expressed ORs [88, 89, 104]. What is also important to point out is the expression of insect ORs is highly stereotyped in a way not seen in mammals. One individual has the same OR distribution as another individual. This perhaps reflects a difference in the way vertebrate and invertebrate ORs are regulated.

Until recently there was little known as to how ORs are regulated in insects. For a long time the only information available was that the transcription factor Acj6 was necessary for expression of a group of *D. melanogaster* ORs [60]. In the most comprehensive study to date Ray et al. has identified several regulatory elements of some *D. melanogaster* maxillary palp ORs [105]. They found via deletion reporter constructs that there are several trans-regulatory elements upstream of these genes involved in tissue and neuron specificity. For example motif Oligo-1 represses expression of maxillary palp ORs in the antenna, while Dyad-1 promotes expression of OR genes in the maxillary palp. One neuron-specific regulatory factor the pb2A-2 element was found to be necessary for expression of receptors in the pb2A neurons. These regulatory elements lie within <450 bp of the transcriptional sites of the genes studied. ORs in general may have very small and compact regulatory regions and this might explain the success of OR gene duplication by non-homologous recombination.

In mosquitoes, honeybees, and beetles there are large numbers of OR gene clusters and tandem arrays [85, 87].

(d) Or83b: The Conserved Insect Odorant Receptor

One of the most striking things about insect ORs is the high level of diversity found in the family between distantly related [85, 86, 99], and even closely related insect lineages [87]. This suggests each lineage if not each species may have a finely tuned sense of smell specifically adapted for their specific ecological niche. Therefore it is surprising that one OR gene is conserved among a wide group of insects including *D. melanogaster* [100], several species of mosquitoes [106-108], *Apis mellifera* [86], *Bombyx mori* [89], *Heliothis virescens* [99], and *Tribolium casterium* [98]. The *Or83b* gene family was first discovered in *D. melanogaster* (*DmOr83b*), and members of this family have been found to have characteristics that are not present in other *D. melanogaster* receptors. *DmOr83b* is significantly larger than the other receptors at 486 amino acids, has very low sequence identity to other receptors except in one region [62], and *in situ* hybridization experiments show expression in ~70-80% of antennal and 100% of palp OSNs [63, 100]. *Or83b* orthologs also have a similar ubiquitous expression profile in the chemosensory tissue of other insects according to both RT-PCR and *in situ* hybridization studies [99, 106-108]. These data suggest the *Or83b* family is critical to insect olfaction.

Two models have been proposed for the function of *Or83b*. One involves *Or83b* acting as a generalized receptor responding to many different olfactory cues. The second theory suggests *Or83b* being a generalized cofactor for some aspect of olfactory signaling. This later theory has been supported by several recent studies. The first study by Larsson and co-authors showed that *D. melanogaster* knockouts of *DmOr83b* were not able to respond to olfactory cues [100]. In these knockouts at least three non-*DmOr83b* ORs are localized to the OSN cell body, but not the dendrites where they are found in wild type flies. Both of these phenotypes could be rescued with a copy of *DmOr83b*. Taken together these data suggest a role for the *Or83b* family in properly localizing some if not all ORs to the dendrites of OSNs allowing them to interact with odors. Subsequent studies have focused on what interaction *DmOr83b* has with non-*DmOr83b* ORs. In two studies it has been established that *DmOr83b* is capable of forming heterodimers with other non-*DmOr83b* ORs both *in vitro* [82] and *in vivo* [92]. The exact site of this heterodimer formation is unknown although it is probably in the C-terminal domain of insect ORs [92] (see Chapter 2). Presumably this heterodimer formation is required for transportation of other ORs to the dendrites of OSNs although this theory has yet to be tested explicitly, and the pathway involved in trafficking ORs has yet to be uncovered.

Although these studies do indicate a function of *Or83b* specifically in membrane trafficking of other OR proteins there may be other functions. Two possible additional

functions are assisting ORs in binding of their ligands, and recruiting/binding G-proteins. Beyond the conserved amino acids located in the C-terminal region of ORs there are no other motifs that non-*Or83b* receptors might use to recruit G-proteins. These receptors may need another protein to bind the G-proteins. It is worth noting the unusually long length of *DmOr83b* (and other *Or83b* proteins) is due to a predicted extended second intracellular loop [76]. This loop may in fact have motifs involved in binding G-proteins.

(e) Determining the Function of Insect Odorant Receptors

Initial studies of insect ORs focused on identification of new receptors and on the expression profile of these receptors [60-63]. These results have driven further studies that identify OR gene regulatory elements [105], and determine the function of OR genes [80]. The most successful approach in determining OR function has utilized a single-unit electrophysiology method known as electroantennograms [109-111]. With this technique an electrode is inserted into the base of a sensilla, and then the organism is exposed to an odor. If the odor is detected the neuron(s) in the sensilla this leads to an excitatory response. Based on the amplitude of the response each neuron in the sensilla can be distinguished from it's neighbor. Electroantennograms allow us to determine the odor(s) that a OSN responds to. Coupled with this technique are three approaches for deducing OR function. The first is to knockout the receptor of

interest, and see what effect this has on the neuron action potential when exposed to odorants. The second technique is to express an OR gene in heterologous expression systems such as *Xenopus* oocytes. The third and more elegant approach is to use the empty neuron system [81]. Both techniques require knowledge of which neuron expresses which receptor, a criteria met in *D. melanogaster* where these techniques have been well-established.

In their study of *DmOr22a/b*, Dobritsa and co-authors developed a knockout fruit fly strain lacking both ORs [81]. Electroantennograms targeted to the ab3A neurons, where these receptors are expressed, found that these neurons were unresponsive to normal olfactory cues. They also found that by using the *DmOr22a* promoter they could drive expression of *DmOr47a* to ab3A neurons. Expression of *DmOr47a* in ab3A neurons leads ab3A neurons to be sensitive to the exact same odorants as the ab5B neurons where *DmOr47a* is normally expressed [54]. The major significance of this work is that some ORs can be expressed in non-native neurons and still function. This suggests ORs can function independent of other chemosensory proteins, but this is directly contradicted by one study showing the requirement of an OBP for proper odorant detection [58]. Since this report many different *D. melanogaster* ORs and some non-*Drosophila* ORs have been expressed in this system. Coupled with electroantennograms, and some knowledge as to which neuron expresses which receptor allows each receptor's ligand(s) to be identified. One

particularly elegant study used these techniques to deorphanize virtually all *D. melanogaster* ORs [80]. Subsequent studies focusing on knockout and behavioral assays will in time validate the results of these ground-breaking studies in determining the function of odorant receptors.

(2) Non-Odorant Receptor Driven Olfaction: Gr21a and Gr63a Sub-family

ORs are just one of several families of possible GPCRs found in insects. The most closely related family to the ORs are insect taste or gustatory receptors (GRs) [76]. *Drosophila* GRs are similar to ORs in both overall size, some very limited sequence similarity at the C-terminal end, and in their lack of sequence similarity with other known GPCR families [112]. *D. melanogaster* GRs were discovered through the same methods as ORs, but phylogenetic analysis has placed GRs and ORs as separate groups with *Or83b* being a possible ancestral copy linking the two families [14, 76]. Initial RT-PCR and *in situ* hybridization experiments have demonstrated that the majority of genes in this family are expressed in neurons located in the proboscis, legs, and wings supporting a role for this family in gustatory (taste) perception [14, 112, 113]. GR-expressing neurons are found to synapse to the subesophageal ganglion (SOG), a insect taste center, further supporting this hypothesis [113]. Since then the *D. melanogaste*r GRs that detect caffeine [114], trehalose [115], and several other sugars [116-118] have been identified. Not all of these GRs are involved in taste perception.

Three *D. melanogaster* GRs were found expressed in antennal neurons [14]. These particular antennal neurons synapse to the AL, not the SOG, suggesting a role for some or all of these receptors in olfactory, not gustatory perception [63, 100].

One of these antennal-expressed GRs: *DmGr21a* is localized on ab1c neurons, and these neurons synapse to a specific glomerulus in the ventral region of the AL. The ab1c neurons have electrophysiologic response to CO_2 [119], and are required for CO_2 avoidance behavior [41]. Two subsequent studies established that *DmGr21a* and the closely related *DmGr63a* are coexpressed in ab1c neurons, and that expressing both GRs in the empty neuron system [81] confers CO_2 responsiveness [120, 121]. If either GR is not present there is an abolition of this response. These two GRs are not found localized with or require *DmOr83b* to function suggesting they are in fact true GRs and not ORs. Lu and co-authors have carried this work further in the mosquito species *An. gambiae* [122]. They found that the *Gr21a* subfamily in mosquitoes and other non-*Drosophila* insect species is composed of three genes. The three *An. gambiae* GRs *AgGr22-24* were found co-expressed in the same neuron in the CO_2 sensing organ of mosquitoes the maxillary palp. Electroantennograms coupled with the empty neuron system support a role in CO_2 detection for all three receptors, but *AgGr23* may not be as necessary as the other two receptors. Taken together these studies have identified a subfamily of GRs that are involved in odor detection. There has yet to be any report of any OR protein responding to a non-olfactory cue or

involved in a non-olfactory process. GRs not only have been found to respond to olfactory cues, but are also found expressed in non-chemosensory neurons [123]. These data suggest the GR family may be a more generalized receptor family in comparison with ORs.

1.3: Project Goals

Female mosquitoes seek out vertebrate host species to extract blood for oocyte development. Host-seeking behavior is a complex process in part involving attraction of female mosquitoes to host olfactory cues. Those mosquitoes attracted preferentially to human host odors tend to be more effective disease vectors. The exact molecular basis of preference for human or non-human odors remains unknown. Closely related mosquito species such as members of the *An. gambiae* species complex display different host-seeking preferences. This suggests that subtle genetic differences are involved in host preference. I hypothesize that differences in distribution, sequence, and regulation of mosquito odorant receptors leads to observed differences in host preference. One objective of this project is to study odorant receptors among different mosquito species to uncover the genetic variations that may affect their behavioral differences. In addition to molecular cloning, comparative genomics, and expression analysis, we are interested in establishing a systematic approach to study the function of mosquito chemosensory receptors

The specific aims of this project have been:

1) Identify the potential site(s) of heterodimer formation (see Chapter 2). This chapter stemmed from a comparative analysis of all odorant receptor protein sequences from

An. gambiae and two other divergent insect species which revealed conserved sequence features of potential functional importance.

2) Compare the genomic and evolutionary features of *An. gambiae* OR gene clusters in the mosquito species *An. stephensi* and *An. quadriannulatus* via bacterial artificial chromosome (BAC) library screening and sequencing (see Chapter 3).

3) Establish an *in vivo* method for determining the function of mosquito odorant and gustatory (olfactory) receptors (see Chapter 4).

Abbreviations

antennal lobe, AL; electroantennograms, EAG; G-protein coupled receptor, GPCR; gustatory receptor, GR; inositol 1,4,5-trisphosphate, IP3; odorant binding protein, OBP; odorant degrading enzyme, ODE; odorant receptor, OR; olfactory sensory neuron, OSN; sensory neuron membrane protein, SNMP; subesophageal ganglion, SOG

1.4: Bibliography

1. Eldridge BF: **Mosquitoes, the Culicidae.** In *Biology of Disease Vectors* Edited by Marquardt WC. Amsterdam: Elsevier Academic Press; 2005:95-111.
2. Nijhout H, Craig G: **Reproductive isolation in Stegomyia mosquitoes. III evidence for a sexual pheromone.** *Entomologia Experimentalis et Applicata* 1971, **14**:399-412.
3. Cabrera M, Jaffe K: **An aggregation pheromone modulates lekking behavior in the vector mosquito Aedes aegypti (Diptera: Culicidae).** *J Am Mosq Control Assoc* 2007, **23**:1-10.
4. Foster WA: **Mosquito sugar feeding and reproductive energetics.** *Annu Rev Entomol* 1995, **40**:443-474.
5. Takken W, Knols BGJ: **Odor-mediated behavior of Afrotropical malaria mosquitoes.** *Annu Rev Entomol* 1999, **44**:131-157.
6. Service MW, Joshi GP, Pradhan GD: **A survey of Anopheles gambiae (species A) and An. arabiensis (species B) of the An. gambiae Giles complex in the Kisumu area of Kenya following insecticidal spraying with OMS-43 (Fenitrothion).** *Ann Trop Med Parasitol* 1978, **72**:377-386.
7. Rudolfs W: **Chemotropism in mosquitoes.** *Bull NJ Agric Exp Stn* 1922, **367**:4-23.
8. Rodrigues SC, Maruniak JE: **Blood meal identification from mosquitoes collected at a commercial alligator farm.** *J Am Mosq Control Assoc* 2006, **22**:557-560.
9. White GB: **Anopheles gambiae complex and disease transmission in Africa.** *Trans R Soc Trop Med Hyg* 1974, **68**:278-301.
10. Zwiebel LJ, Takken W: **Olfactory regulation of mosquito-host interactions.** *Insect Biochem Mol Biol* 2004, **34**:645-652.
11. Dekker T, Takken W: **Differential responses of mosquito sibling species Anopheles arabiensis and An. quadriannulatus to carbon dioxide, a man or a calf.** *Med Vet Entomol* 1998, **12**:136-140.
12. Takken W, Dekker T, Wijnholds YG: **Odor-Mediated Flight Behavior of Anopheles gambiae Giles Sensu Stricto and An. stephensi Liston in Response to CO_2, Acetone, and 1-Octen-3-ol (Diptera: Culicidae).** *J Insect Behav* 1997, **10**:395-407.

13. Justice RW, Biessmann H, Walter MF, Dimitratos SD, Woods DF: **Genomics spawns novel approaches to mosquito control.** *BioEssays* 2003, **25**:1011-1020.
14. Scott K, Brady R, Cravchik A, Morozov PS, Rzhetsky A, Zuker C, Axel R: **A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila.** *Cell* 2001, **104**:661-673.
15. Knols BGJ, Bossin HC, Mukabana WR, Robinson AS: **Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world.** *Am J Trop Med Hyg* 2007, **77**:232-242.
16. Besansky NJ, Fahey GT: **Utility of the white gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae).** *Mol Biol Evol* 1997, **14**:442-454.
17. Harbach RE, Kitching IJ: **Phylogeny and classification of the Culicidae (Diptera).** *Syst Ent* 1998, **23**:327-370.
18. Krzywinski J, Wilkerson RC, Besansky NJ: **Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence.** *Syst Biol* 2001, **50**:540-556.
19. Sallum M, Schultz T, Foster P, Aronstein K: **Phylogeny of Anophelinae(Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences.** *Syst Ent* 2002, **27**:361-382.
20. Collins FH, Paskewitz SM: **A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic Anopheles species.** *Insect Mol Biol* 1996, **5**:1-9.
21. Coluzzi M, Sabatini A, Petrarca V, Di Deco M: **Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex.** *Trans R Soc Trop Med Hyg* 1979, **73**:483-497.
22. Coetzee M, Craig M, le Sueur D: **Distribution of African malaria mosquitoes belonging to the Anopheles gambiae complex.** *Parasitology today (Personal ed)* 2000, **16**:74-77.
23. Hunt RH, Coetzee M, Fettene M: **The Anopheles gambiae complex: a new species from Ethiopia.** *Trans R Soc Trop Med Hyg* 1998, **92**:231-235.
24. Tripet F, Touré YT, Taylor CE, Norris DE, Dolo G, Lanzaro GC: **DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of Anopheles gambiae.** *Mol Ecol* 2001, **10**:1725-1732.
25. Tripet F, Touré YT, Dolo G, Lanzaro GC: **Frequency of multiple inseminations in field-collected Anopheles gambiae females revealed by DNA analysis of transferred sperm.** *Am J Trop Med Hyg* 2003, **68**:1-5.
26. Gentile G, Della Torre A, Maegga B, Powell JR, Caccone A: **Genetic differentiation in the African malaria vector, Anopheles gambiae s.s., and the problem of taxonomic status.** *Genetics* 2002, **161**:1561-1578.
27. Coluzzi M, Petrarca V, Di Deco M: **Chromosomal inversion intergradation and incipient speciation in Anopheles gambiae.** *Bollettino di Zoologia* 1985, **52**:45-63.
28. della Torre A, Merzagora L, Powell JR, Coluzzi M: **Selective introgression of paracentric inversions between two sibling species of the Anopheles gambiae complex.** *Genetics* 1997, **146**:239-244.
29. Turner TL, Hahn MW, Nuzhdin SV: **Genomic islands of speciation in Anopheles gambiae.** *PLoS Biol* 2005, **3**:e285.
30. Costantini C, Gibson G, Sagnon N, della Torre A, Brady J, Coluzzi M: **Mosquito responses to carbon dioxide in a west African Sudan savanna village.** *Med Vet Entomol* 1996, **10**:220-227.
31. Sachs J, Malaney P: **The economic and social burden of malaria.** *Nature* 2002, **415**:680-685.

32. WHO: **WHO Expert Committee on Malaria: Twentieth report.** *World Health Organization technical report series* 2000, **892**:i-v, 1-74.
33. Gould EA, Solomon T: **Pathogenic flaviviruses.** *Lancet* 2008, **371**:500-509.
34. Pinheiro FP, Corber SJ: **Global situation of dengue and dengue haemorrhagic fever, and its emergence in the Americas.** *World health statistics quarterly* 1997, **50**:161-169.
35. Trape JF: **The public health impact of chloroquine resistance in Africa.** *Am J Trop Med Hyg* 2001, **64**:12-17.
36. Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ: **Different evolutionary processes shaped the mouse and human olfactory receptor gene families.** *Hum Mol Genet* 2002, **11**:535-546.
37. Gilad Y, Man O, Pääbo S, Lancet D: **Human specific loss of olfactory receptor genes.** *Proc Natl Acad Sci U S A* 2003, **100**:3324-3327.
38. Wolf H, Wehner R: **Pinpointing food sources: olfactory and anemotactic orientation in desert ants, Cataglyphis fortis.** *J Exp Biol* 2000, **203**:857-868.
39. Karlson P, Luscher M: **'Pheromones': a new term for a class of biologically active substances.** *Nature* 1959, **183**:55-56.
40. Mehren JE: **Mate recognition: should fly stay or should fly go?** *Curr Biol* 2007, **17**:R240-R242.
41. Suh GSB, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ: **A single population of olfactory sensory neurons mediates an innate avoidance behaviour in Drosophila.** *Nature* 2004, **431**:854-859.
42. Ray S, Ferneyhough B: **Behavioral development and olfactory learning in the honeybee (Apis mellifera).** *Dev Psychobiol* 1999, **34**:21-27.
43. Reinhard J, Srinivasan MV, Zhang S: **Olfaction: scent-triggered navigation in honeybees.** *Nature* 2004, **427**:411.
44. Bush GL: **Mating Behavior, Host Specificity, and the Ecological Significance of Sibling Species in Frugivorous Flies of the Genus Rhagoletis (Diptera-Tephritidae).** *Am Nat* 1969, **103**:669-672.
45. Feder JL, Opp SB, Wlazlo B, Reynolds K, Go W, Spisak S: **Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly.** *Proc Natl Acad Sci U S A* 1994, **91**:7990-7994.
46. Feder JL, Chilcote CA, Bush GL: **The Geographic Pattern of Genetic Differentiation Between Host Associated Populations of Rhagoletis Pomonella (Diptera: Tephritidae) In the Eastern United States and Canada.** *Evolution* 1990, **44**:570-594.
47. Linn C, Feder JL, Nojima S, Dambroski HR, Berlocher SH, Roelofs W: **Fruit odor discrimination and sympatric host race formation in Rhagoletis.** *Proc Natl Acad Sci U S A* 2003, **100**:11490-11493.
48. Firestein S: **How the olfactory system makes sense of scents.** *Nature* 2001, **413**:211-218.
49. Ruťzler M, Zwiebel LJ: **Molecular biology of insect olfaction: recent progress and conceptual models.** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2005, **191**:777-790.
50. Vosshall LB, Stocker RF: **Molecular architecture of smell and taste in Drosophila.** *Annu Rev Neurosci* 2007, **30**:505-533.
51. Kwon HW, Lu T, Ruťzler M, Zwiebel LJ: **Olfactory responses in a gustatory organ of the malaria vector mosquito Anopheles gambiae.** *Proc Natl Acad Sci U S A* 2006, **103**:13526-13531.
52. Stocker RF: **The organization of the chemosensory system in Drosophila melanogaster: a review.** *Cell Tissue Res* 1994, **275**:3-26.

53. Shanbhag SR, Muller B, Steinbrecht RA: **Atlas of olfactory organs of Drosophila melanogaster - 1. Types, external organization, innervation and distribution of olfactory sensilla.** *Int J Insect Morphol Embryol* 1999, **28**:377-397.
54. Couto A, Alenius M, Dickson BJ: **Molecular, anatomical, and functional organization of the Drosophila olfactory system.** *Curr Biol* 2005, **15**:1535-1547.
55. Pitts RJ, Zwiebel LJ: **Antennal sensilla of two female anopheline sibling species with differing host ranges.** *Malar J* 2006, **5**:26.
56. Steinbrecht RA: **Experimental morphology of insect olfaction: tracer studies, X-ray microanalysis, autoradiography, and immunocytochemistry with silkmoth antennae.** *Microsc Res Tech* 1992, **22**:336-350.
57. Shanbhag SR, Muller B, Steinbrecht RA: **Atlas of olfactory organs of Drosophila melanogaster 2. Internal organization and cellular architecture of olfactory sensilla.** *Arthropod Struct Dev* 2000, **29**:211-229.
58. Kim MS, Repp A, Smith DP: **LUSH odorant-binding protein mediates chemosensory responses to alcohols in Drosophila melanogaster.** *Genetics* 1998, **150**:711-721.
59. Smith DP: **Odor and pheromone detection in Drosophila melanogaster.** *Pflugers Arch* 2007, **454**:749-758.
60. Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR: **A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila.** *Neuron* 1999, **22**:327-338.
61. Gao Q, Chess A: **Identification of candidate Drosophila olfactory receptors from genomic DNA sequence.** *Genomics* 1999, **60**:31-39.
62. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R: **A spatial map of olfactory receptor expression in the Drosophila antenna.** *Cell* 1999, **96**:725-736.
63. Vosshall LB, Wong AM, Axel R: **An olfactory sensory map in the fly brain.** *Cell* 2000, **102**:147-159.
64. Rogers ME, Sun M, Lerner MR, Vogt RG: **Snmp-1, a novel membrane protein of olfactory neurons of the silk moth Antheraea polyphemus with homology to the CD36 family of membrane proteins.** *J Biol Chem* 1997, **272**:14792-14799.
65. Rogers ME, Krieger J, Vogt RG: **Antennal SNMPs (sensory neuron membrane proteins) of Lepidoptera define a unique family of invertebrate CD36-like proteins.** *J Neurobiol* 2001, **49**:47-61.
66. Benton R, Vannice KS, Vosshall LB: **An essential role for a CD36-related receptor in pheromone detection in Drosophila.** *Nature* 2007, **450**:289-293.
67. Fishilevich E, Vosshall LB: **Genetic and functional subdivision of the Drosophila antennal lobe.** *Curr Biol* 2005, **15**:1548-1553.
68. Kaissling KE, Klein U, de Kramer JJ, Keil TA, Kanaujia S, Hemberger J: **Insect olfactory cells: electrophysiological and biochemical studies.** In *Molecular Basis of Nerve Activity. Proceedings of the International Symposium in Memory of D. Nachmansohn*. Edited by Changeux JP, Hucho F. Berlin: Walter de Gruyter; 1985:173-183.
69. Vogt RG, Riddiford LM, Prestwich GD: **Kinetic properties of a sex pheromone-degrading enzyme: the sensillar esterase of Antheraea polyphemus.** *Proc Natl Acad Sci U S A* 1985, **82**:8827-8831.
70. Ruťzler M, Lu T, Zwiebel LJ: **Galpha encoding gene family of the malaria vector mosquito Anopheles gambiae: expression analysis and immunolocalization of AGalphaq and AGalphao in female antennae.** *J Comp Neurol* 2006, **499**:533-545.
71. Merrill CE, Riesgo-Escovar J, Pitts RJ, Kafatos FC, Carlson JR, Zwiebel LJ: **Visual arrestins in olfactory pathways of Drosophila and the malaria vector mosquito**

Anopheles gambiae. *Proc Natl Acad Sci U S A* 2002, **99**:1633-1638.
72. Gomez-Diaz C, Martin F, Alcorta E: **The Inositol 1,4,5-triphosphate kinase1 gene affects olfactory reception in Drosophila melanogaster.** *Behav Genet* 2006, **36**:309-321.
73. Breer H, Boekhoff I, Tareilus E: **Rapid kinetics of second messenger formation in olfactory transduction.** *Nature* 1990, **345**:65-68.
74. Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K: **Insect olfactory receptors are heteromeric ligand-gated ion channels.** *Nature* 2008, **452**:1002-1006.
75. Wicher D, Schäfer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS: **Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels.** *Nature* 2008, **452**:1007-1011.
76. Robertson HM, Warr CG, Carlson JR: **Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster.** *Proc Natl Acad Sci U S A* 2003, **100 Suppl 2**:14537-14542.
77. Warr CG, Vosshall LB, Amrein HO, Carlson JR, Gao Q, Smith DP: **A unified nomenclature system for the Drosophila odorant receptors.** *Cell* 2000, **102**:145-146.
78. Kreher SA, Kwon JY, Carlson JR: **The molecular basis of odor coding in the Drosophila larva.** *Neuron* 2005, **46**:445-456.
79. Fishilevich E, Domingos AI, Asahina K, Naef F, Vosshall LB, Louis M: **Chemotaxis behavior mediated by single larval olfactory neurons in Drosophila.** *Curr Biol* 2005, **15**:2086-2096.
80. Hallem EA, Carlson JR: **Coding of odors by a receptor repertoire.** *Cell* 2006, **125**:143-160.
81. Dobritsa AA, Van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR: **Integrating the molecular and cellular basis of odor coding in the Drosophila antenna.** *Neuron* 2003, **37**:827-841.
82. Neuhaus EM, Gisselmann G, Zhang W, Dooley R, Störtkuhl K, Hatt H: **Odorant receptor heterodimerization in the olfactory system of Drosophila melanogaster.** *Nat Neurosci* 2005, **8**:15-17.
83. Wetzel CH, Behrendt HJ, Gisselmann G, Stortkuhl KF, Hovemann B, Hatt H: **Functional expression and characterization of a Drosophila odorant receptor in a heterologous cell system.** *Proc Natl Acad Sci U S A* 2001, **98**:9377-9380.
84. Kiely A, Authier A, Kralicek AV, Warr CG, Newcomb RD: **Functional analysis of a Drosophila melanogaster olfactory receptor expressed in Sf9 cells.** *J Neurosci Methods* 2007, **159**:189-194.
85. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ: **G protein-coupled receptors in Anopheles gambiae.** *Science* 2002, **298**:176-178.
86. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family.** *Genome Res* 2006, **16**:1395-1403.
87. Bohbot J, Pitts RJ, Kwon HW, Ruťzler M, Robertson HM, Zwiebel LJ: **Molecular characterization of the Aedes aegypti odorant receptor gene family.** *Insect Mol Biol* 2007, **16**:525-537.
88. Fox AN, Pitts RJ, Robertson HM, Carlson JR, Zwiebel LJ: **Candidate odorant receptors from the malaria vector mosquito Anopheles gambiae and evidence of down-regulation in response to blood feeding.** *Proc Natl Acad Sci U S A* 2001, **98**:14693-14697.

89. Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, Tanoue S, Yasukochi Y, Touhara K, Nishioka T: **Identification and functional characterization of a sex pheromone receptor in the silkmoth Bombyx mori.** *Proc Natl Acad Sci U S A* 2004, **101**:16653-16658.
90. Benton R: **On the ORigin of smell: odorant receptors in insects.** *Cell Mol Life Sci* 2006, **63**:1579-1585.
91. Wistrand M, Käll L, Sonnhammer EL: **A general model of G protein-coupled receptor sequences and its application to detect remote homologs.** *Protein Sci* 2006, **15**:509-521.
92. Benton R, Sachse S, Michnick SW, Vosshall LB: **Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo.** *PLoS Biol* 2006, **4**:e20.
93. Lundin C, Käll L, Kreher SA, Kapp K, Sonnhammer EL, Carlson JR, Heijne GV, Nilsson I: **Membrane topology of the Drosophila OR83b odorant receptor.** *FEBS Lett* 2007, **581**: 5601–5604.
94. Jones KA, Borowsky B, Tamm JA, Craig DA, Durkin MM, Dai M, Yao WJ, Johnson M, Gunwaldsen C, Huang LY, Tang C, Shen Q, Salon JA, Morse K, Laz T, Smith KE, Nagarathnam D, Noble SA, Branchek TA, Gerald C: **GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2.** *Nature* 1998, **396**:674-679.
95. Bush CF, Jones SV, Lyle AN, Minneman KP, Ressler KJ, Hall RA: **Specificity of olfactory receptor interactions with other G protein-coupled receptors.** *J Biol Chem* 2007, **282**:19042-19051.
96. Vosshall LB: **Diversity and expression of odorant receptors in Drosophila.** In *Insect Pheromone Biochemistry and Molecular Biology* Edited by Blomquist GJ, Vogt RG. Amsterdam: Elsevier; 2003:567-591.
97. Wanner KW, Anderson AR, Trowell SC, Theilmann DA, Robertson HM, Newcomb RD: **Female-biased expression of odourant receptor genes in the adult antennae of the silkworm, Bombyx mori.** *Insect Mol Biol* 2007, **16**:107-119.
98. Engsontia P, Sanderson AP, Cobb M, Walden KK, Robertson HM, Brown S: **The red flour beetle's large nose: An expanded odorant receptor gene family in Tribolium castaneum.** *Insect Biochem Mol Biol* 2008, **38**:387-397.
99. Krieger J, Raming K, Dewer YM, Bette S, Conzelmann S, Breer H: **A divergent gene family encoding candidate olfactory receptors of the moth Heliothis virescens.** *Eur J Neurosci* 2002, **16**:619-628.
100. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB: **Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction.** *Neuron* 2004, **43**:703-714.
101. Jefferis GS, Hummel T: **Wiring specificity in the olfactory system.** *Semin Cell Dev Biol* 2006, **17**:50-65.
102. Goldman AL, Van der Goes van Naters W, Lessing D, Warr CG, Carlson JR: **Coexpression of two functional odor receptors in one neuron.** *Neuron* 2005, **45**:661-666.
103. Fox AN, Pitts RJ, Zwiebel LJ: **A cluster of candidate odorant receptors from the malaria vector mosquito, Anopheles gambiae.** *Chem Senses* 2002, **27**:453-459.
104. Krieger J, Grosse-Wilde E, Gohl T, Breer H: **Candidate pheromone receptors of the silkmoth Bombyx mori.** *Eur J Neurosci* 2005, **21**:2167-2176.
105. Ray A, van Naters WG, Shiraiwa T, Carlson JR: **Mechanisms of odor receptor gene choice in Drosophila.** *Neuron* 2007, **53**:353-369.
106. Melo ACA, Rutzler M, Pitts RJ, Zwiebel LJ: **Identification of a chemosensory receptor**

from the yellow fever mosquito, Aedes aegypti, that is highly conserved and expressed in olfactory and gustatory organs. *Chem Senses* 2004, **29**:403-410.
107. Pitts RJ, Fox AN, Zwiebel LJ: **A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector Anopheles gambiae.** *Proc Natl Acad Sci U S A* 2004, **101**:5058-5063.
108. Xia Y, Zwiebel LJ: **Identification and characterization of an odorant receptor from the West Nile virus mosquito, Culex quinquefasciatus.** *Insect Biochem Mol Biol* 2006, **36**:169-176.
109. Schneider D: **Elektrophysiologische untersuchungen von chemo- und mechanorezeptoren der antenne des seidenspinners Bombyx mori L.** *L Z vergl Physiol* 1957, **40**:8-41.
110. Morita H, Yamashita S: **Receptor Potentials Recorded From Sensilla Basiconica on the Antenna of the Silkworm Larve, Bombyx mori.** *J Exp Biol* 1961, **38**:851-856.
111. Venard R, Pichon Y: **Étude électroantennographique de la réponse péripherique de l'antenne de Drosophila melanogster faite à l'aide de stimulations odorantes.** *C R Acad Sci Paris* 1981, **293**:839-842.
112. Clyne PJ, Warr CG, Carlson JR: **Candidate taste receptors in Drosophila.** *Science* 2000, **287**:1830-1834.
113. Dunipace L, Meister S, McNealy C, Amrein HO: **Spatially restricted expression of candidate taste receptors in the Drosophila gustatory system.** *Curr Biol* 2001, **11**:822-835.
114. Moon SJ, Köttgen M, Jiao Y, Xu H, Montell C: **A taste receptor required for the caffeine response in vivo.** *Curr Biol* 2006, **16**:1812-1817.
115. Chyb S, Dahanukar A, Wickens A, Carlson JR: **Drosophila Gr5a encodes a taste receptor tuned to trehalose.** *Proc Natl Acad Sci U S A* 2003, **100 Suppl 2**:14526-14530.
116. Jiao Y, Moon SJ, Montell C: **A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging.** *Proc Natl Acad Sci U S A* 2007, **104**:14110-14115.
117. Slone J, Daniels J, Amrein HO: **Sugar receptors in Drosophila.** *Curr Biol* 2007, **17**:1809-1816.
118. Dahanukar A, Lei YT, Kwon JY, Carlson JR: **Two Gr genes underlie sugar reception in Drosophila.** *Neuron* 2007, **56**:503-516.
119. de Bruyne M, Foster K, Carlson JR: **Odor coding in the Drosophila antenna.** *Neuron* 2001, **30**:537-552.
120. Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB: **Two chemosensory receptors together mediate carbon dioxide detection in Drosophila.** *Nature* 2007, **445**:86-90.
121. Kwon JY, Dahanukar A, Weiss LA, Carlson JR: **The molecular basis of CO2 reception in Drosophila.** *Proc Natl Acad Sci U S A* 2007, **104**:3574-3578.
122. Lu T, Qiu YT, Wang G, Ruťzler M, Kwon HW, Pitts RJ, van Loon JJ, Takken W, Carlson JR, Zwiebel LJ: **Odor Coding in the Maxillary Palp of the Malaria Vector Mosquito Anopheles gambiae.** *Curr Biol* 2007, **17**:1533-1544.
123. Thorne N, Amrein HO: **Atypical expression of Drosophila gustatory receptor genes in sensory and central neurons.** *J Comp Neurol* 2008, **506**:548-568.

Chapter 2. Odorant receptor C-terminal motifs in divergent insect species

2.1: Abstract

Insect odorant receptors are a large family of seven transmembrane proteins believed to be G-protein coupled receptors. The peptide sequences of two odorant receptors within a given species may share as little as 17% identity, and there is limited similarity between receptors of divergent species. One exception is *DmOr83b*, which is found in *Drosophila melanogaster* and is highly conserved in at least ten other insect species. *DmOr83b* is broadly expressed in most of the olfactory sensory neurons of *D. melanogaster* at most developmental stages, while other odorant receptors tend to have more restricted and specific expression patterns. *DmOr83b* is critical for *D. melanogaster* olfaction, and it is involved in properly localizing other odorant receptors possibly by forming heterodimers with these receptors. The C-terminal region has been implicated as sites for such heterodimer formation. Multiple em for motif elicitation (MEME), a hidden markov model based program, was used to uncover three conserved motifs in the C-termini of a vast majority of the odorant receptor peptides from *Anopheles gambiae*, *D. melanogaster*, and *Apis mellifera*. These motifs are also found in *DmOr83b* and its orthologs and the order of these motifs is conserved as well. The conservation of these motifs among divergent odorant receptors in divergent species suggests functional importance. We propose that these motifs are involved in

receptor- receptor protein interactions, contributing to the heterodimer formation between *DmOr83b* (or its orthologs) and other odorant receptors.

2.2: Introduction

Insect olfaction and olfactory signaling is a rapidly growing area of research [1]. Several protein families are being studied that include odorant binding proteins, sensory neuron membrane proteins, odorant degrading enzymes, and odorant receptors. A large body of recent literature has been written on insect odorant receptors [2-6]. Most of the papers addressing insect odorant receptors report either the discovery of receptor genes in an insect species [4, 7], and/or the expression of selected odorant receptor genes at various points of the organism's life cycle [8]. Odorant receptor gene expression is usually localized to the insect sensory organs such as antenna and maxillary palp [9, 10], and more recently have been found to be expressed in the proboscis [11]. There are also a growing number of papers addressing the specific functions of several odorant receptor proteins [5, 12-14].

Insect odorant receptors have been reported to be putative G-protein coupled receptors [2, 4, 15], but recently this status has been questioned [16, 17]. The most extensively researched insect odorant receptor is *DmOr83b* in *Drosophila melanogaster*. A highly conserved ortholog of *DmOr83b* has been found in all insect species with sufficient genomic sequence information. This list includes *D.*

melanogaster [9], *Anopheles gambiae* [18], *An. stephensi* (R. Miller and Z. Tu, unpublished data), *An. quadriannulatus* (R. Miller and Z. Tu, unpublished data), *Aedes aegypti* [8], *Culex quinquefasciatus* [19], *Bombyx mori* [5], *Heliothis virescens* [20], *Apis mellifera* [6], and *Tribolium castaneum* (GenBank Accession XP_973196). Note that the GenBank name for OR is GPROR. This is in contrast to the vast majority of the other insect odorant receptors, which are not conserved between species of different genera. *DmOr83b* is broadly expressed in most of the olfactory sensory neurons of *D. melanogaster* at most stages of development [9, 14]. This again is in contrast to other odorant receptors, which have been reported to have a restrictive expression pattern [9, 10]. *D. melanogaster* lacking a copy of *DmOr83b* are not able to respond to olfactory cues, and other odorant receptors are not properly localized to the membrane of olfactory sensory neurons [14]. *DmOr83b* is capable of forming a heterodimer with at least one *D. melanogaster* odorant receptor: *DmOr43a* [21]. The requirement of a heterodimer of two G-protein coupled receptors has only been previously observed in the GABA complex where heterodimer formation is required for the function of potassium/calcium channels [17, 22]. Benton and co-authors provides further evidence of heterodimer formation involving *DmOr83b* with DmOr22a/b, and additionally point to the C-terminal domain of odorant receptor peptides as being the site of heterodimer formation [23]. The specific location(s) of the protein-protein interaction(s) were not explored. However, previous reports have indicated limited amino acid conservation

occurring in the C-terminal end of *D. melanogaster* odorant receptor (DmOr) peptides, including a nearly invariable tryptophan residue [2, 24, 25].

Using a hidden markov model based program called multiple em for motif elicitation (MEME) [26], we have discovered three C-terminal motifs in 76 of the 79 previously annotated *An. gambiae* odorant receptor peptides [4]. Subsequent analysis indicates that these motifs are conserved within the odorant receptor peptides of *D. melanogaster* and *Ap. mellifera* [6]. This is significant given that it has been reported that insect odorant receptor peptides are highly divergent within and between species [2-4]. For example, the amino acid identity between insect odorant receptors of the same species is only 17% in some cases [25]. We hypothesize that these motifs are protein-protein interaction sites involved in odorant receptor-odorant receptor interactions or potentially heterodimer formation between *DmOr83b* and other odorant receptors.

2.3: Materials and Methods

Alignment of *An. gambiae* odorant receptor peptides with ClustalW. All 79 *An. gambiae* odorant receptor peptides [4] were aligned using ClustalW v1.83.1 [27]. Default parameters were used (multiple alignment gap opening penalty = 10, gap extension penalty = 0.2). Alignments were illustrated using the Jalview java alignment editor [28].

Motif discovery in odorant receptors peptides using MEME. *An. gambiae* and *Ap. mellifera* odorant receptor peptide sequences were obtained from the supplementary material of two separate studies [4, 6]. Fifty-nine *D. melanogaster* odorant receptor peptides were obtained from the Ensembl database (http://www.ensembl.org) and were used in the analysis. The program multiple em for motif elicitation (MEME) [26](http://meme.sdsc.edu/meme/) version 3.5.1 was compiled on a Macintosh computer running Mac OS 10.4.8. Each MEME analysis was run with peptide data set from each species as input. For all three data sets MEME was run using the following command line: meme dataset_name -protein -mod zoops -minw 15 -maxw 45 -wg 8 -ws 0.2 -evt .00001 -nmotifs 8. The program command call is meme, while dataset_name identifies the input dataset, -protein indicates the dataset contained peptide sequences, and -mod defines the search model. The remaining parameters were -minw, which sets the minimum possible motif width at 15 residues, -maxw, which sets the maximum possible motif width at 45 residues, -wg, which is the gap opening penalty, -ws, which is the gap extension penalty, -evt, which is the maximum e-value for a motif to be reported, and -nmotifs, which indicated the number of motifs that are searched for in the input dataset. Gap opening and extension penalties were reduced from the default values of wg=11 and ws=1 to wg=8 and ws=0.2 to reduce artificial breakup of the motifs due to small insertions or deletions. In addition to

searching for the top 3 motifs as set by default, -nmotifs 8 was used to determine whether more than three motifs existed in each dataset.

MAST searching of *An. gambiae* gustatory receptor peptides for odorant receptor motifs. The motif alignment and search tool (MAST) [29], another program in the MEME package, was used to search for AgOr motifs in all 76 *An. gambiae* GRs [4]. MAST version 3.5.1 was installed as part of the MEME package (see above). Command line used for MAST was: mast motif_matrices_found_by_meme –d database_of_AgGrs. The motif_matrices_found_by_meme are the profile matrices of the motifs found in a previous MEME analysis and they effectively define the motifs. These matrices were used to search the database_of_AgGrs, where AgGrs stands for *An. gambiae* gustatory receptors. No other parameters were used.

Weblogo diagrams. All weblogo diagrams were constructed using the weblogo program [30] (http://weblogo.berkeley.edu/). MEME output includes BLOCKS of the motifs. If an odorant receptor peptide sequence was found to have a motif, the part of the peptide sequence that contains that motif was used in an alignment, which produced an aligned BLOCK. The aligned BLOCK was used to construct weblogos.

2.4: Results

2.4.1: ClustalW alignments of *An. gambiae* odorant receptors

An alignment of all 79 *An. gambiae* odorant receptor (AgOr) peptides using the multiple sequence alignment program ClustalW [27] revealed very little strict sequence conservation (Figure 2.1). There were a small number of conserved or highly prevalent residues located in the C-terminal region (Figure 2.1, blue-colored residues). One of these highly conserved residues is a tryptophan residue found in all but four AgORs. The lack of strict sequence conservation in AgOrs, and the prevalence of the conserved tryptophan residue is consistent with what has been previously reported for DmOr peptides [25].

2.4.2: MEME identifies c-terminal motifs in *An. gambiae* odorant receptors

To locate conserved patterns a hidden markov model based program named multiple em for motif elicitation (MEME) was used [26]. MEME has been used to locate potential regulatory sites in sequences upstream of genes [31], potential protein-protein interaction domains [32], and homologous genes missed by homology search [33]. One key advantage of MEME over common alignment programs is its ability to find motifs that are not absolutely conserved in consensus sequence. Other advantages of MEME are its speed, no need for prior knowledge about a dataset, and

its ability to locate motifs that may not be in the same order through all members of a dataset.

All 79 AgOr peptides were used as input for MEME run using a gap opening penalty of eight and a gap extension parameter of 0.2. Three motifs were identified within the dataset all with highly significant e-values ($4.2e^{-401}$, $4.5e^{-367}$, and $1.1e^{-332}$) (Figure 2.2, Table 2.1). All three motifs were present within the last 70 or 90 amino acid residues of the C-terminal end of AgOr peptides, and 76 out of 79 (96%) AgOr peptides had all three motifs. The order of the motifs from N-terminal to C-terminal is motif 3, motif 2, and motif 1. MEME numbers the motifs according to their relative e-values with motif 1 having the best e-value. In subsequent discussions, the three motifs are referred to as motif A, motif B, and motif C, with motif A being furthest of the three from the C-terminus and motif C being the closest to the C-terminus. This naming system is used to allow meaningful comparison between results from different species where these motifs have different ranks of e-values relative to each other. The combined p-value of finding all of the identified motifs in a given odorant receptor peptide in the dataset ranged from $2.49e^{-12}$ to $2.62e^{-38}$. The combined p-value was the probability of finding a match of a sequence in the dataset to a group of motifs by random chance (Figure 2.2). Significantly, *AgOr7*, the mosquito ortholog of *DmOr83b*, has all three motifs at the C-terminal end (Figure 2.2, asterisk). When the number of motifs for MEME was increased to search for from three to eight motifs, only one

additional motif was found with a significant distribution ($2.9e^{-295}$, present in 63 of 79 AgOr peptides). This motif had limited sequence conservation with the notable exception of a histidine residue located approximately 70 residues to the N-terminal of *An. gambiae* motif A.

Weblogo diagrams of motif A (Figure 2.3A), motif B (Figure 2.3B) and motif C (Figure 2.3C) illustrate the level of amino acid conservation within each motif at each position [30]. It is apparent from the weblogo diagrams that only a small portion of each motif consists of highly prevalent residues although there are additional areas where the chemical properties of the residues such as hydrophobicity, charge, and side chain structure are conserved. For example, in motif B of *An. gambiae* (Figure 2.3B) residue 1 and 2 are predominately positively charged residues while residues 4, 6, 7, and 8 are hydrophobic. The most highly conserved residues in the AgOr motifs are the tryptophan residue in motif A (Figure 2.3A), and a tyrosine/serine dyad in motif C (Figure 2.3C) mentioned above.

2.4.3: Odorant receptor c-terminal motifs are not found in gustatory receptors

Insect gustatory receptors are another family of putative G-protein coupled receptors. Insect gustatory receptors and odorant receptors are the closest relatives to each other in evolutionary terms [24, 34]. *DmGr21a* in *D. melanogaster* is able to confer response to carbon dioxide [35] in conjunction with *DmGr63a* [36, 37]. The motif

alignment and search tool (MAST) [29] was used to search for the previously identified AgOr motifs in all 76 *An. gambiae* GRs [4]. The best hit showed an e-value of 0.033 for a motif in the C-terminal region of one gustatory receptor. The poor e-value of the hit as well as further manual inspection suggests that it is not a true match. Thus this analysis indicates that the AgOr motifs are specific to odorant receptors and not a feature of G-protein coupled receptors.

2.4.4: Odorant receptor c-terminal motifs are found in *D. melanogaster* and *Ap. mellifera* odorant receptors

A DmOr peptide database of 59 DmOrs was used as input into MEME to determine if any similar motifs existed in these odorant receptors. This analysis revealed three motifs found in the C-terminal end of a vast majority of these peptides (Table 2.1). All three motifs were found in 54 of 59 (92%) DmOrs. As was the case in AgOr peptides these three motifs are in the same order in all DmOr peptides. A side-by-side comparison of the weblogo diagrams from motifs A and B in both species reveals obvious similarities in sequence (Figure 2.3A and Figure 2.3B). Most significant is the highly conserved tryptophan residue in motif A of both species. Part of motif A in DmOr peptides has been previously identified as the sequence of Phe-Pro-X-Cys-Tyr-(X)20-Trp [25]. The analysis showed several additionally conserved residues such as a glycine (residue 9) and a tyrosine/phenylalanine (residue 24). Motif C is very similar in both species in terms of their sequences and boundaries (Figure 2.3C).

Eight motifs were found in *Ap. mellifera* odorant receptor peptides (AmOr). Three of the motifs are apparent orthologs to the dipteran motifs A, B, and C (Table 2.1 and Figure 2.3) both in terms of their sequence and relative location. Among the eight AgOr motifs, motifs A, B, and C ranked as number 1, 4, and 2 in terms of the significance of their respective e-values. The motif that had the third best e-value was near the middle of the receptor peptide, and is not shared with the dipteran receptors. Motifs ranked number 5 to 8 appear to have limited distribution in subgroups of AmOr peptides, and thus are not universal motifs in all AmOrs. These motifs are not further discussed in this paper. All three motifs are present in 147 of 170 (86%) AmOr peptides (Table 2.1 and Figure 2.3). Motifs A, B, and C in AmOrs share similar sequence with dipteran Motifs A, B, and C respectively (Figure 2.3). For example there is a highly prevalent glycine residue followed by two variable residues, and then a highly prevalent leucine residue in motif A of all three species in addition to the conserved tryptophan residue. AmOr motif C is again very similar to the dipteran motifs (Figure 2.3C). However, instead of a tyrosine/serine dyad there is a phenylalanine/serine dyad in AmOr. The MEME analysis has therefore found three C-terminal motifs that are located in *An. gambiae*, *D. melanogaster* and *Ap. mellifera* odorant receptor peptides. Most of the residues in these motifs are not highly conserved, but several are highly prevalent across these diverse insect species.

2.5: Discussion

Three motifs were located in the C-terminal ends of the odorant receptor peptides of three divergent insect species *An. gambiae*, *D. melanogaster*, and *Ap. mellifera* using a hidden markov model program. Table 2.1 lists the number of odorant receptors containing these motifs in each species, the e-value of the motifs, and the prevalent amino acid sequences of these motifs. The vast majority of insect odorant receptor peptides analyzed contain these C-terminal motifs. This is interesting considering that insect odorant receptor proteins are a very diverse family having very little conservation between species or within one species [2, 3, 7, 25]. These motifs were not found in *An. gambiae* GRs despite the close evolutionary relationship between the odorant receptor and GR families [24, 34].

Although all of the motifs described above had wide distribution in odorant receptors of the three species, motif B was not present in a small, but significant number of odorant receptors, especially in *Ap. mellifera* (Table 2.1). The absence of motif B may be explained by either technical or biological reasons, or both, as described below. Eleven of the 18 AmOrs lacking motif B had incomplete C-termini in current annotation, and two of the peptides were clearly pseudogenes [6]. Motif B was also not found in *AmOr2*, which is the honeybee ortholog of *DmOr83b*. However, a close inspection of the *AmOr2* sequence revealed no amino acid substitution in the motif B region in comparison with *DmOr83b* and one substitution in comparison with

AgOr7 (Figure 2.4). Therefore, sequence variation between motif B of the three species may explain why nearly identical sequences were recognized as motif B in *DmOr83b* and *AgOr7* but not in *AmOr2*. Motif B was also lacking in two AgOrs and four DmOrs. Motif B was not as well conserved as the other two motifs (Figure 2.3). It is possible that the specific sequence of motif B is not as important as the chemical or structural properties of the residues in this motif. In comparing motif B in all three insect species (Figure 2.3B), some amino acid residues are present that are highly variable, but most of the residues in this region are hydrophobic in character. This conservation of hydrophobicity in these five residues may be functionally significant, while at the same time are difficult to be recognized by computer programs. It is also possible that motif B serves a role in enhancing a biological process, but is not absolutely required. For example, based on the working hypothesis that these C-terminal motifs are involved in protein-protein interactions, odorant receptor proteins lacking motif B might have a lower binding efficiency.

Having identified these motifs it is appropriate to ask why these motifs are present in the highly diverse insect odorant receptor family? As mentioned above, one possibility is that these motifs are involved in protein-protein interactions. There have been many efforts to identify protein-protein interaction sites through *in silico* methods, which resulted in the identification of several key characteristics. Protein-protein interaction sites are exposed on the surface of proteins and are hydrophobic, circular,

and protruding [38-41]. Within these interaction areas are small “hot-spots” of a few residues contributing greatly to the overall binding energy of protein-protein interactions [42]. In one survey it was found that tryptophan, tyrosine, and arginine are highly prevalent in these “hot-spots” [43]. Another study reported that tryptophan, phenylalanine, and methionine residues are significantly conserved in binding sites, but not on other exposed surfaces of proteins [44]. Highly conserved and prevalent tryptophan, tyrosine, phenylalanine, and arginine residues were located in the C-terminal motifs of odorant receptors (Figure 2.3). We hypothesize that these motifs are protein-protein interaction sites, which would explain the conservation of only a few residues across the highly diverse insect odorant receptor protein family.

Unfortunately, at present there is no X-ray crystal structure of any insect odorant receptor or gustatory receptor that may illuminate the exact positioning of the newly discovered motifs and their potential role in protein-protein interaction. Hydrophobicity analysis can be useful at least in determining where residues are in relation to transmembrane helices. In this study the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict the transmembrane helices of five odorant receptors each from *An. gambiae*, *D. melanogaster*, and *Ap. mellifera* along with *DmOr83b* (data not shown). Motif A in all three species was found within helix 6, and perhaps part of the helix 6-7 loop. The difficulty of pinning down the exact positioning of helix 7 in particular makes this conclusion hard to draw unequivocally. It also makes

further analysis of the positions of motif B and C uninformative except the supposition that these motifs lie near or in helix 7.

The vast majority of these insect odorant receptors maintained these motifs across hundreds of millions of years of evolution. This is impressive considering that the identity between insect odorant receptor peptides of the same species in some cases is as low as 17% [25]. These motifs have several highly conserved amino acids that were identified as being important in protein-protein interactions in other models. It is possible that these motifs allow odorant receptor-odorant receptor interactions as has been reported in vitro [21]. A more tantalizing prospect is that all or some of these motifs are involved in the formation of a heterodimer complex between *DmOr83b* or its ortholog and other odorant receptors [21, 23], a hypothesis that may be tested experimentally.

Acknowledgements

We thank the anonymous reviewers for their comments. This work was supported by a NIH grant AI063252 to Z. Tu.

Abbreviations

multiple em for motif eliciation, MEME; motif alignment and search tool, MAST; *Drosophila melanogaster* odorant receptor, DmOr; *Anopheles gambiae* odorant receptor, AgOr; *Apis mellifera* odorant receptor, AmOr

2.6: Figures

Figure 2.1: Multiple sequence alignment of the C-terminal region of all 79 *Anopheles gambiae* odorant receptors. AgOr peptides were aligned using ClustalW, and the subsequent alignment visualized using Jalview. A residue present at a given site in 50% or more of the AgOr peptides is boxed in blue. The more intense the blue the more often the residue is found at that site. Only the C-terminal region of the alignment is shown. The positions of motifs A, B, and C are shown. These motifs were not identified using alignment shown here. Instead they were identified using MEME. See Table 2.1 and Figure 2.3 for details.

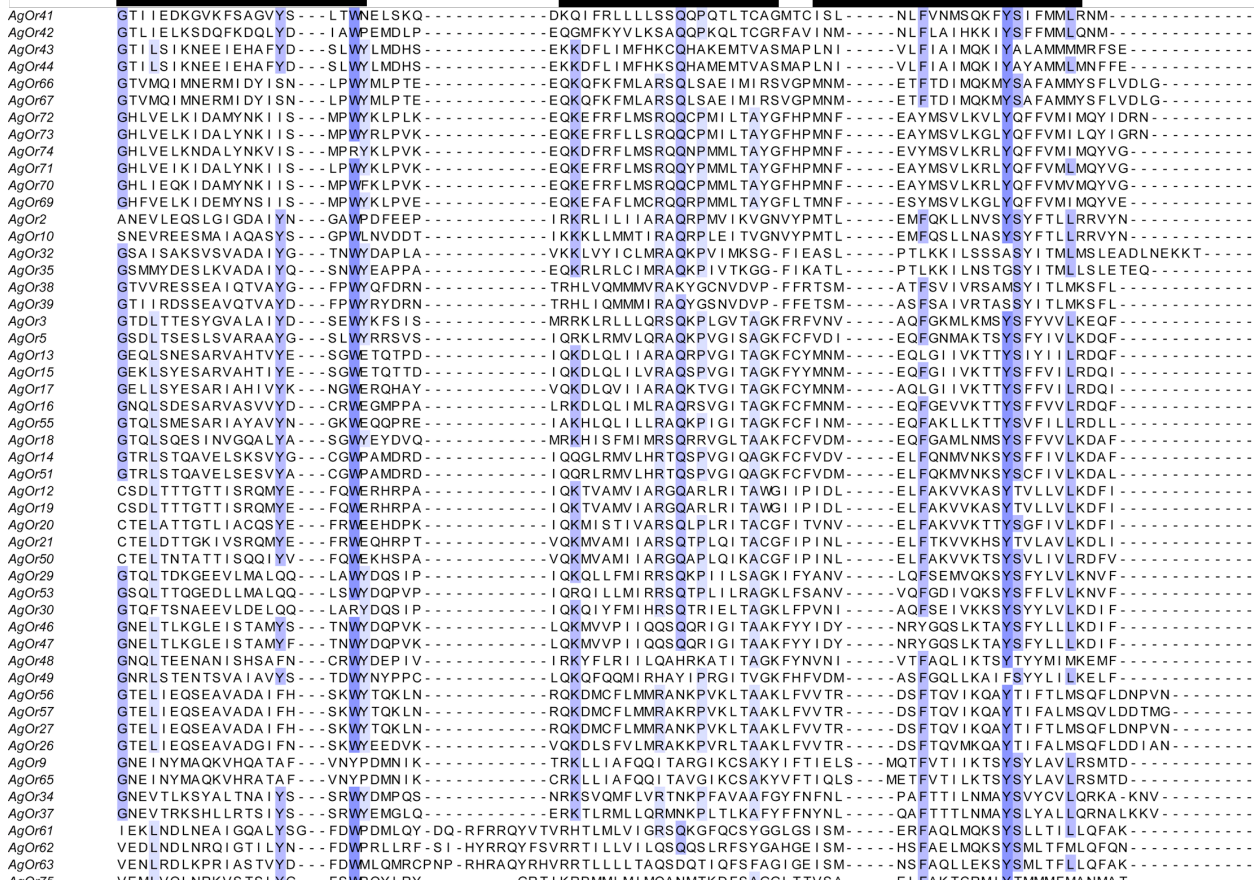

Motif A
Motif B
Motif C
AgOr41
AgOr42
AgOr43
AgOr44
AgOr66
AgOr67
AgOr72
AgOr73
AgOr74
AgOr71
AgOr70
AgOr69
AgOr2
AgOr10
AgOr32
AgOr35
AgOr38
AgOr39
AgOr3
AgOr5
AgOr13
AgOr15
AgOr17
AgOr16
AgOr55
AgOr18
AgOr14
AgOr51
AgOr12
AgOr19
AgOr20
AgOr21
AgOr50
AgOr29
AgOr53
AgOr30
AgOr46
AgOr47
AgOr48
AgOr49
AgOr56
AgOr57
AgOr27
AgOr26
AgOr9
AgOr65
AgOr34
AgOr37
AgOr61
AgOr62
AgOr63
AgOr75
AgOr76
AgOr78
AgOr77
AgOr79
AgOr60
AgOr68
AgOr4
AgOr31
AgOr1
AgOr64
AgOr54
AgOr23
AgOr24
AgOr59
AgOr11
AgOr52
AgOr36
AgOr25
AgOr45
AgOr40
AgOr28
AgOr8
AgOr6
AgOr22
AgOr58
AgOr7
AgOr33

Figure 2.2: C-terminal motifs found in *Anopheles gambiae* odorant receptors. The image was taken directly from the MEME output and shows the position of three C-terminal motifs located in the first 20 AgOr peptides. Only the first 20 AgOrs were shown to save space. The asterisk points to *AgOr7*, which is the *An. gambiae* ortholog of *DmOr83b*. The combined p-value is the probability of finding a match of a sequence in this dataset to a group of motifs by random chance.

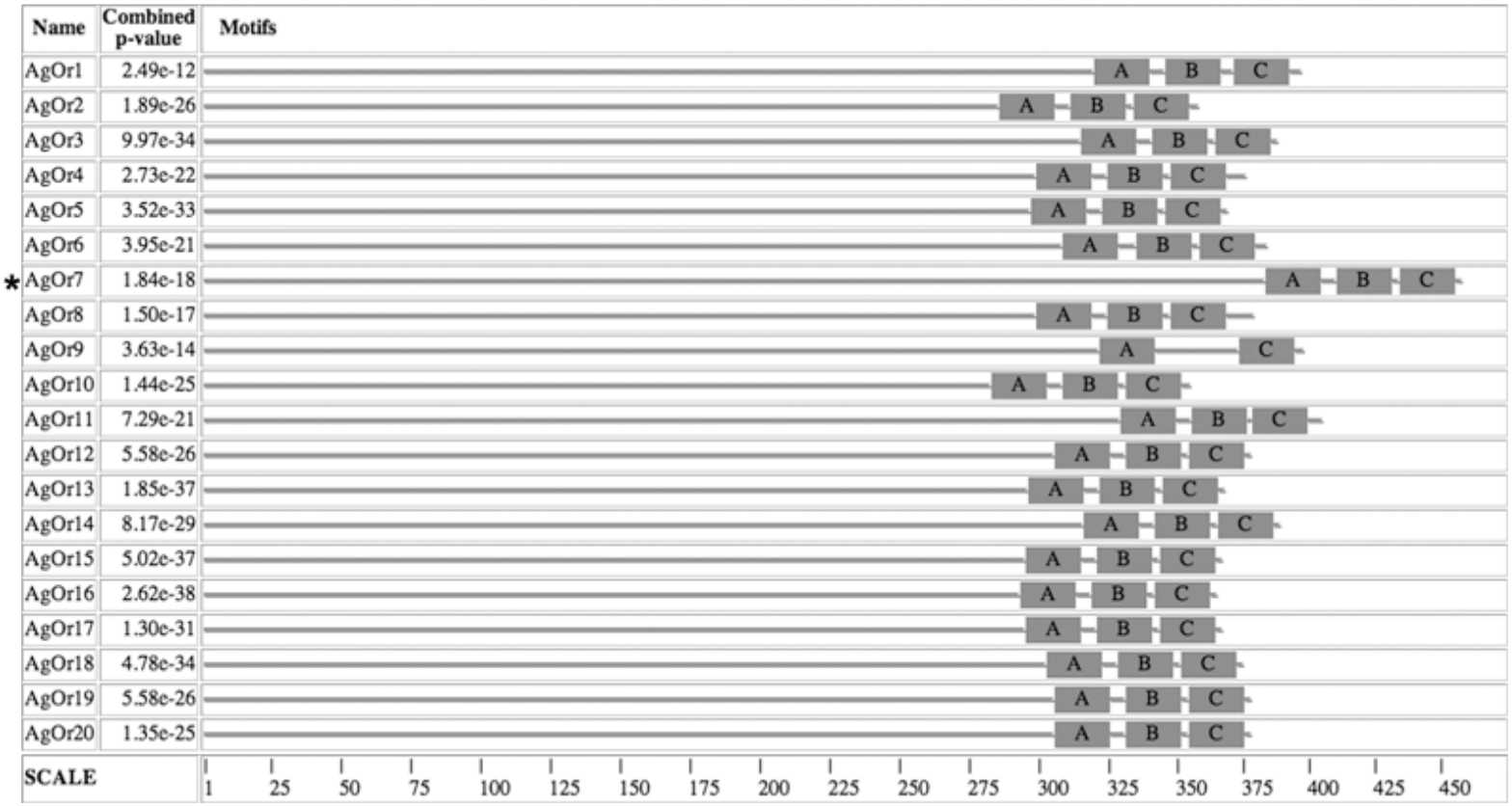
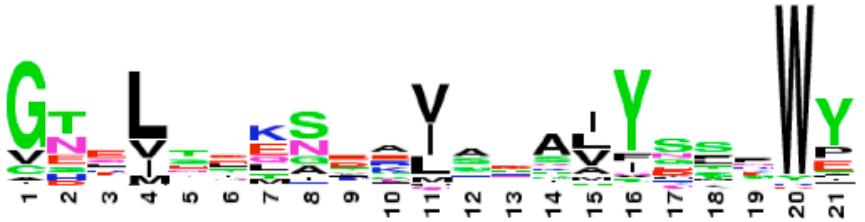

Name
Combined p-value
Motifs
AgOr1 2.49e-12
AgOr2 1.89e-26
AgOr3 9.97e-34
AgOr4 2.73e-22
AgOr5 3.52e-33
AgOr6 3.95e-21
*AgOr7 1.84e-18
AgOr8 1.50e-17
AgOr9 3.63e-14
AgOr10 1.44e-25
AgOr11 7.29e-21
AgOr12 5.58e-26
AgOr13 1.85e-37
AgOr14 8.17e-29
AgOr15 5.02e-37
AgOr16 2.62e-38
AgOr17 1.30e-31
AgOr18 4.78e-34
AgOr19 5.58e-26
AgOr20 1.35e-25
SCALE
1 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450

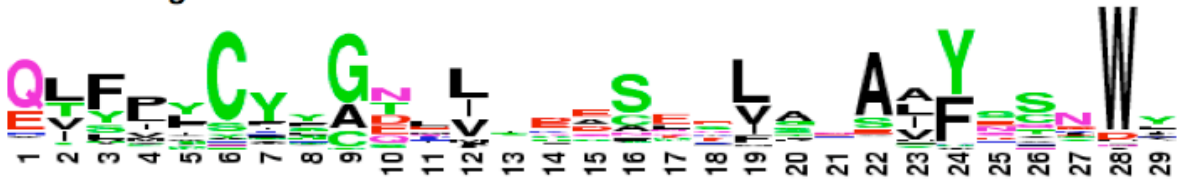
Figure 2.3: Weblogo presentation of motifs A, B, C in *Anopheles gambiae*, *Drosophila melanogaster*, and *Apis mellifera* odorant receptor peptides. Each line contains weblogo diagrams for motifs A, B, or C in one species. Weblogo diagrams indicate the prevalence of amino acids at specific positions. A). Weblogo presentation of motif A in all three species. B). Weblogo presentation of motif B in all three species. C) Weblogo presentation of motif C in all three species. Shown are weblogo diagrams indicating the prevalence of amino acids at specific positions in each motif.

Motif A

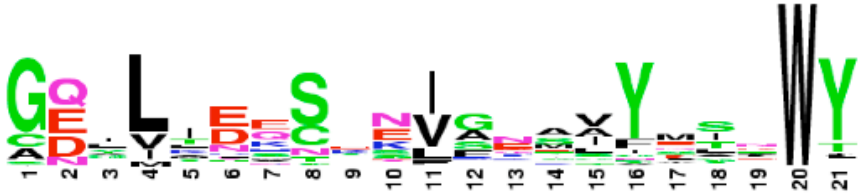
An. gambiae



D. melanogaster

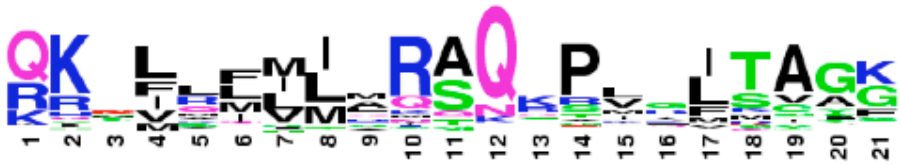


Ap. mellifera



Motif B

An. gambiae



D. melanogaster

Ap. mellifera

Motif C

An. gambiae

D. melanogaster

Ap. mellifera

Figure 2.4: Conservation at the C-terminal regions of *DmOr83b* and its orthologs. Shown here is a ClustalW alignment of the last ~90 amino acid residues of the *Or83b* family members in *Drosophila melanogaster* (*DmOr83b*), *Anopheles gambiae* (*AgOr7*), and *Apis mellifera* (*AmOr2*). The relative position of motifs A, B, and C are shown.


DmOr83b QVFHFCIFGNRLIEESSSVMEAAYSCHWYDGSEEAKTFVQIVCQQCQKAMSISGAKFFTVSLDLFASVLGAVVTYFMVLVQLK
AgOr7 QVFLFCIFGNRLIEESSSVMKAAYSCHWYDGSEEAKTFVQIVCQQCQKAMTISGAKFFTVSLDLFASVLGAVVTYFMVLVQLK
AmOr2 QVFMLCIFGNRLIEESSSVMEAAYSCHWYDGSEEAKTFVQIVCQQCQKAMSISGAKFFTVSLDLFASVLGAMVTYFMVLVQLK

2.7: Table

Table 2.1: Three conserved C-terminal motifs in mosquito, fruit fly, and honeybee OR peptides.

	Motif[1]	No. of ORs Containing Motif [2]	Motif E-value [3]	Prevalent Amino Acid Sequence [4]
An. gambiae	Motif A	78/79	$1.1e^{-332}$	G[T|N]ELTX[K|E][S|N]EX[V|I][A|S]DA[I|L|V]YSSPWY
	Motif B	77/79	$4.5e^{-367}$	[Q|R]KXL[L|R][F|L][M|I|L][I|L|M]MR[A|S]QKP[L|V]G[I|L]TAG[K|G]
	Motif C	79/79	$4.2e^{-401}$	[V|M][N|S][M|L]E[T|L]FAX[I|V][L|V]K[T|K]SYS[Y|F]F[T|M][V|L][L|M]
D. melanogaster	Motif A	59/59	$6.0e^{-368}$	[Q|E][L|T]F[P|L]YCYY[G|A][N|T]L[L|I|V]XXESEX[L|V]AXA[A|L][Y|F]SSNWY
	Motif B	54/59	$1.1e^{-177}$	Y[R|K][R|K]XLL[L|F]F[I|L|M]MR[A|S]Q[R|K|Q]PVX[L|I][K|R|T]AG
	Motif C	59/59	$1.7e^{-214}$	[S|N][L|M]XTFX[A|S][I|V]L[K|R]X[A|S][Y|F]F[T|A]LL[R|K]SM
Ap. mellifera	Motif A	164/170	$6.8e^{-1091}$	G[Q|E|D][I|L]LI[E|D][E|Q][S|C]X[N|E][I|V][G|A]NAVYMSNWY
	Motif B	152/170	$3.7e^{-880}$	[K|L]DLIL[I|V][I|M]IRS[S|N]XPCK[L|I]TAG[K|G][L|I|F]
	Motif C	160/170	$4.8e^{-1086}$	D[L|M|V]SLETF[T|G]S[I|V][L|I][K|S]T[S|A]FSY[L|F][N|T][L|V]LR

Notes
1. The naming of motifs A, B, and C is described in Results section. They are from N- to C- terminus.
2. The number of ORs containing a given motif is given as a fraction of the total number of ORs in a species.
3. An estimate of the likelihood of each motif being found in the dataset by random chance.
4. The sequences are from MEME output and they reflect the amino acid residues that are most frequent at these positions. For example, having a G in the output sequence does not mean that all ORs of that species will have a G residue at that position. A [|] indicates more than one amino acid residue is frequent at this position.

2.8: References

1. Ruťzler M, Zwiebel LJ: **Molecular biology of insect olfaction: recent progress and conceptual models.** *Journal of comparative physiology A, Neuroethology, sensory, neural, and behavioral physiology* 2005, **191**:777-790.
2. Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR: **A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila.** *Neuron* 1999, **22**:327-338.
3. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R: **A spatial map of olfactory receptor expression in the Drosophila antenna.** *Cell* 1999, **96**:725-736.
4. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ: **G protein-coupled receptors in Anopheles gambiae.** *Science* 2002, **298**:176-178.
5. Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, Tanoue S, Yasukochi Y, Touhara K, Nishioka T: **Identification and functional characterization of a sex pheromone receptor in the silkmoth Bombyx mori.** *Proc Natl Acad Sci U S A* 2004, **101**:16653-16658.
6. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family.** *Genome Res* 2006, **16**:1395-1403.
7. Robertson HM, Warr CG, Carlson JR: **Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster.** *Proc Natl Acad Sci U S A* 2003, **100 Suppl 2**:14537-14542.
8. Melo ACA, Ruťzler M, Pitts RJ, Zwiebel LJ: **Identification of a chemosensory receptor from the yellow fever mosquito, Aedes aegypti, that is highly conserved and expressed in olfactory and gustatory organs.** *Chem Senses* 2004, **29**:403-410.
9. Vosshall LB, Wong AM, Axel R: **An olfactory sensory map in the fly brain.** *Cell* 2000, **102**:147-159.
10. Fox AN, Pitts RJ, Robertson HM, Carlson JR, Zwiebel LJ: **Candidate odorant receptors from the malaria vector mosquito Anopheles gambiae and evidence of down-regulation in response to blood feeding.** *Proc Natl Acad Sci U S A* 2001, **98**:14693-14697.
11. Kwon H-W, Lu T, Ruťzler M, Zwiebel LJ: **Olfactory responses in a gustatory organ of the malaria vector mosquito Anopheles gambiae.** *Proc Natl Acad Sci U S A* 2006, **103**:13526-13531.
12. Wetzel CH, Behrendt HJ, Gisselmann G, Störtkuhl K, Hovemann B, Hatt H: **Functional expression and characterization of a Drosophila odorant receptor in a heterologous cell system.** *Proc Natl Acad Sci U S A* 2001, **98**:9377-9380.
13. Hallem EA, Fox AN, Zwiebel LJ, Carlson JR: **Olfaction: mosquito receptor for human-sweat odorant.** *Nature* 2004, **427**:212-213.
14. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB: **Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction.** *Neuron* 2004, **43**:703-714.
15. Gao Q, Chess A: **Identification of candidate Drosophila olfactory receptors from genomic DNA sequence.** *Genomics* 1999, **60**:31-39.
16. Benton R: **On the ORigin of smell: odorant receptors in insects.** *Cellular and molecular*

life sciences : CMLS 2006, **63**:1579-1585.
17. Wistrand M, Käll L, Sonnhammer ELL: **A general model of G protein-coupled receptor sequences and its application to detect remote homologs.** *Protein science : a publication of the Protein Society* 2006, **15**:509-521.
18. Pitts RJ, Fox AN, Zwiebel LJ: **A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector Anopheles gambiae.** *Proc Natl Acad Sci U S A* 2004, **101**:5058-5063.
19. Xia Y, Zwiebel LJ: **Identification and characterization of an odorant receptor from the West Nile virus mosquito, Culex quinquefasciatus.** *Insect Biochem Mol Biol* 2006, **36**:169-176.
20. Krieger J, Raming K, Dewer YME, Bette S, Conzelmann S, Breer H: **A divergent gene family encoding candidate olfactory receptors of the moth Heliothis virescens.** *The European journal of neuroscience* 2002, **16**:619-628.
21. Neuhaus EM, Gisselmann G, Zhang W, Dooley R, Störtkuhl K, Hatt H: **Odorant receptor heterodimerization in the olfactory system of Drosophila melanogaster.** *Nat Neurosci* 2005, **8**:15-17.
22. Jones KA, Borowsky B, Tamm JA, Craig DA, Durkin MM, Dai M, Yao WJ, Johnson M, Gunwaldsen C, Huang LY, Tang C, Shen Q, Salon JA, Morse K, Laz T, Smith KE, Nagarathnam D, Noble SA, Branchek TA, Gerald C: **GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2.** *Nature* 1998, **396**:674-679.
23. Benton R, Sachse S, Michnick SW, Vosshall LB: **Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo.** *PLoS Biol* 2006, **4**:e20.
24. Scott K, Brady R, Cravchik A, Morozov PS, Rzhetsky A, Zuker C, Axel R: **A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila.** *Cell* 2001, **104**:661-673.
25. LB V: **Diversity and expression of odorant receptors in Drosophila.** In Insect Pheromone Biochemistry and Molecular Biology Edited by G B, R V. Elsevier Academic Press; 2003:567-591.
26. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *Proc Int Conf Intell Syst Mol Biol* 1994, **2**:28-36.
27. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
28. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20**:426-427.
29. Bailey TL, Gribskov M: **Combining evidence using p-values: application to sequence homology searches.** *Bioinformatics* 1998, **14**:48-54.
30. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**:1188-1190.
31. Ohler U, Liao GC, Niemann H, Rubin GM: **Computational analysis of core promoters in the Drosophila genome.** *Genome Biol* 2002, **3**:Research0087.
32. Fang J, Haasl RJ, Dong Y, Lushington GH: **Discover protein sequence signatures from protein-protein interaction data.** *BMC Bioinformatics* 2005, **6**:277.
33. Janssen CS, Phillips RS, Turner CM, Barrett MP: **Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites.** *Nucleic Acids Res* 2004, **32**:5712-5720.
34. Clyne PJ, Warr CG, Carlson JR: **Candidate taste receptors in Drosophila.** *Science* 2000,

287:1830-1834.
35. Suh GSB, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ: **A single population of olfactory sensory neurons mediates an innate avoidance behaviour in Drosophila.** *Nature* 2004, **431**:854-859.
36. Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB: **Two chemosensory receptors together mediate carbon dioxide detection in Drosophila.** *Nature* 2007, **445**:86-90.
37. Kwon JY, Dahanukar A, Weiss LA, Carlson JR: **The molecular basis of CO2 reception in Drosophila.** *Proc Natl Acad Sci U S A* 2007, **104**:3574-3578.
38. Janin J, Chothia C: **The structure of protein-protein recognition sites.** *J Biol Chem* 1990, **265**:16027-16030.
39. Young L, Jernigan RL, Covell DG: **A role for surface hydrophobicity in protein-protein recognition.** *Protein Sci* 1994, **3**:717-729.
40. Jones S, Thornton JM: **Protein-protein interactions: a review of protein dimer structures.** *Prog Biophys Mol Biol* 1995, **63**:31-65.
41. Jones S, Thornton JM: **Analysis of protein-protein interaction sites using surface patches.** *J Mol Biol* 1997, **272**:121-132.
42. Clackson T, Wells JA: **A hot spot of binding energy in a hormone-receptor interface.** *Science* 1995, **267**:383-386.
43. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces.** *J Mol Biol* 1998, **280**:1-9.
44. Ma B, Elkayam T, Wolfson H, Nussinov R: **Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces.** *Proc Natl Acad Sci U S A* 2003, **100**:5772-5777.

Chapter 3. Duplication and differential expression of genes in an odorant receptor gene cluster in mosquitoes

3.1: Abstract

Introduction

The current model of insect olfactory signaling, derived mostly from work on *Drosophila melanogaster*, suggests that a critical step is the recognition of odors by cognate odorant receptors (ORs). In addition to *D. melanogaster*, a number of ORs have been identified in several insect genomes including those of the Africa malaria mosquito *Anopheles gambiae*. With a few exceptions, insect ORs are highly divergent within and between species, suggesting that these genes may be rapidly evolving. Most comparative work involving insect ORs has been done between highly divergent species with sequenced genomes. Comparative genomics analysis targeted at OR gene clusters from closely related *Anopheles* species provides an opportunity to evaluate evolution and function of these genes in mosquitoes.

Results

Here we report the isolation and characterization of genes in the genomic regions that contain the *Or2*/*Or10* gene cluster in *An. stephensi* and *An. quadriannulatus*. Multi-species comparison of these orthologous regions in *An. gambiae*, *An. quadriannulatus*, and *An. stephensi* revealed highly conserved gene structure among the OR genes and a novel OR gene *AsOr10x*, which is only present in *An. stephensi*. *AsOr10x* showed a different expression pattern than *AsOr2* and

AsOr10, the other members of this gene subfamily in *An. stephensi*. Therefore *AsOr10x* might be adapting or has adapted a new function. Analysis of the phylogeny and physical location of all known members of the *Or2*/*Or10* gene subfamily in *Anopheles*, *Aedes*, and *Culex* mosquitoes suggest that a few events of gene duplication and loss resulted in the current gene distribution.

Conclusion

We have identified and characterized five OR genes from *An. stephensi* and *An. quadriannulatus*, which are all members of the *Or2*/*Or10* gene subfamily. This gene subfamily has expanded in mosquito species through at least three duplication events when compared with fruit flies. More importantly, by comparing closely related species of the same genus, we were able to identify a gene duplication event (*Or10* and *Or10x*) that produced two differentially expressed OR genes that may have adapted different functions. This study demonstrates that OR gene duplication and loss can occur in closely related *Anopheles* mosquitoes, which may provide them with a rapidly changing gene set that could facilitate adaptation. In the long term, comparative analysis of OR gene clusters will help illustrate the underlying mechanism of behavioral differences between *Anopheles* mosquitoes.

3.2: Introduction

Mosquitoes transmit a number of pathogens that cause infectious diseases such as malaria, yellow fever, dengue fever, and West Nile encephalitis. These pathogens are transmitted when a female mosquito bites a host to extract blood for egg development. Most malaria vectors belong to the genus *Anopheles* and many are members of species complexes that include sibling species that are isomorphic or morphologically similar [1]. *Anopheles gambiae*, the primary malaria vector in Sub-Saharan Africa, is the founding member of the *An. gambiae* species complex that includes six additional reproductively isolated cryptic species, *An.* arabiensis, *An.* bwambae, *An.* merus, *An.* melus, *An. quadriannulatus* A, and *An. quadriannulatus* B [2, 3]. The *An. gambiae* species is referred to as *An. gambiae* sensu stricto, or *An. gambiae* s.s. (*An. gambiae* hereafter). Members of this species complex show differential host preferences [4, 5]. *An. gambiae* is highly anthropophilic, attracted to humans over other vertebrates such as cattle [2, 6, 7]. *An. quadriannulatus* has been reported as either highly zoophilic, preferring animals to humans [2, 5] or opportunistic [8, 9]. *An. stephensi*, the Asian malaria mosquito, is in the same subgenus Cellia as the *An. gambiae* species complex [10]. *An. stephensi* is attracted to humans, but it has also been shown to be attracted to large amounts of CO_2 and 1-octen-3-ol indicating an attraction to cattle [11]. There are several reports attempting to identify the specific odorants involved in attraction of these and other mosquito species to different host species [11-14].

Odorant binding directly or indirectly to odorant receptors (ORs) is thought to be the key step in insect olfactory signaling [15]. Insect ORs were first discovered in *Drosophila melanogaster* through bioinformatic approaches [16-18]. Insect ORs have been reported in the past to be putative G-protein coupled receptors (GPCRs) [16, 17, 19], but recently this classification has been questioned [20-22]. Subsequent studies in *D. melanogaster* have begun to identify the ligands bound by these receptors. One study in particular has identified potential ligands for most of the *D. melanogaster* ORs [23]. The identification of *Drosophila* ORs has helped in the identification of OR genes in several additional insect species including *An. gambiae* [19, 24, 25], *Aedes aegypti* [26, 27], *Bombyx mori* [28-30], *Apis mellifera* [31], and *Heliothis virescens* [32, 33]. With a few exceptions, insect ORs are highly divergent within and between species. Several mosquito [34], moth [28, 29], and honeybee [35] ORs have identified ligands. In mosquitoes there have been lineage specific expansion of some OR gene families and many ORs are physically clustered (e.g., [19, 27]). We are interested in using comparative genomics targeted at mosquito OR gene clusters to identify potential gene regulatory elements, and to shed light on evolutionary mechanism(s) of OR gene expansion/loss within a cluster. In the long term such comparative analysis will help illustrate the underlying mechanism of behavioral differences between closely related insect species such as the mosquitoes of the *An. gambiae* species complex and the Cellia subgenus.

The *An. gambiae Or2* (*AgOr2*) gene cluster contains two genes, *AgOr2* and *AgOr10*, which form a monophyletic clade that is related to *DmOr43a*, a *D. melanogaster* homolog [19, 24]. The potential ligands of both *AgOr2* and *DmOr43a* have been identified in several studies [34, 36, 37]. Thus we decided to initiate a comparative genomics study of the *Or2* gene cluster in different mosquito species to trace the evolutionary history of this gene cluster. We have screened bacteria artificial chromosome (BAC) DNA libraries to locate the *Or2* gene cluster in two other mosquito species, *An. stephensi* and *An. quadriannulatus*. Comparative analysis between *An. gambiae*, *An. quadriannulatus*, and *An. stephensi* as well as *Ae. aegypti* and *Culex pipiens quinquefasciatus* suggest that genes within the *Or2* cluster have undergone at least three duplication events. We have also identified a new member of the *Or2*/*Or10* gene subfamily, which is found only in *An. stephensi*. This gene, which we named *AsOr10x*, has a different expression profile than the other genes in the *Or2*/*Or10* subfamily, indicating it might have adapted or is adapting a new function. This study demonstrates that OR gene duplication and loss can occur in closely related *Anopheles* mosquitoes. Thus the birth-and-death process as described for *Drosophila* OR genes [38] may also be used in mosquitoes which provides a rapidly changing gene set that could facilitate adaptation.

3.3: Results

3.3.1: Identification and annotation of the *Or2* gene cluster in *An. stephensi* and *An. quadriannulatus*

We designed gene specific primers to amplify a 209bp region of the third exon of *AgOr2*. Polymerase chain reaction (PCR) products were successfully obtained using the above primer set and *An. gambiae* and *An. stephensi* genomic DNA as the templates. Sequencing confirmed that the PCR products are derived from *Or2* genes in their respective species and they are 85.9% identical to each other. The *Or2* fragment obtained from *An. stephensi* was then labeled with digoxigenin (DIG) in an asymmetric PCR reaction. The labeled probe was used to screen BAC DNA libraries created from *An. stephensi* and *An. quadriannulatus* genomic DNA. In the *An. quadriannulatus* BAC library we identified five positive clones among 9,216 clones screened. In the *An. stephensi* BAC library we identified six positive clones among 18,432 clones screened. Amplicon Express (Pullman, WA) sequenced one positive clone from each library. Sequencing of the *An. quadriannulatus* BAC colony insert produced approximately 148 kb of DNA sequence in five contigs. The *An. stephensi* BAC clone insert produced approximately 102 kb of DNA sequence in four contigs.

We used the program mVISTA [39] to perform global nucleotide alignments of the *Or2* clusters in *An. gambiae* (*AgOr2*), *An. stephensi* (*AsOr2*), and *An. quadriannulatus* (*AqOr2*) (Figure 3.1). The presence of well aligned *Or2* and *Or10* as well as neighboring genes suggest that the two sequenced BACs are orthologous to

the reference *AgOr2* cluster. Fluorescent *in situ* hybridization (FISH) results of *An. stephensi* polytene chromosomes have shown that the positive *An. stephensi* BAC hybridizes specifically to the 3R chromosome (band 3R:31B) (Igor Sharakhov, Virginia Tech, personal communication). This suggests all *An. stephensi Or2*/10 genes are localized at one discrete loci, and that our sequencing of one BAC clone provides us with all *An. stephensi Or2*/10 sequences. The BAC that contains *AqOr2* showed an approximately 90% identity on average at the nucleotide level to the *AgOr2* region with an even higher conservation in the areas predicted to be genes. There are several areas of sequence that have no detectable conservation between *An. gambiae* and *An. quadriannulatus*, which represents insertion/deletions (indels) between the two species (Figure 3.1). The BAC that contains *AsOr2* showed an average 85% nucleotide conservation in predicted coding sequences of the *AgOr2* region, but there are several areas where indels are present. One of these is a 3.35kb indel immediately downstream of the predicted *AsOr10* gene (Figure 3.1, black bar). This indel is present in *An. stephensi*, but not in either *An. gambiae* or *An. quadriannulatus*. tBLASTn [40] analysis indicates this region is significantly similar to *AgOr10* with an e-value of 6.0 x10-49, thus this may be a duplicated paralog of *Or10*. Further tBLASTn analyses using the peptide sequences of *AgOr2* and *AgOr10* as well as the other 77 *An. gambiae* ORs as query against the entire BAC sequences of both species revealed no additional OR sequences. Thus we have found one *Or2* and one *Or10* gene in *An.*

quadriannulatus, and one *Or2* and two *Or10* genes in *An. stephensi*. The copy of the *An. stephensi Or10* nearest to *AsOr2* is phylogenetically closer to *AgOr10* than the other copy (see phylogenetic analysis below). Thus we named the two *An. stephensi Or10* genes *AsOr10* and *AsOr10x* respectively. *AsOr10x* had a premature stop codon in the predicted exon 2 according to the BAC sequence. To verify this, we sequenced PCR products covering the *AsOr10x* coding region, which were obtained by using *An. stephensi* genomic DNA as template. These sequencing results revealed that the *An. stephensi* BAC DNA sequence was missing a 109bp piece that resulted in an artificial stop codon in exon 2 of the *AsOr10x* gene. This is likely due to a mis-assembly of our BAC DNA sequence. The corrected *AsOr10x* sequence is used in all subsequent analysis. We performed additional PCR with *An. stephensi* genomic DNA as template using a primer based in the *AsOr10* coding sequence and another primer based in the *AsOr10x* coding sequence. We were able to obtain a PCR product consistent with the positions of *AsOr10* and *AsOr10x* genes as predicted. Thus *AsOr10x* is not an artifact of sequencing or sequence assembly.

AsOr2 and *AsOr10* showed 94.1% and 93.9% amino acid identity to their respective *An. gambiae* homologs (Table 3.1). *AsOr10x* showed 77.7% amino acid identity to *AsOr10*, which is significantly higher than the 51.2% amino acid identity it shares with *AsOr2*. *AqOr2* and *AqOr10* showed higher amino acid identities at 98.7% and 99.7% to their respective *An. gambiae* homologs. It is important to note there is a

sequencing/assembly gap in our *AqOr2* BAC sequence, which is predicted to be approximately 2kb according to the Amplicon Express assembly. This 2kb region is located in the sixth (last) intron of the *AqOr2* gene. We have represented this region in Figure 3.2 as a dashed wedge.

3.3.2: Phylogenetic and genomic analyses suggest multiple duplication events in the mosquito *Or2* gene subfamily.

Phylogenetic analysis of mosquito *Or2* and *Or10* peptides was performed to illuminate the evolutionary relationship between *AsOr10x* and other members of the *Or2* gene subfamily. Included in this analysis were the peptide sequences of the *Or2* and *Or10* genes of *An. gambiae*, *An. quadriannulatus*, *An. stephensi*, *Ae. aegypti*, and *C.p. quinquefasciatus*. The *Ae. aegypti* OR genes were just recently reported [27], and the *C.p. quinquefasciatus* OR genes were obtained from the newly released genome assembly (http://cpipiens.vectorbase.org/SequenceData/Genome/). Detailed information regarding exon/intron boundaries of all mosquito *Or2*/10 genes were obtained using the programs GeneQuest (DNASTAR, Inc. Madison, WI) and Genewise [41], along with comparisons to current annotations, and our subsequent cDNA cloning. Exon/intron boundaries were confirmed by sequence data obtained from reverse transcription polymerase chain reaction (RT-PCR) experiments discussed below. The structure of all *Or2*, *Or10*, and *Or10x* genes is very similar (Figure 3.2). All of these genes except Cp*Or2* have the same first and last introns suggesting they are

ancestral to this subfamily. Introns are typically very small at 60-80 bp with a few exceptions. The peptide sequences coded by these genes and *DmOr43a*, the *D. melanogaster* ortholog of the *Or2* gene subfamily were aligned using ClustalX [42] (Figure 3.3). The alignment showed high amino acid conservation overall, which may be a characteristic of these closely related mosquito species. *DmOr30a*, *DmOr49b*, *AgOr21*, *AgOr26*, and *AgOr53* peptide sequences were then added as outgroup sequences and all sequences were realigned [19, 24].

Phylogeny was inferred from this alignment using the Bayesian program MrBayes [43]. It is clear that *DmOr43a* and the mosquito *Or2* (*Or2*/*Or10*) gene subfamily form a well-supported monophyletic clade (Figure 3.4A). Within the mosquito *Or2* (*Or2*/*Or10*) gene subfamily, *Or2* and *Or10* form two distinct clades. There is a single copy of the *Or2* gene in each mosquito species, and the phylogeny of the *Or2* genes perfectly reflects the mosquito phylogeny [10, 44]. In the *Or10* clade, there are two *Or10*-type genes in *An. stephensi*, *Ae. aegypti* and *C.p. quinquefasciatus*. For the two *Ae. aegypti* ORs in the *Or10* clade, we are using the nomenclature of Bohbot et al. [27], which referred to them as *AaOr9* and *AaOr10*. It appears that there are at least three gene duplication events that are responsible for the current distribution of genes in the *Or2* gene cluster. The first duplication (D1, Figure 3.4A) created the *Or2* and *Or10* clades, which happened after the separation between *D. melanogaster* and the Culicidae. The second and third duplications (D2 and D3, Figure

3.4A) happened after the separation between Culicinae (*Aedes*/*Culex*) and Anophelinae mosquitoes. One is responsible for duplicating *Or10*/*Or10*.1 and Or9/*Or10*.2 in *Aedes*/*Culex* species (D2, Figure 3.4A), and the other is responsible for duplicating *Or10* and *Or10x* in the Anophelinae species (D3, Figure 3.4A). The *Or10x* gene appears to have been lost in *An. gambiae* and *An. quadriannulatus*. Gene conversion is unlikely as there are sufficient levels of divergence between the different *Or2*/*Or10* genes within a given species (Table 3.1), and the relative similarities between these genes are higher in the exons than in the introns as expected (data not shown).

Figure 3.4B illustrates the current physical relationship of the *Or2* gene cluster in *An. gambiae*, *An. stephensi*, and *Ae. aegypti*. The genes flanking *Or2* and *Or10* are conserved in all three species and they are in the same orientation. Therefore it is likely that these genes were present and linked in the last common ancestor of Anophelinae and Culicinae. The different organizations of the *Or2* cluster between *An. gambiae* and *An. stephensi* may be explained by a simple tandem duplication of the *Or10* gene. The different organizations of the *Or2* cluster between *An. gambiae* and *Ae. aegypti* is more difficult to deduce and may involve duplications as well as chromosomal inversions that did not effect the neighboring genes. There is also potentially another duplication occurring independently in the *Culex* lineage because there is a third *Or10* gene in the current assembly of *C.p. quinquefasciatus*, which is decayed with multiple stop

codons. Given the number of stop codons in this sequence, we decided not to include it in further analysis.

3.3.3: Expression analysis of *AsOr2*, *AsOr10*, and *AsOr10x* and intron retention in *AsOr10x*.

The expression profile of *AsOr2*, *AsOr10*, and *AsOr10x* in *An. stephensi* mosquitoes at various time points was determined using non-quantitative RT-PCR (Figure 3.5 and Figure 3.6). Primer sequences are shown in Table 3.2 and the exons covered by these primers are illustrated in Figure 3.2. These time points were 1st instar larvae, pupae, 4-day-old sugar fed adult females, 4-day-old adult females that were 24 hr post blood feeding, and 4-day-old sugar fed adult males. Total RNA samples were obtained from dissected adults antenna, maxillary palp, and proboscis as well as whole larvae and pupae. *AsOr2* and *AsOr10* were found to have very similar expression patterns as far as could be determined with non-quantitative RT-PCR (Figure 3.5, Figure 3.6, and Table 3.3). Both genes were robustly expressed in adult antenna and maxillary palp/proboscis. Both genes were not expressed in adult bodies without appendages but appeared to be expressed in adult wings (data not shown). There is no apparent difference between males and females and between sugar-fed and blood-fed females. Both *AsOr2* and *AsOr10* were expressed in larvae and pupae although the mRNA level of *AsOr2* appeared to be low in pupae (Figure 3.6). *AsOr10x* has a significantly different expression pattern from *AsOr2* and *AsOr10*. The main

difference is that although *AsOr10x* is robustly expressed in adult antenna of male as well as sugar-fed and blood-fed females, it is not expressed in adult maxillary palp or proboscis (Figure 3.5). RT-PCR using *AsOr10x* primers revealed three products at 600bp, 510bp, and 450bp. Cloning and sequencing of *AsOr10x* RT-PCR products showed that the 600bp fragment was an *AsOr10x* product primed from genomic DNA contamination. The remaining two products were amplified from cDNA templates. The predominant 450bp product was from *AsOr10x* cDNA with all introns excised while the 510bp product was from *AsOr10x* cDNA with one intron retained (Figure 3.2 and Figure 3.5). The *AsOr10x* RT-PCR was repeated twice and the 510bp remained recognizable in both cases. The addition of the intron leads to a premature stop codon and this transcript could only code for a substantially smaller protein of 284 amino acids. Further experimentation is needed to determine if this transcript serves any biological role in olfactory perception.

3.4: Discussion

We have identified and characterized five OR genes in the *Or2/Or10* cluster from *An. stephensi* and *An. quadriannulatus* through the use of BAC libraries. We have shown that the *Or2/Or10* gene cluster has expanded in mosquito species through at least three duplication events when compared with fruit flies. Lineage-specific OR gene expansion has been previously noted when comparing *An. gambiae* with *D.*

melanogaster [19] and *An. gambiae* with *Ae. aegypti* [27] gene repertoires. However, even the comparison between the *Anopheles* and *Aedes* mosquitoes is between two divergent mosquito genera spanning 145-200 MYA of evolution. Thus the resolution or power of such comparisons are limited especially considering the fast-evolving nature of the OR gene family. A recent study, on the other hand, compared ORs from the 12 sequenced *Drosophila* genomes and identified frequent gains and losses of OR genes through a "birth-and-death" process [38]. Our current study focuses on the *Or2* gene cluster and included both divergent and closely related mosquito species. We uncovered the complexity of gene duplication and loss that led to the current distribution and expansion of the cluster. In the case of *Anopheles Or10* and *Or10x* genes, we have shown evidence of gene duplication and loss within closely related *Anopheles* mosquitoes. Such a process may provide mosquitoes with a rapidly changing gene set that could facilitate adaptation. Such duplication and loss could also result in lineage-specific and even species-specific distribution of OR genes. For example, *AsOr10x* is only found in the Asian malaria mosquito *An. stephensi*, not in *An. gambiae* or *An. quadriannulatus*. It is possible that our library screening has not uncovered all copies of *Or2*/*Or10* in *An. stephensi* and *An. quadriannulatus*, but this is unlikely. Preliminary FISH data shows the localization of the *An.* stephensi *Or2* BAC to one region on the 3R chromsome. In both the *An. stephensi* and *An. quadriannulatus* BACs there are conserved genes flanking the ORs found in all three species indicating

these are orthologous regions. One or both species may have additional *Or2* and *Or10* genes at different loci, but closely related OR genes in mosquitoes are located within gene clusters and are not randomly distributed throughout the genome [19, 27]. BAC Library screening of targeted ORs can allow us to identify targeted and closely related ORs in many mosquito species without the drawbacks of whole-genome sequencing.

Another interesting finding of this study is that the above mentioned gene duplication event appears to have created genes that are differentially expressed and thus these genes may have adapted different functions. The expression of *AsOr2*, *AsOr10*, and *AsOr10x* mRNA is persistent throughout the mosquito life cycle, appearing early in larvae and continuing to adults. Although *AsOr2* and *AsOr10* have accumulated significant differences in their amino acid sequences they have a similar expression pattern. However, *AsOr10x* has a different expression pattern in the adults compared with *AsOr2* and *AsOr10*. In contrast to *AsOr2* and *AsOr10*, *AsOr10x* is not expressed in the maxillary palp or proboscis. In *An. gambiae*, *AgOr2* mRNA is detected in both male and females in the antenna and/or maxillary palp but not the proboscis [24]. *AgOr10* mRNA is also expressed in antenna and/or maxillary palp, but there is no data as to if it is sexually dimorphic [19]. In *Ae. aegypti*, *AaOr2* and *AaOr10* are expressed in both male and female adult antenna, and additionally *AaOr2* is present in larvae antenna [27]. *AaOr9* mRNA is only found in larvae antenna. In contrast to the mosquito *Or2*/10 genes, the *D. melanogaster* ortholog, *DmOr43a*, is

expressed in pupae and in adult antenna, but not in larvae, not in adult maxillary palp, or proboscis [36, 45, 46]. The expression profile of members of the *Or2/Or10* gene subfamily is summarized in Table 3.3. There is no obvious conservation of the expression profile of members of this subfamily [19, 24, 27].

Existing data suggest that members of the *Or2/Or10/DmOr43a* subfamily encode receptors that bind benzaldehyde [36, 37] and structurally similar compounds [34]. Studies have indicated that benzaldehyde generates an avoidance response in *D. melanogaster* that is not sexually dimorphic [47, 48], and this behavior is dependent on *DmOr43a* [46]. One behavioral report found that the presence of the structurally related chemical 2-methylphenol makes ovipositioning sites more attractive for both the mosquitoes *Toxorhynchites moctezuma* and *Toxorhynchites amboinensis* [49]. Another electrophysiological study confirms a response to 2-methylphenol in *An. gambiae* [50]. Further behavioral studies will be necessary to understand the biological context of these responses.

Another application of the comparative genomics analysis between gene sequences among closely related species is the identification of regulatory elements. Although we have not discussed it explicitly in this paper, our comparison of the upstream sequences of *Or2/Or10* genes has revealed many conserved non-coding sequences, which may be the regulatory elements of these genes (data not shown). Only two papers report information on OR gene regulation. One of these papers

reports that the *D. melanogaster* Acj6 transcription factor regulates a subset of *D. melanogaster* OR genes [16]. The second and very elegant study identified several elements that contribute to tissue and neuron specificity of some *D. melanogaster* ORs [51]. This study took advantage of the completed expression profile of virtually all *D. melanogaster* ORs. As the expression profile as well as comparative genomics data accumulate, it will become feasible to systematically identify potential regulatory sequences that control mosquito OR gene expression. A better understanding of the evolutionary change of the OR gene repertoire as well as their regulation will shed light on the genetic basis of mosquito host-seeking behavior and thus contribute to the control of mosquito-borne infectious diseases.

3.5: Conclusions

We have identified and characterized five OR genes from *An. stephensi* and *An. quadriannulatus*, which are all members of the *Or2/Or10* gene subfamily. This gene subfamily has expanded in mosquito species through at least three duplication events when compared with fruit flies. More importantly, by comparing closely related species of the same genus, we were able to identify a gene duplication event (*Or10* and *Or10x*) that produced two differentially expressed OR genes that may have adapted different functions. This study demonstrates that OR gene duplication and loss can occur in closely related *Anopheles* mosquitoes, which may provide them with

a rapidly changing gene set that could facilitate adaptation. In the long term, comparative analysis of OR gene clusters will help illustrate the underlying mechanism of behavioral differences between *Anopheles* mosquitoes.

3.6: Experimental Procedures

BAC library construction, screening, and sequencing. Amplicon Express (Pullman, WA) prepared BAC libraries from genomic DNA of *An. stephensi* Indian strain and *An. quadriannulatus* using a previously reported method [52]. The *An. stephensi* BAC library constructed has clones with an average insert size of 125.0kb giving this library a 9x fold coverage. The *An. quadriannulatus* BAC library had clones containing an average insert size of 123.0kb giving this library a 17x fold coverage. Screening of both libraries was accomplished by using a randomly DIG-labeled single-stranded DNA probe. The template for the probe was a ~200bp region of the *AsOr2* gene amplified by using the primers 2F1 5' CCTGCTTTGTGACCTATC 3' and 2R1 5' GGCCGCTATCTGGACGAG 3' and *An. stephensi* genomic DNA. The template, DIG DNA labeling mix (Roche Basel, Switzerland), primer 2F1, and Taq polymerase (Takara, Otsu, Japan) were used in an asymmetric PCR reaction that produced randomly DIG-labeled single-stranded probe. Probe concentration was determined by following the procedure "Estimating the Yield of DIG-labeled DNA" listed in The Genius System User's Guide for Filter Hybridization Version 2.0 (Boehringer Mannheim Corporation,

Indianapolis, IN). Determining probe specificity and library screening was accomplished following a standard protocol. 9,216 *An. quadriannulatus* and 18,432 *An. stephensi* clones were screened and this reflects a 4x fold and 9x fold coverage respectively. We found five and six clones respectively, which is close to what was predicted. One positive *An. stephensi* and *An. quadriannulatus* BAC were submitted to Amplicon Express for sequencing. Sequencing of the *An. stephensi* BAC colony insert produced approximately 102 kb of DNA sequence in four contigs. Sequencing of the *An. quadriannulatus* BAC colony insert produced approximately 148 kb of DNA sequence in five contigs.

PCR and Re-sequencing of the *AsOr10x* gene. A ~4 kb fragment spanning from *AsOr10* to *AsOr10x* was obtained through PCR using the primers *Or10*LA-F1 5' TGTTTGGGTTGGTGCAGATA 3' and *Or10x*LA-R1 5' GAAGATGCTTATGCCCGGTA 3', *An. stephensi* genomic DNA as template, and LA (long accurate) Taq polymerase (Takara). A smaller fragment (~1 kb) was obtained using the primers *Or10x*SeqF 5' CTACCGAAGGATTGCCAGAC 3' and *Or10x*SeqR 5' TTATGCGCAAGTGATCTTCG3', *An. stephensi* genomic DNA as template, and Taq polymerase (Takara). The smaller fragment was cloned via TA cloning into the pGEM T-Easy vector (Promega, Madison, WI) and sequenced.

Phylogenetic analysis. Phylogenetic analysis was carried out using peptide sequences from mosquito *Or2* and *Or10* genes. *AgOr2* and *AgOr10* peptide

sequences were obtained from the paper by Hill et al [19]. *AqOr2*, *AqOr10*, *AsOr2*, *AsOr10*, and *AsOr10x* sequences were obtained through BAC library screening as described in detail above. The *Ae. aegypti AaOr2*, *AaOr9*, and *AaOr10* sequences were obtained from a recent report [27]. The *C.p. quinquefasciatus* OR sequences were obtained from BLAST [40] analysis of the *C.p. quinquefasciatus* genome assembly (assembly version: CpipJ1) located on the Vectorbase website (http:/ /cpipiens.vectorbase.org/SequenceData/Genome/). In addition to the mosquito *Or2* and *Or10* genes the *D. melanogaster* OR peptides *DmOr43a*, *DmOr30a*, and *DmOr49b* were included as the closest related non-mosquito ORs, and these sequences were also obtained from another report [53]. *AgOr21*, *AgOr26*, and *AgOr53*, which are related to *AgOr2*/*AgOr10* [19], were also included in our analysis.

Peptide sequences were aligned by the ClustalX program (v1.83.1) [42] installed on a MacBook running Mac OS 10.4.9. All parameters were left at default values except for the following changes: pairwise gap opening penalty was set to 35, multiple alignment gap opening penalty was set to 15, and multiple alignment gap extension penalty was set to 0.3. Two alignments were made one containing all of the sequences mentioned above, and one containing all of the mosquito *Or2*/10 sequences and *DmOr43a*. This later alignment was made into Figure 3.3. via importing into Jalview [54] for enhancement of presentation value. The former alignment was saved to the nexus format. The nexus file header was altered to make it conform to

the modified nexus format used by the program MrBayes [43]. MrBayes version 3.1.2 was complied on a PC with an Athlon 64-bit processor, 2 gigs of RAM, and running Ubuntu Linux version 7.04. The nexus file was imported into MrBayes and the prior was set with the command prset aamodelpr=mixed. This allowed MrBayes to use multiple fixed rate amino acid models in inferring a phylogenetic tree. The program was run for 100,000 generations (command mcmcp ngen=100000) with a sampling frequency of every 100 generations. A tree consensus file was generated using the commands sump burnin=250 and sumt burnin=250. This file was imported into Treeview (v1.6.6) [55] for visualization. Only labels and probability values were moved around for the sake of clear presentation.

ClustalX Alignment. Amino acid sequences of *Or2*/10 genes from *An. gambiae*, *An. quadriannulatus*, and *An. stephensi* were aligned using ClustalX version 2.0 [42] with default parameters. Alignment was imported into Jalview [54] for enhancement of presentation value.

Mosquitoes. *An. stephensi* Indian strain mosquitoes were used in all experiments. Mosquitoes were reared in an incubator at 27°C, ~80% humidity, and with a 12 hour light and 12 hour dark cycle. Larvae were feed for two days on 3mL of 66.6% Sera Micron Fry Food and 33.3% brewer's yeast suspended at 1g/50mL. Adults were fed a 10% sugar solution from soaked cotton balls. The blood source for the female

mosquitoes were mice, and all females that were fed blood did so on the third day post emergence.

RT-PCR of *AsOr2*, *AsOr10*, and *AsOr10x*. Total RNA for RT-PCR experiments was obtained from *An. stephensi* mosquitoes at various time points. Larvae were collected within a few hours of emergence, while pupae were collected 24 hours after start of pupation to insure an approximately equal amount of males and females. Adults chemosensory tissues were dissected from ~200 live four-day-old sugar-fed female, four-day-old 24 hours post-blood feeding female, and four-day-old sugar-fed male *An. stephensi* mosquitoes. Tissues were divided into two fractions: 1) antenna and 2) maxillary palp plus proboscis. Tissue samples were processed in 80µL of Trizol (Invitrogen, Carlsbad, CA) following the standard protocol included with the reagent, and total RNA pellets were resuspended in 20µL of DEPC treated H_2O. Approximately 1.3µg of total RNA from each sample was DNaseI treated using the DNA-free kit (Ambion Austin, TX). These DNaseI treated total RNA samples were then divided into two equal pools one pool for a negative (lacking reverse transcriptase) control and one for cDNA synthesis. Two sets of negative control and cDNA template were used in the course of these RT-PCR experiments. Two different sets of *AsRPS4* RT-PCR reactions were done to indicate this. The SuperScript II Reverse Transcriptase kit (Invitrogen) was then used according to that kit's protocol with the exception of a 90 not 50 minute incubation at 42°C to generate cDNA template.

The gene-specific primer sets used in RT-PCR are listed in Table 3.2. These gene-specific primer sets all have a melting temperature of 60°C, and span at least one intron to distinguish between products primed from genomic DNA or cDNA template. PCR was then performed using 1µL (*AsRPS4*, *AsOr2*, and *AsOr10*) or 2µL (*AsOr10x*) of template from each pool, a gene-specific primer set, and Taq polymerase (Takara). 25 cycles at 60°C were used for *AsRPS4* RT-PCR reactions while 35 cycles were used for *AsOr2*, *AsOr10*, and *AsOr10x* RT-PCR reactions. The lower number of cycles for *AsRPS4* RT-PCR was to prevent saturation of the *AsRPS4* product. Products that lacked introns according to size prediction were gel purified and cloned into pGEM T-Easy vector (Promega) and sequenced. In this way cDNA products were confirmed to be valid for all genes examined (*AsRPS4*, *AsOr2*, *AsOr10*, and *AsOr10x*).

List of Abbreviations

Anopheles gambiae odorant receptor, AgOr; *Anopheles quadriannulatus* odorant receptor, AqOr; *Anopheles stephensi* odorant receptor, AsOr; bacterial artificial chromosome, BAC; dioxygenin, DIG; polymerase chain reaction, PCR; fluorescent in situ hybridization, FISH; G-protein coupled receptor, GPCR; reverse transcription polymerase chain reaction, RT-PCR; odorant receptor, OR;

3.7: Figures

Figure 3.1: VISTA alignment of *An. gambiae Or2* gene cluster with *An. stephensi* and *An. quadriannulatus Or2* gene clusters. A VISTA alignment is a global nucleotide alignment with a graphical overview. The first row contains two brackets, and each bracket represents a pairwise alignment that is continued in its respective bracket on the second row. The upper bracket of each row is the alignment between the *AgOr2* and *AqOr2* gene clusters, while the bottom bracket of each row is the alignment between the *AgOr2* and AsOr2 gene clusters. The height of a peak indicates the amount of nucleotide conservation within a 100bp reading frame around that point. Pink peaks are those peaks above a 70% nucleotide conservation threshold, and are considered conserved non-coding sequences (CNSs). Blue peaks are predicted coding sequences. There is one ~2kb gap in the final intron of the *AqOr2* gene in the *An. quadriannulatus* assembly as indicated by a gap in the thick grey line below the peaks. *An. gambiae* and *An. quadriannulatus* are expected to have high nucleotide conservation as they are sibling species. The region indicated by the thick black bar is the location of the *AsOr10x* gene.

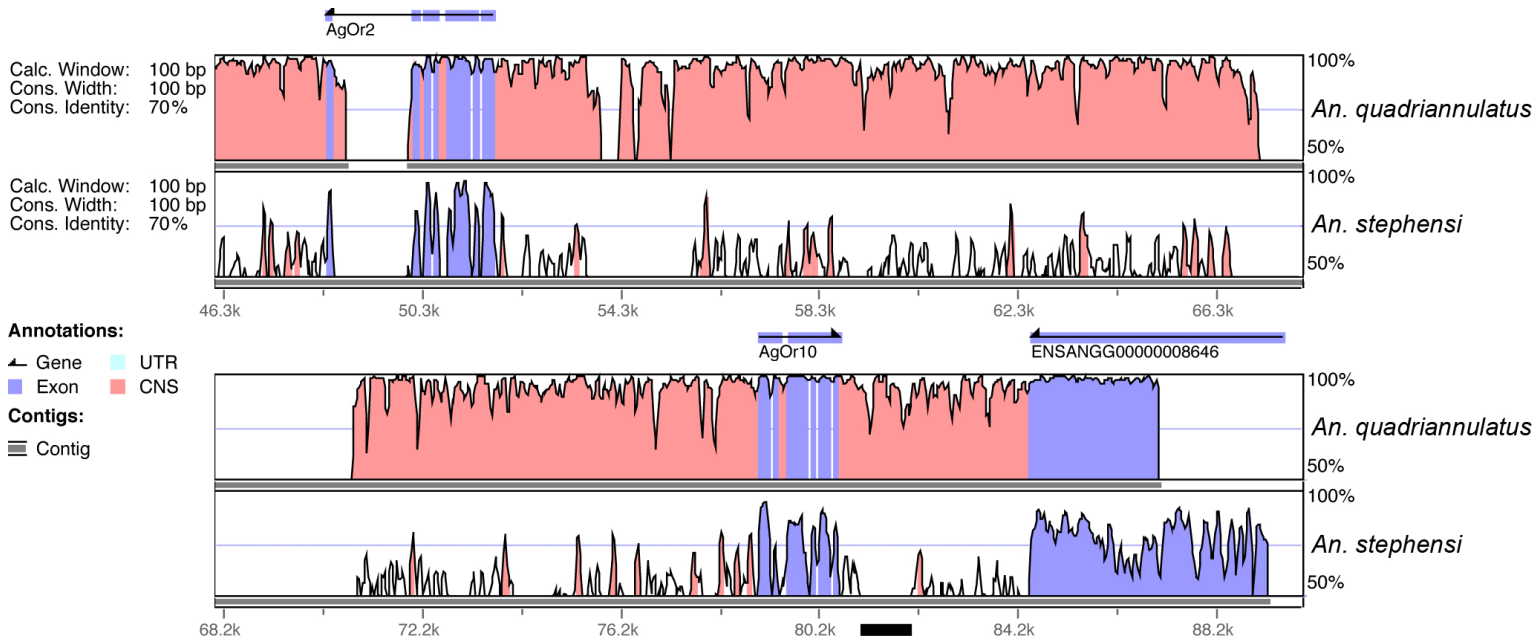

AgOr2
Calc. Window: 100 bp
Cons. Width: 100 bp
Cons. Identity: 70%
100%
An. quadriannulatus
50%
Calc. Window: 100 bp
Cons. Width: 100 bp
Cons. Identity: 70%
100%
An. stephensi
50%
46.3k
50.3k
54.3k
58.3k
62.3k
66.3k
Annotations:
Gene
UTR
Exon
CNS
Contigs:
Contig
AgOr10
ENSANGG00000008646
100%
An. quadriannulatus
50%
100%
An. stephensi
50%
68.2k
72.2k
76.2k
80.2k
84.2k
88.2k

Figure 3.2: *Or2* and *Or10* gene structure diagrams. Diagrams are drawn approximately to scale with boxes representing exons and wedges representing introns. RT-PCR products were obtained from all three of the *An. stephensi Or2*/*Or10* genes. These RT-PCR products are represented by diagrams below their respective *An. stephensi* OR genes. In the case of *AsOr10x* there were two products, a RT-PCR product with one intron retained and a product with no introns.

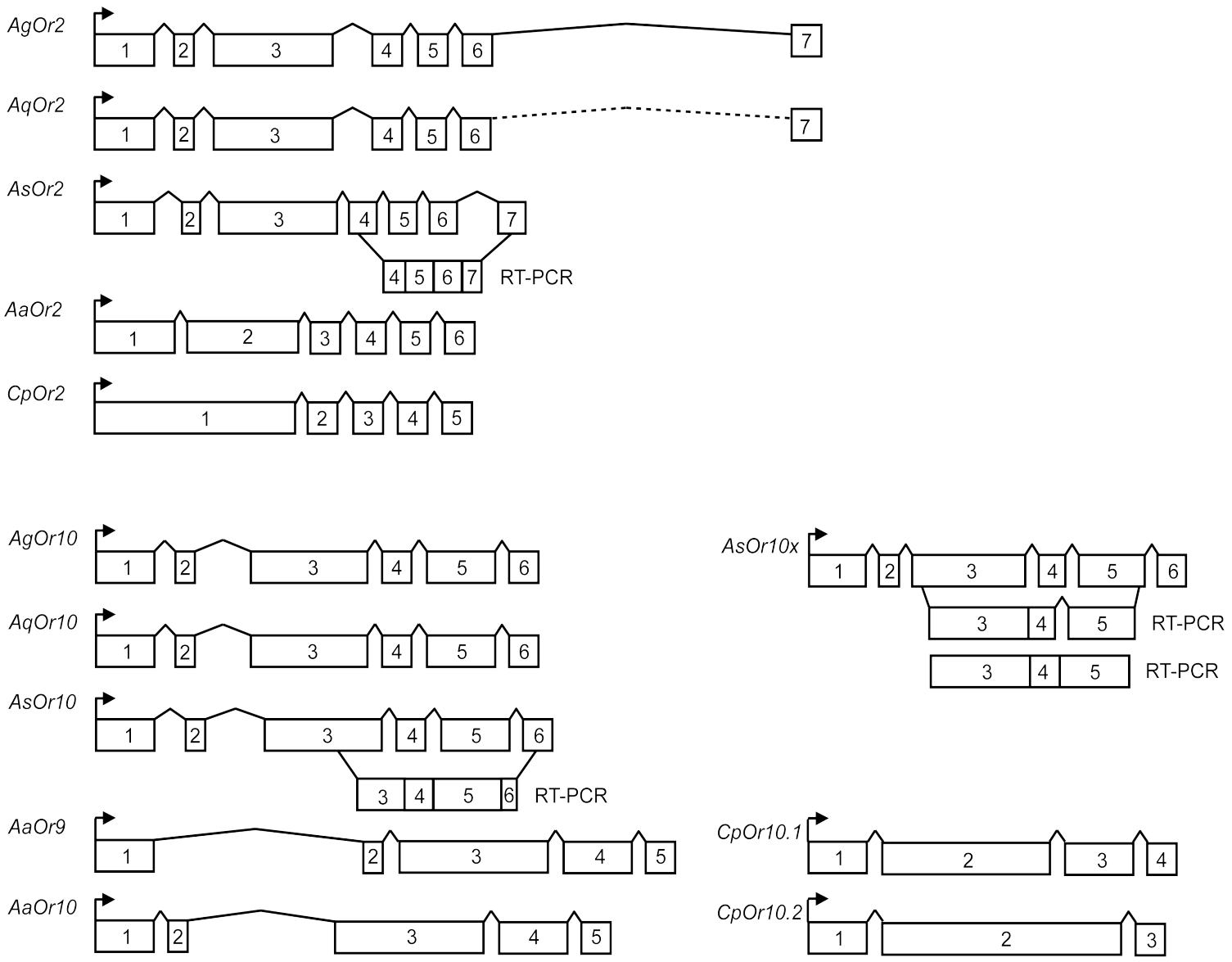

AgOr2
1
2
3
4
5
6
7
AqOr2
1
2
3
4
5
6
7
AsOr2
1
2
3
4
5
6
7
4
5
6
7
RT-PCR
AaOr2
1
2
3
4
5
6
CpOr2
1
2
3
4
5
AgOr10
1
2
3
4
5
6
AqOr10
1
2
3
4
5
6
AsOr10
1
2
3
4
5
6
3
4
5
6
RT-PCR
AaOr9
1
2
3
4
5
AaOr10
1
2
3
4
5
AsOr10x
1
2
3
4
5
6
3
4
5
RT-PCR
3
4
5
RT-PCR
CpOr10.1
1
2
3
4
CpOr10.2
1
2
3

Figure 3.3: ClustalX alignment of *Or2* subfamily genes from *An. gambiae*, *An. quadriannulatus*, *An. stephensi*, *Ae. aegypti*, *C.p. quinquefasciatus*, and *D. melanogaster*. Alignment was visualized using Jalview. The more conserved a residue is at a given position the more intense the blue coloring.

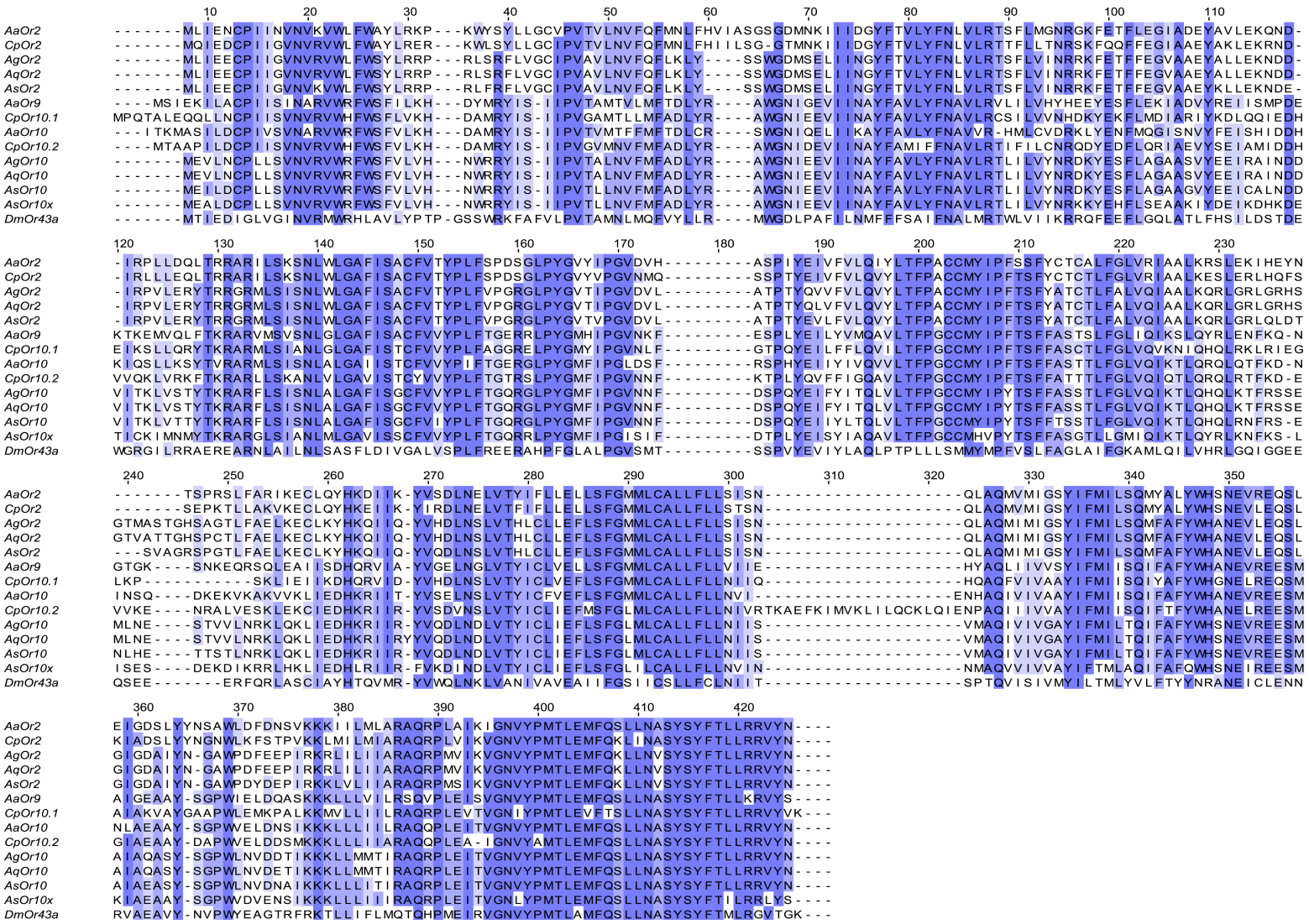
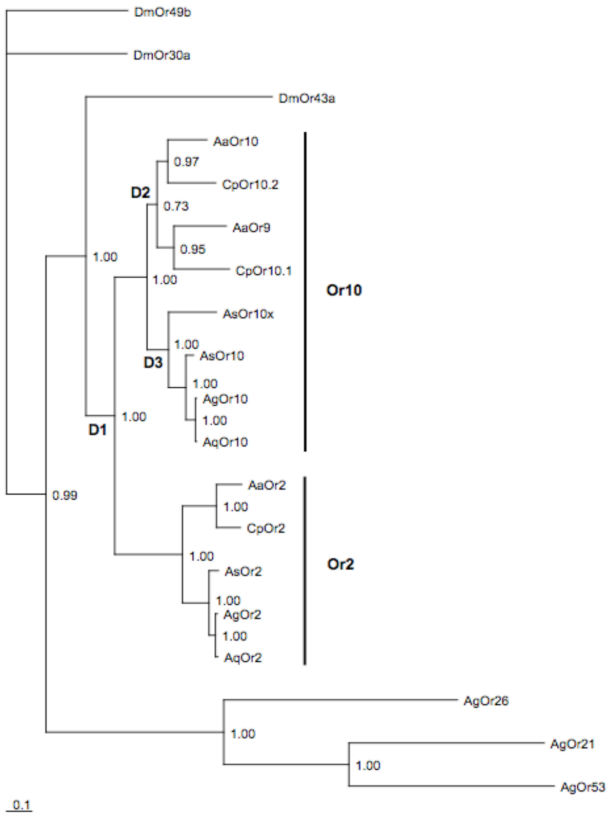

AaOr2
CpOr2
AgOr2
AqOr2
AsOr2
AaOr9
CpOr10.1
AaOr10
CpOr10.2
AgOr10
AqOr10
AsOr10
AsOr10x
DmOr43a

Figure 3.4: Phylogenetic tree of *Or2* and *Or10* gene subfamily in four mosquito species. A) Peptide sequences of all *Or2* and *Or10* genes of *An. gambiae*, *An. stephensi*, *Ae. aegypti*, and *C.p. quinquefasciatus* were aligned with closely related *D. melanogaster* OR genes. *AgOr26*, *AgOr21*, and *AgOr53* were included as an outgroup. This phylogenetic tree was generated from the Bayesian phylogenetic program MrBayes. Values displayed are posterior probabilities indicating likelihood of branch placement. Scale at the bottom left is for substitutions per site. B) The orientation of genes in and near the *Or2*/*Or10* gene cluster in *An. gambiae*, *An. stephensi*, and *Ae. aegypti*. Genes are represented as boxes, and boxes above the line are genes read left to right and vice versa for boxes below the line. Annotations for non-olfactory genes come from the Ensembl database when provided. Arrows connect orthologous genes in different species.

A)


DmOr49b
DmOr30a
DmOr43a
AaOr10
0.97
CpOr10.2
D2
0.73
AaOr9
0.95
CpOr10.1
1.00
1.00
Or10
AsOr10x
1.00
D3
AsOr10
1.00
AgOr10
1.00
AqOr10
1.00
D1
AaOr2
1.00
CpOr2
0.99
1.00
AsOr2
Or2
1.00
AgOr2
1.00
AqOr2
AgOr26
1.00
AgOr21
1.00
AgOr53
0.1

B)

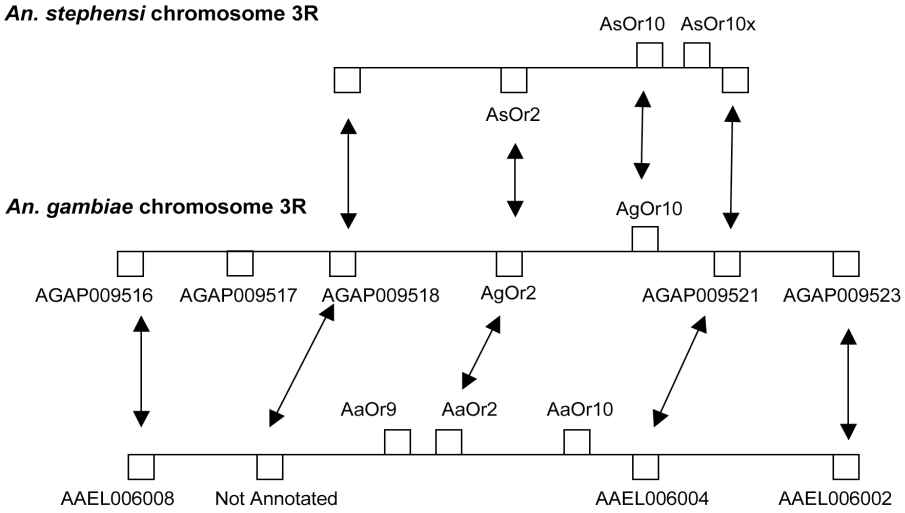
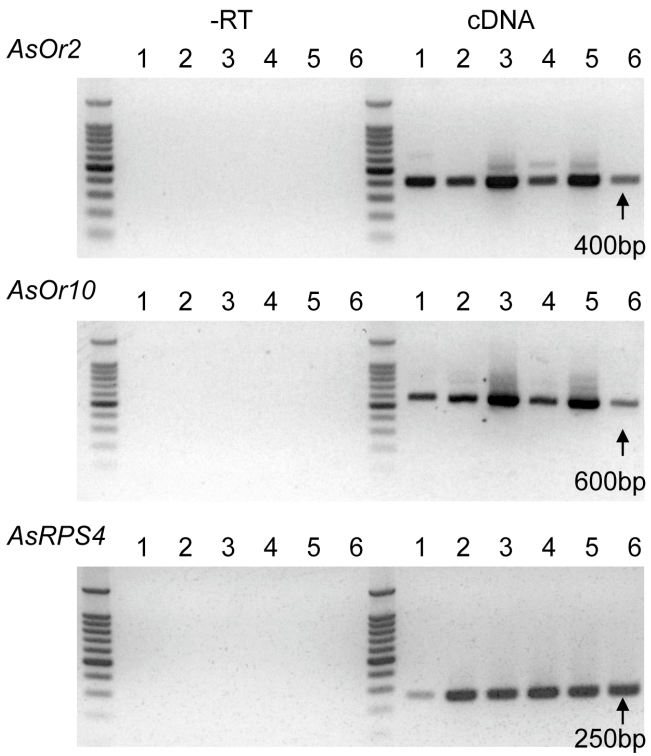

An. stephensi chromosome 3R
AsOr10
AsOr10x
AsOr2
An. gambiae chromosome 3R
AgOr10
AGAP009516
AGAP009517
AGAP009518
AgOr2
AGAP009521
AGAP009523
AaOr9
AaOr2
AaOr10
AAEL006008
Not Annotated
AAEL006004
AAEL006002
Ae. aegypti supercongtig 1.185

Figure 3.5: *AsOr2*, *AsOr10*, and *AsOr10x* mRNA expression in 4-day-old adult *An. stephensi* mosquitoes. RT-PCR reactions were performed using gene specific primer sets for *AsOr2*, *AsOr10*, *AsOr10x*, and *AsRPS4* (40s ribosomal protein s4) (Table 3.2). AsRPS4 is used as a positive control. The templates were either negative control (lacking reverse transcriptase, -RT) or cDNA. Two different sets of negative control and cDNA template were used, one for *AsOr2/Or10* RT-PCR and the other for the *AsOr10x* RT-PCR (hence the presence of two sets of AsRPS4 reactions). RT-PCR reactions were run on 1.5% agarose gels with -RT reactions on the left and cDNA reactions on the right with 100bp DNA markers (M) separating both sets. The tissues used were (1) antenna from 4-day-old sugar fed females, (2) maxillary palp/proboscis from 4-day-old sugar fed females, (3) antenna from 4-day-old females 24 hours post blood feeding, (4) maxillary palp/proboscis from 4-day-old females 24 hours post blood feeding, (5) antenna from 4-day-old sugar fed males, and (6) maxillary palp/proboscis from 4-day-old sugar fed males. See Table 3.2 for RT-PCR product sizes versus genomic DNA contamination product sizes. The 510bp RT-PCR product of *AsOr10x* is generated from AsOr10x template containing one intron but lacking another one.


-RT
cDNA
AsOr2
1 2 3 4 5 6 1 2 3 4 5 6
400bp
AsOr10
1 2 3 4 5 6 1 2 3 4 5 6
600bp
AsRPS4
1 2 3 4 5 6 1 2 3 4 5 6
250bp

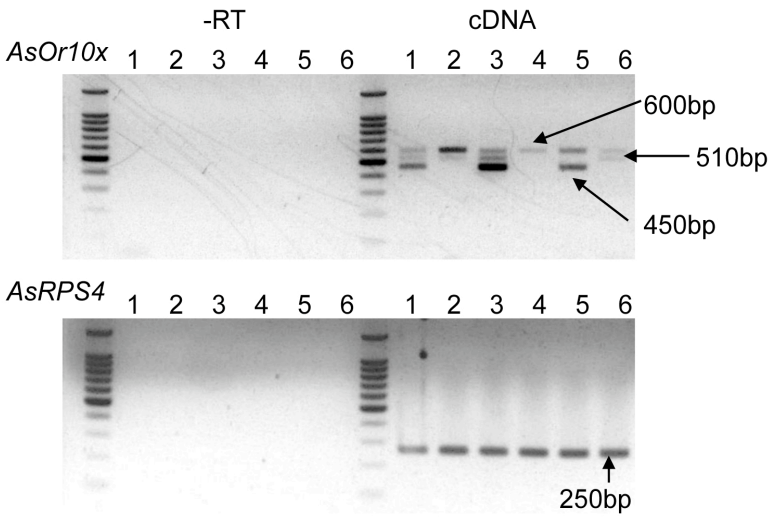
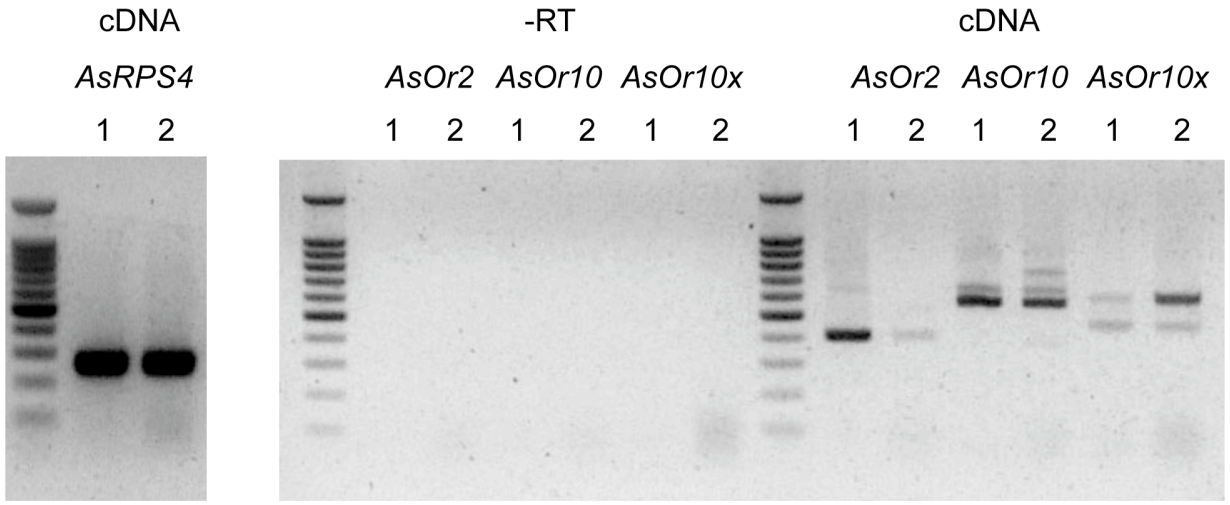

-RT
cDNA
AsOr10x
1 2 3 4 5 6 1 2 3 4 5 6
600bp
510bp
450bp
AsRPS4
1 2 3 4 5 6 1 2 3 4 5 6
250bp

Figure 3.6: Expression profile of *AsOr2*, *AsOr10*, and *AsOr10x* in *An. stephensi* larvae and pupae. RT-PCR reactions were performed on RNA isolated from *An. stephensi* larvae (1) and pupae (2). See legend of Figure 3.5 for details.


cDNA
AsRPS4
1 2
-RT
AsOr2 AsOr10 AsOr10x
1 2 1 2 1 2
cDNA
AsOr2 AsOr10 AsOr10x
1 2 1 2 1 2

3.8: Tables

Table 3.1: Percent identity between *Anopheles Or2*/*Or10* amino acid sequences

		1	2	3	4	5	6	7
AgOr2	1		98.7	94.1	55.8	56	56.5	51.5
AqOr2	2			93.9	55.8	56	56.5	51.5
AsOr2	3				56.3	56.6	57	51.2
AgOr10	4					99.7	93.9	77.2
AqOr10	5						93.9	77.2
AsOr10	6							77.7
AsOr10x	7							

Table 3.2: RT-PCR primers and size of predicted products

Gene	Forward Primer	Reverse Primer	PCR from Genomic DNA (with intron)	PCR from cDNA
AsRPS4	CACGAGGATGGATGTTGGAC	GGTGAATACTTCCGCCTGAT	1368 bp	261 bp
AsOr2	GATCGAAGAGTGCCCGATAA	GGTCACAAAGCACGCACTAA	633 bp	403 bp
AsOr10	CAGCTGGTGCTTACCTTTCC	GTTCAGCAGCGATTGAAACA	755 bp	570 bp
AsOr10x	CATACCGGGCATAAGCATCT	TTCCTCCCGTATCTCATTCG	595 bp	457 bp

Table 3.3: Summary of the expression profile of genes in the Or2/Or10 subfamily

			Adult Female (NBF)			Adult Male		
	Larvae	**Pupae**	**Antenna**	**Palp**	**Proboscis**	**Antenna**	**Palp**	**Proboscis**
AgOr21	unknown	unknown	present	unknown	absent	present	unknown	absent
AgOr10	unknown	unknown	unknown	unknown	unknown	unknown	unknown	unknown
AsOr2[2]	present	present	present	present	present	present	present	present
AsOr10[2]	present	present	present	present	present	present	present	present
AsOr10x[2]	present	present	present	absent	absent	present	absent	absent
AaOr2[3]	present	unknown	present	absent	absent	present	absent	absent
AaOr9[3]	present	unknown	absent	absent	absent	absent	absent	absent
AaOr10[3]	absent	absent	present	absent	absent	present	absent	absent
DmOr43a[4]	absent	present	present	absent	absent	present	absent	absent

1. See reference 24 for data
2. This study
3. See reference 27 for data
4. See references 34, 43, 44

Acknowledgements

Thanks to Dr. Igor Sharakhov for information about the chromosomal location of the *Or2* cluster in *An. stephensi*, and suggestions on changes to the manuscript. Thanks to Dr. Jim Biedler for helpful comments on the manuscript. Thanks to Dr. Senay Sengul for her kind gift of the *An. stephensi RPS4* primers. This work is supported by a NIH grant AI063252 to Z.T.

3.9: Bibliography

1. Krzywinski J, Besansky NJ: **Molecular systematics of *Anopheles*: from subgenera to subpopulations.** *Annu Rev Entomol* 2003, **48**:111-139.
2. White GB: ***Anopheles gambiae* complex and disease transmission in Africa.** *Trans R Soc Trop Med Hyg* 1974, **68**:278-301.
3. Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V: **A polytene chromosome analysis of the *Anopheles gambiae* species complex.** *Science* 2002, **298**:1415-1418.
4. Githeko AK, Adungo NI, Karanja DM, Hawley WA, Vulule JM, Seroney IK, Ofulla AV, Atieli FK, Ondijo SO, Genga IO, Odada PK, Situbi PA, Oloo JA: **Some observations on the biting behavior of *Anopheles gambiae* s.s., *Anopheles* arabiensis, and *Anopheles* funestus and their implications for malaria control.** *Exp Parasitol* 1996, **82**:306-315.
5. Dekker T, Takken W: **Differential responses of mosquito sibling species *Anopheles* arabiensis and *An. quadriannulatus* to carbon dioxide, a man or a calf.** *Med Vet Entomol* 1998, **12**:136-140.
6. Costantini C, Sagnon NF, della Torre A, Diallo M, Brady J, Gibson G, Coluzzi M: **Odor-mediated host preferences of West African mosquitoes, with particular reference to malaria vectors.** *Am J Trop Med Hyg* 1998, **58**:56-63.
7. Dekker T, Takken W, Braks MA: **Innate preference for host-odor blends modulates degree of anthropophagy of *Anopheles gambiae* sensu lato (Diptera: Culicidae).** *J Med Entomol* 2001, **38**:868-871.
8. Pates HV, Takken W, Curtis CF, Huisman PW, Akinpelu O, Gill GS: **Unexpected anthropophagic behaviour in *Anopheles quadriannulatus*.** *Med Vet Entomol* 2001, **15**:293-298.
9. Pates HV, Takken W, Curtis CF: **Laboratory studies on the olfactory behaviour of *Anopheles quadriannulatus*.** *Entomologia Experimentalis et Applicata* 2005, **114**:153-159.
10. Sallum MAM, Schultz TR, Foster PG, Aronstein K, Wirtz RA, Wilkerson RC: **Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences.** *Systematic Entomology* 2002, **27**:361-382.

11. Takken W, Dekker T, Wijnholds YG: **Odor-Mediated Flight Behavior of *Anopheles gambiae* Giles Sensu Stricto and *An. stephensi* Liston in Response to CO_2, Acetone, and 1-Octen-3-ol (Diptera: Culicidae).** *J Insect Behav* 1997, **10**:395-407.
12. Acree FJ, Turner RB, Gouck HK, Beroza M, Smith N: **L-Lactic acid: a mosquito attractant isolated from humans.** *Science* 1968, **161**:1346-1347.
13. Geier M, Bosch OJ, Boeckh J: **Ammonia as an attractive component of host odour for the yellow fever mosquito, *Aedes aegypti*.** *Chem Senses* 1999, **24**:647-653.
14. Meijerink J, Braks MAH, Brack AA, Adam W, Dekker T, Posthumus MA, Beek TAV, Loon JJAV: **Identification of Olfactory Stimulants for *Anopheles gambiae* from Human Sweat Samples.** *J Chem Ecol* 2000, **26**:1367-1382.
15. Zwiebel LJ, Takken W: **Olfactory regulation of mosquito-host interactions.** *Insect Biochem Mol Biol* 2004, **34**:645-652.
16. Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR: **A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*.** *Neuron* 1999, **22**:327-338.
17. Gao Q, Chess A: **Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence.** *Genomics* 1999, **60**:31-39.
18. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R: **A spatial map of olfactory receptor expression in the *Drosophila* antenna.** *Cell* 1999, **96**:725-736.
19. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ: **G protein-coupled receptors in *Anopheles gambiae*.** *Science* 2002, **298**:176-178.
20. Benton R: **On the ORigin of smell: odorant receptors in insects.** *Cell Mol Life Sci* 2006, **63**:1579-1585.
21. Wistrand M, Kall L, Sonnhammer EL: **A general model of G protein-coupled receptor sequences and its application to detect remote homologs.** *Protein Sci* 2006, **15**:509-521.
22. Lundin C, Käll L, Kreher SA, Kapp K, Sonnhammer EL, Carlson JR, Heijne GV, Nilsson I: **Membrane topology of the *Drosophila* OR83b odorant receptor.** *FEBS Lett* 2007, **581**: 5601–5604.
23. Hallem EA, Carlson JR: **Coding of odors by a receptor repertoire.** *Cell* 2006, **125**:143-160.
24. Fox AN, Pitts RJ, Robertson HM, Carlson JR, Zwiebel LJ: **Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding.** *Proc Natl Acad Sci U S A* 2001, **98**:14693-14697.
25. Fox AN, Pitts RJ, Zwiebel LJ: **A Cluster of Candidate Odorant Receptors from the Malaria Vector Mosquito, *Anopheles gambiae*.** *Chem Senses* 2002, **27**:453-459.
26. Melo ACA, Ruťzler M, Pitts RJ, Zwiebel LJ: **Identification of a chemosensory receptor from the yellow fever mosquito, *Aedes aegypti*, that is highly conserved and expressed in olfactory and gustatory organs.** *Chem Senses* 2004, **29**:403-410.
27. Bohbot J, Pitts RJ, Kwon HW, Rutzler M, Robertson HM, Zwiebel LJ: **Molecular characterization of the *Aedes aegypti* odorant receptor gene family.** *Insect Mol Biol* 2007, **16**:525-537.
28. Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, Tanoue S, Yasukochi Y, Touhara K, Nishioka T: **Identification and functional characterization of a sex pheromone receptor in the silkmoth Bombyx mori.** *Proc Natl Acad Sci U S A* 2004, **101**:16653-16658.
29. Nakagawa T, Sakurai T, Nishioka T, Touhara K: **Insect sex-pheromone signals mediated**

by specific combinations of olfactory receptors. *Science* 2005, **307**:1638-1642.
30. Wanner KW, Anderson AR, Trowell SC, Theilmann DA, Robertson HM, Newcomb RD: **Female-biased expression of odourant receptor genes in the adult antennae of the silkworm, Bombyx mori.** *Insect Mol Biol* 2007, **16**:107-119.
31. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family.** *Genome Res* 2006, **16**:1395-1403.
32. Krieger J, Raming K, Dewer YM, Bette S, Conzelmann S, Breer H: **A divergent gene family encoding candidate olfactory receptors of the moth Heliothis virescens.** *Eur J Neurosci* 2002, **16**:619-628.
33. Krieger J, Grosse-Wilde E, Gohl T, Dewer YM, Raming K, Breer H: **Genes encoding candidate pheromone receptors in a moth (Heliothis virescens).** *Proc Natl Acad Sci U S A* 2004, **101**:11845-11850.
34. Hallem EA, Nicole Fox A, Zwiebel LJ, Carlson JR: **Olfaction: mosquito receptor for human-sweat odorant.** *Nature* 2004, **427**:212-213.
35. Wanner KW, Nichols AS, Walden KK, Brockmann A, Luetje CW, Robertson HM: **A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid.** *Proc Natl Acad Sci U S A* 2007, **104**:14383-14388.
36. Stortkuhl KF, Kettler R: **Functional analysis of an olfactory receptor in *Drosophila melanogaster*.** *Proc Natl Acad Sci U S A* 2001, **98**:9381-9385.
37. Wetzel CH, Behrendt HJ, Gisselmann G, Stortkuhl KF, Hovemann B, Hatt H: **Functional expression and characterization of a *Drosophila* odorant receptor in a heterologous cell system.** *Proc Natl Acad Sci U S A* 2001, **98**:9377-9380.
38. Nozawa M, Nei M: **Evolutionary dynamics of olfactory receptor genes in *Drosophila* species.** *Proc Natl Acad Sci U S A* 2007, **104**:7122-7127.
39. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: **VISTA: computational tools for comparative genomics.** *Nucleic Acids Res* 2004, **32**:W273-279.
40. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
41. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome Res* 2004, **14**:988-995.
42. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
43. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
44. Besansky NJ, Fahey GT: **Utility of the white gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae).** *Mol Biol Evol* 1997, **14**:442-454.
45. Fishilevich E, Domingos AI, Asahina K, Naef F, Vosshall LB, Louis M: **Chemotaxis behavior mediated by single larval olfactory neurons in *Drosophila*.** *Curr Biol* 2005, **15**:2086-2096.
46. Stortkuhl KF, Kettler R, Fischer S, Hovemann BT: **An increased receptive field of olfactory receptor Or43a in the antennal lobe of *Drosophila* reduces benzaldehyde-driven avoidance behavior.** *Chem Senses* 2005, **30**:81-87.
47. Helfand SL, Carlson JR: **Isolation and characterization of an olfactory mutant in *Drosophila* with a chemically specific defect.** *Proc Natl Acad Sci U S A* 1989, **86**:2908-2912.

48. Anholt RR, Lyman RF, Mackay TF: **Effects of single P-element insertions on olfactory behavior in *Drosophila melanogaster*.** *Genetics* 1996, **143**:293-301.
49. Collins LE, Blackwell A: **Olfactory cues for oviposition behavior in Toxorhynchites moctezuma and Toxorhynchites amboinensis (Diptera: Culicidae).** *J Med Entomol* 2002, **39**:121-126.
50. Blackwell A, Johnson SN: **Electrophysiological investigation of larval water and potential oviposition chemo-attractants for *Anopheles gambiae* s.s.** *Ann Trop Med Parasitol* 2000, **94**:389-398.
51. Ray A, van Naters WG, Shiraiwa T, Carlson JR: **Mechanisms of odor receptor gene choice in *Drosophila*.** *Neuron* 2007, **53**:353-369.
52. Tao Q, Wang A, Zhang HB: **One large-insert plant-transformation-competent BIBAC library and three BAC libraries of Japonica rice for genome research in rice and other grasses.** *Theor Appl Genet* 2002, **105**:1058-1066.
53. Robertson HM, Warr CG, Carlson JR: **Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*.** *Proc Natl Acad Sci U S A* 2003, **100**:14537-14542.
54. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20**:426-427.
55. Page RD: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**:357-358.

Chapter 4. Expression profile and knockdown of *AaOr7*, *AaGr1*, *AaGr2*, and *AaGr3* in *Aedes aegypti*

4.1: Abstract

Insect odorant receptors are believed to be critical components for both odorant recognition and olfactory signaling. The closely related family of gustatory receptors code both taste and olfactory receptors. Recent whole genome sequencing projects have led to the identification of odorant and gustatory receptor genes in several insect species. Most of these genes code orphan proteins with unknown functions. Many recent studies have focused on determining the function of *Drosophila melanogaster* olfactory receptors. The techniques used in these studies are not transferable to other insect species such as mosquitoes.

In this report we have studied the dengue mosquito *Aedes aegypti*'s odorant receptor *Or7* (*AaOr7*) and gustatory receptors *Gr1* and Gr2 (*AaGr1* and *AaGr2*). These receptors are orthologs of receptors of known function in other species. We have found the distribution of these genes in chemosensory tissue using quantitative and non-quantitative RT-PCR. *AaOr7* is widely expressed in antenna, maxillary palp, and proboscis of adults. There is very limited expression in bodies and legs, and surprisingly lower amount of *AaOr7* in male antenna compared to female antennas. The putative CO_2-detecting receptors *AaGr1-3* are expressed almost exclusively in the maxillary palp in addition to expression in *Ae. aegypti* larvae. The presence of these

receptors in larvae suggests they are able to respond to CO_2, something that hasn't been previously reported. To determine the actual function of these receptors in mosquitoes we decided to lower their mRNA levels which in turn will eventually lower protein levels generating loss-of-function mosquitoes to assay. Recombinant double subgenomic Sindbis viruses were able to lowere mRNA levels of these receptors. Recombinant Sindbis viruses are able to target odorant and gustatory receptors and knockdown mRNA levels to 15% of uninfected mosquitoes. This work is ongoing as we improve upon this knockdown technique and develop effective behavior assays to test the potential loss-of-function mosquitoes.

4.2: Introduction

Insect odorant receptors (ORs) are a large and diverse family of seven-transmembrane proteins. These receptors were first discovered in *Drosophila melanogaster* through a combination of bioinformatics and cDNA library screening [1-3]. This family has also been identified in several non-*Drosophila* insect species including *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera*, *Bombyx mori*, and *Tribolium castaneum* [4-8]. Functional studies have shown that ORs detect several biologically significant odorants. One *Ap. mellifera* receptor responds to a major component of queen retinue pheromone [9]. The AgOr1 receptor in *An. gambiae* responds to a component of human sweat [10]. ORs are shown to be necessary for

detection of many other specific odorants [11, 12]. There is an ongoing effort to identify ORs that bind other biologically important odorants through functional studies.

Three types of approaches have been used to determine the function of *D. melanogaster* ORs. One approach is to disrupt an OR gene and assay the loss-of-function phenotype. A recent study explored the function of DmOr83b by using a combination of gene targeting and behavior assays [13]. Another group misexpressed the *DmOr43a* gene to identify the function of the DmOr43a protein [14]. The second approach is to express receptors in heterologous expression systems (using *Xenopus* oocytes, HEK293 cells, or Sf9 cells) and perform either two-electrode voltage-clamp electrophysiology or calcium imaging [11, 15, 16]. These heterologously expressed receptors are able to generate observable changes in voltage potential in the presence of a recognized ligand. Recent reports have indicated that this is likely due to the ability of ORs to form ion channels and utilize cell signaling components [17, 18]. Most *D. melanogaster* OR function studies have employed a different strategy of expressing a transgenic receptor gene in a neuron with no endogenously expressed receptor (the "empty neuron" system) and then performing electroantennograms [19]. Primarily through this approach almost every *D. melanogaster* OR has an assigned putative function [20-23]. The gene disruption and knockout techniques available in *D. melanogaster* are unfortunately not always present in other insect species. Although this system has been used to deorphanize ORs from non-*Drosophila* species this was

only accomplished by inserting receptors into *D. melanogaster* neurons [10]. This is a less than ideal approach as ORs are expressed in an environment lacking other native (non-*Drosophila*) olfactory proteins such as odorant binding proteins, odorant degrading enzymes, and sensory neuron membrane proteins. Receptor specificity for a ligand might be affected without these proteins or in the presence of *D. melanogaster* olfactory proteins.

A second family of chemoreceptor proteins, the taste or gustatory receptor (GR) family, has been identified in insects. Both OR and GR families encode seven-transmembrane proteins, are highly divergent, have few obvious orthologs conserved between different insect lineages, and display similarity to other protein families besides each other. GRs are expressed in insect gustatory tissues including the proboscis, legs, and wings [24-26]. Several of these receptors are also expressed in olfactory tissue [27]. Given their expression profile, GRs have been hypothesized to be predominately taste receptors, and several functional studies have confirmed this [28-30]. The most well studied receptors are members of a small subfamily that are found conserved throughout many different species (see supplementary material in [31]). *DmGr21a* and *DmGr63a* are the *D. melanogaster* GRs of this subfamily. Both receptors are co-expressed in ab1C antennal neurons and these neurons synapse to the V glomeruli in the antennal lobe [27, 32, 33]. These studies have found that both GRs are required for detection of CO_2 in fruit flies, making them functionally olfactory

receptors although not phylogenetically ORs. This family has also been identified in the mosquito *An. gambiae* [4], and a preliminary report has found that these GRs confer CO_2 detection [31]. Therefore both the insect OR and GR families encode at least some olfactory receptors, which detect biologically important odorants. Given their similarity a method to test the function of receptors of one family should work with the other family as it pertains to detection of odorants.

Several different mosquito species serve as disease vectors transmitting causative agents that result in malaria, dengue, West Nile encephalitis, and yellow fever. Not all mosquito species are disease vectors in large part due to their host-seeking behavior. Hosts are detected by mosquitoes in response to host olfactory cues [34]. There is a great deal of interest in identifying how olfaction in mosquitoes works, and more specifically which mosquito ORs/GRs are involved in detection of host-odors. We decided to study the function of mosquito ORs and GRs by knocking down the mRNA levels of select receptors, and using assays to measure behavioral differences. It is well established that mosquitoes utilize an RNAi pathway, and that this pathway can be exploited to lower genes of interest mRNA levels. In *Ae. aegypti* the double subgenomic Sindbis virus (SINV) has been used to knockdown a number of non-olfactory endogenous and exogenous genes [35-37]. We have developed several recombinant SINVs targeting the *AaOr7*, *AaGr1*, and *AaGr2* genes alone and in different combinations. These genes are orthologs of olfactory receptors with known

function in other insect species. *AaOr7* is an ortholog of *DmOr83b*, which is required for many olfactory driven behaviors in *D. melanogaster* [13]. *AaGr1* and *AaGr2* are *Ae. aegypti* members of the *DmGr21a* and *DmGr63a subfamily*. By knocking down ORs/ GRs with known or possible functions already assigned, we can test the validity of our approach.

4.3: Results

4.3.1: Expression Profile of AaOr7, AaGr1, AaGr2, and AaGr3

We first wanted to determine the expression profile of the genes of interest to determine at which stages of life and which chemosensory tissues they are expressed. A previous study has already identified the members of the CO_2 detecting GR subfamily in *Ae. aegypti*: *AaGr1*, *AaGr2*, and *AaGr3* [38]. We designed gene specific primers to amplify regions from these genes that cross at least one predicted intron to distinguish products primed from cDNA with those primed from genomic DNA. An additional set of primers was designed to amplify cDNA products of the *AaOr7* gene. Although this gene already has a described expression profile, we include it in this analysis for consistency and to serve as a positive control [39]. To determine effectiveness of cDNA synthesis we used the *AaRPS7* (ribosomal protein S7) gene as it is a housekeeping gene, which should have unchanged mRNA levels at each stage of development.

We isolated total RNA from *Ae. aegypti* mosquitoes at several life stages and from several different tissues. From young mosquitoes we obtained total RNA from the whole bodies of 1st, 2nd, 3rd, and 4th instar larvae and from pupae. Five to six days post-emergence adult male and female mosquitoes were dissected, and then total RNA was extracted from antenna, maxillary palp, proboscis, and legs. Both non-bloodfed (NBF) and 24 hours post-bloodfed (BF) female mosquitoes were used to determine if any of the receptors were knocked down after blood feeding. Total RNA was used to synthesize cDNA and this was used in several non-quantitative RT-PCR reactions along with gene specific primers (Figure 4.1A and Figure 4.2A). We found that *AaOr7* is robustly expressed in both young and adult mosquitoes. In adults the three major olfactory tissues of mosquitoes (antenna, maxillary palp, and proboscis) showed *AaOr7* expression (Figure 4.2A), which is consistent with previous data [39]. Of significant note we saw no expression of *AaOr7* in BF females or males legs, but only a very low level in NBF female legs. The expression of *AaGr1-3* was much more restricted than *AaOr7*. All three GRs were found expressed in larvae and pupae, but at low levels suggesting transitory expression or stable expression in a very small number of larvae neurons (Figure 4.1A). In adults *AaGr1-3* expression is localized almost exclusively to the maxillary palp with significantly lower levels in the antenna and proboscis (Figure 4.2A).

To determine more precisely differences in mRNA levels between tissues we decided to perform real-time PCR. Custom-made Taqman assays for *AaRPS7*, *AaOr7*, *AaGr1*, *AaGr2*, and *AaGr3* were developed by Applied Biosystems. These assays were used in conjunction with our cDNA pools to determine relative mRNA levels of a given gene in several different tissues (Figure 4.1B, Figure 4.2B, and Table 4.1-Table 4.8). All genes were compared with their relative gene levels in 1st instar larvae therefore the mRNA level of each gene in 1st instar larvae was set to one. In larvae we found that all four receptor genes mRNA levels were relatively unchanged from 1st to 4th instar. This was not true of pupae where we saw an increase of mRNA levels in all four genes. The most dramatic example is *AaGr1* which displayed 40.6 fold increase in mRNA levels compared to 1st instar larvae (Figure 4.1B and Table 4.1). *AaOr7* is expressed at much higher levels in female antenna than in either palp or proboscis, which is consistent with its role as a primary olfactory tissue (Table 4.8). Male antennas display much lower levels of *AaOr7* perhaps reflecting a different distribution or number of neurons in female *Ae. aegypti*. Additionally our analysis of leg tissues has shown that *AaOr7* is expressed in NBF, BF, and male legs, but at <3% of NBF proboscis *AaOr7* mRNA levels (data not shown). *AaGr1-3* was indeed expressed in maxillary palp at 36.10 to 113.2 times the levels in NBF Proboscis (Table 4.2, Table 4.4, and Table 4.6). Surprisingly we found that only *AaGr1* shows any reduction of mRNA levels in BF females while *AaGr2* and *AaGr3* stay the same. It is possible *AaGr1* mRNA knockdown

eventually destabilizes the CO_2 detecting complex leading to the reported lack of interest in CO_2 post-bloodfeding. More experiments must be conducted before this can be conclusively proven. As far as could be determined with both non-quantitative and quantitative RT-PCR *AaGr1-3* expression profiles are very similar supporting previous data that receptors of this family colocalize in the same neurons [31-33].

4.3.2: Mosquito chemosensory tissues are infected by recombinant SINVs

Previous data has shown that SINVs are able to infect a large range of tissues in different insect species [40]. We wanted to determine if the SINV could infect chemosensory tissues, and possibly olfactory sensory neurons. We obtained a plasmid encoding a SINV that expresses green fluorescent protein (GFP), and then the virus (GFP-SINV) was raised as above. GFP-SINV was injected into the abdomen of 0-24 hour post-emergence *Ae. aegypti* female mosquitoes. Four days post infection, uninfected and infected mosquitoes were examined for green fluorescence (Figure 4.3). Comparing our uninfected control with infected mosquitoes we observed an obvious difference with infected mosquitoes showing green fluorescence in a wide range of tissues including abdomen, thorax, eyes, etc. Of importance for this study we saw florescence between the antennal segments (Figure 4.3, pink arrows), and the tips of the maxillary palp (Figure 4.3, blue arrows), and the proboscis (Figure 4.3, orange arrows). Fluorescence is possibly localized throughout the entirety of these tissues,

but the mosquito cuticle may block the fluorescence. In another study (Sengul and Tu, unpublished it was shown that another family of olfactory proteins (odorant binding proteins) could be knocked down by SINVs. These data suggests that mosquito olfactory sensory neurons can be infected by SINVs.

4.3.3: Developing Recombinant SINVs

Previous studies have shown that by inserting gene fragments into SINVs and then infecting mosquitoes with these recombinant viruses results in targeted gene knockdown [35-37]. We decided to focus on knocking down *AaOr7*, *AaGr1*, and *AaGr2* by developing a series of recombinant SINVs. The first step in producing the viruses was to PCR amplify regions of these three genes from genomic DNA or non-bloodfed female antenna cDNA (Table 4.9). We produced four inserts for *AaOr7* in different regions of the *AaOr7* gene to insure sufficient knockdown (Figure 4.4A). As an *Or83b* ortholog, knocking down *AaOr7* should lead to mosquitoes defficient in most normal olfactory cues [13]. The gene fragments were then cloned into the TE3'2Jmcs vector, a SINV-encoding plasmid. Figure 4.4B lists all the different plasmids constructed, and the different combination of receptor gene fragments cloned into each plasmid. Plasmids and the recombinant viruses they produce were named according to the gene fragments they contain. Plasmids were used as template in an *in vitro* transcription reaction to produce viral RNA, and it was then electroporated into

BHK21 cells. These cells were allowed to produce the recombinant viruses for 48 hours, and then the viruses were harvested by centrifugation.

4.3.4: SINV mediated knockdown of *AaOr7* and *AaGr1*

The AaOr7 SINV, AaGr1 SINV, AaGr1+AaOr7 SINV, and AaGr1+AaGr2 SINV recombinant SINVs (Figure 4.4B) were injected into the abdomen of 0-24 hour post-emergence *Ae. aegypti* female mosquitoes. As a negative control we injected mosquitoes with the TE3'2JE2 virus, which contains an E2 insert but no receptor gene inserts. Mosquitoes were feed on sugar water for 19 days post infection to ensure sufficient time for viral proliferation. Uninfected, TE3'2JE2 infected, and recombinant SINV infected mosquitoes were surgically dissected removing antenna, maxillary palp, and proboscis. Total RNA was extracted from these tissues and converted into cDNA for mRNA expression analysis.

Real-time PCR assays were obtained from Applied Biosystems for the *AaRPS7* (ribosomal protein S7) gene as an endogenous control along with the *AaGr1*, *AaGr2*, and *AaOr7* genes. We used these assays along with our cDNA template to determine relative quantities of our genes of interest in each tissue, and determine if infected mosquitoes showed lower mRNA levels of these genes (Figure 4.5 and Table 4.10). Each graph on Figure 4.5 represents a separate reaction run with separate negative controls and calibrators. Each graph has at least one negative control which is

uninfected and/or TE3'2JE2 infected tissue. Calibrators for *AaOr7* assays are negative control antenna tissue, while calibrators for *AaGr1* and *AaGr2* assays are negative control maxillary palp. These tissues were chosen as calibrators because they have the highest relative levels of the genes assayed (Figure 4.2). All viruses except the AaOr7 SINV (data not shown) produced some level of gene knockdown. The most significant knockdown was that of *AaGr2* in the maxillary palp of AaGr1+AaGr2 SINV infected mosquitoes. *AaGr2* mRNA levels in AaGr1+AaGr2 SINV infected maxillary palp were only 15.04% of *AaGr2* mRNA levels in uninfected maxillary palp. Less significant knockdowns of *AaGr1* were found in AaGr1 SINV infected maxillary palp and in AaGr1+AaOr7 SINV infected maxillary palp. *AaOr7* mRNA levels were also depressed in AaGr1+AaOr7 SINV infected antenna. From these data we've determined that OR and GR mRNA levels can be depressed to as low as 15.0% of normal using recombinant SINVs.

4.3.5: Behavior Assay

Having shown that ORs and GRs mRNA levels can be lowered using recombinant SINVs we then develop a behavioral assay setup to test for behavioral defects in infected mosquitoes. The assay setup we choose is illustrated on Figure 4.6A. This setup is a dual-choice olfactometer allowing mosquitoes to associate in a non-random manner with one side of the apparatus or the other in response to test

odors. For example if an attractive odor is used on one side then mosquitoes will associate with that side of the container and attempt to touch/interact with the odor source. The opposite would be true of a repulsive odor. Currently we are working to further refine this assay with known attractive chemicals and non-bloodfed *Ae. aegypti* female mosquitoes. Figure 4.6B illustrates that *Ae. aegypti* females are attracted to palm sweat and CO_2 versus palm sweat alone.

4.4: Discussion

The study of insect ORs has greatly expanded in the recent years. Currently most studies focus on how ORs interact with other proteins to generate a signaling pathway and also which ORs respond to biologically important odorants. In this current report we have accomplished several things. The first was to determine the expression profile of the *Ae. aegypti* CO_2 detecting subfamily members *AaGr1*, *AaGr2*, and *AaGr3*. Using non-quantitative RT-PCR we were able to determine that these three receptors have nearly identical expression profiles in keeping with previous data from *An. gambiae* [31]. In adults all three receptors are expressed almost exclusively in the maxillary palp the CO_2-sensing organ of mosquitoes. Surprisingly these receptors are expressed in larvae and pupae suggesting young mosquitoes may respond to CO_2. As far as we know this has not been observed in any mosquito species. A recent study has shown that *D. melanogaster* larvae do respond to CO_2, and that the *Drosophila*

orthologs of *AaGr1-3* are present [41]. Behavior assays of *Ae. aegypti* larvae should determine if CO_2 elicits a response.

One major purpose of this study was to develop a method to determine the function of mosquito chemosensory receptors. We have used recombinant SINVs containing fragments of receptor genes of known or suspected function (*AaOr7*, *AaGr1*, and *AaGr2*) to lower mRNA levels of these genes which may generate loss-of-function behavioral phenotypes. *Ae. aegypti* female mosquitoes infected with these viruses display lower receptor mRNA levels than those not infected and/or those infected with non-recombinant SINVs. The most significant knockdown resulted in *AaGr2* mRNA levels being only 15.0% of uninfected levels. In order to obtain better knockdowns we have developed several more *AaOr7* recombinant SINVs containing inserts of *AaOr7* in different regions of the gene and at different sizes. Currently these viruses have been used to infect another group of 0-24 hour post-emergence mosquitoes and results are pending. We hope to identify from these series of viruses features that lead to more effective knockdowns of ORs/GRs with recombinant SINVs. The final thrust of this study is to establish a behavior setup to assay our potential loss-of-function behavioral knockdown *Ae. aegypti*. Given the infectious nature of the SINV the main requirement for this assay is that it is easy to assemble/disassemble for cleaning. Currently we have a very preliminary setup that works with non-bloodfed *Ae. aegypti* females.

4.5: Methods

Non-quantitative RT-PCR. Total RNA was extracted from *Ae. aegypti* (Liverpool strain) mosquitoes at various life stages. The whole bodies of 1st, 2nd, 3rd, and 4th instar larvae along with pupae were homogenized in Trizol (Invitrogen; Carlsbad, CA), and total RNA extracted according to the included protocol. Live 5-6 day old *Ae. aegypti* adults were dissected on ice with the antenna, maxillary palp, proboscis, and legs being removed and placed immediately into Trizol. Adult tissues were homogenized and total RNA extracted according to the included protocol including the optional steps for handling small volumes of tissues. Total RNA was then treated to remove genomic DNA contamination using the Turbo DNA Free kit (Ambion; Austin, TX), and total RNA concentrations measured by spectrometry. cDNA was synthesized from 500ng of adult total RNA or 2ug of larvae or pupae total RNA using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). Non-quantitative RT-PCR reagents were composed of 1uL of cDNA, Taq polymerase and reagents (Takara; Otsu, Japan), and gene specific primers for AaRPS7 (JB29 5' ATGGTTTTCGGATCAAAGG 3', JB31 5' GGAATTCGAACGTAACGTCAC 3'), AaOr7 (AaOr7F 5' GGTCTATGCTCTGGCTCAGG 3', AaOr7R 5' AACTGCACCAACACCATGAA 3') AaGr1 (AaGr1F 5' TCTACTGCATGGGTCTGCTG 3', AaGr1R 5' CAGAGCTGGGATGAGAGCTT 3'), AaGr2 (AaGr2F 5' TGGACACGGTCTACTGTTCG 3', AaGr2R 5'

AACCAGGTAGATCGCAATGG 3'), and AaGr3 (AaGr3F 5' GGAGAGTGTTCGGTTTGGAA 3', AaGr3R 5' CAGGCCGTAGATCGAAAGAG 3'). Reactions for AaRPS7 were amplified for 25 cycles, while all other reactions were amplified for 30 cycles. Reactions were analyzed on 1.5% agarose gels.

Developing Recombinant SINVs. Inserts for SINVs were PCR amplified from *Ae. aegypti* genomic DNA or non-bloodfed female antenna cDNA using custom primers designed according to *AaOr7*, *AaGr1*, and *AaGr2* gene sequences. XbaI/PacI restriction sites were added to the 5' ends of primers to facilitate directional cloning. Directional cloning of gene fragments into the TE3'2Jmcs plasmid resulted in gene fragments being in the anti-sense orrientation. Subsequent recombinant plasmids were linearized by cleaving with XhoI. An *in vitro* transcription reaction was performed using the linearized plasmid as template along with components of the T7/SP6 Riboprobe System (Promega; Fitchburg, WI). Viral RNA was stored at -80C to minimize RNA degridation. BHK21 cells were grown in T-150 flasks with 20mL of 1x Dulbecco/Vogt modified Eagle's minimal essential medium (DMEM) (Mediatech; Manassas, VA) supplimented with 10% FBS and 1% penicillin/streptomycin at 37°C untill cells reached a 100% confluency. Cells were washed twice with 10mL of 1x PBS (Mediatech) and then removed from flask surface by incubation with 5mL of 0.25% Trypsin-EDTA (Gibco/Invitrogen) for no more than three minutes followed by firm tapping. 15mL of 1x DMEM supplimented with 10% FBS and 1% penicillin/

streptomycin was added to cells and medium and pipeted till throughly mixed. Cells were transfered to 50mL tubes and then centrifuged at 1000RPM at 4°C for five minutes to pellet cells. Cell pellet was washed twice using 1x PBS (Mediatech) followed by centrifugation at 000RPM at 4°C for five minutes. Cell count was estimated using a hemocytometer and pellet was resuspended in 1x PBS to give a final concentration of $1*10^7$ cells/mL. 400µL of cells were needed for each virus plus two more 400µL aliquotes for negative controls. Each aliquiote of cells was pipeted into electroporation vials, and then 12µL of viral RNA were added to all but two vials. Vials were pulse electroporated using the following conditions: 460 V, 725 Ω, 0075 µF except for one vial lacking virus. Electroporated cells were sowed in T-25 flasks with 5mL of 1x Dulbecco/Vogt modified Eagle's minimal essential medium (DMEM) (Mediatech; Manassas, VA) with 10% FBS and 1% penicillin/streptomycin. After 48 hours the medium was removed from the flasks and centrifuged at 1000 RPM for 6 minutes at 4°C to remove cell debris. Virus-containing media were placed into 1mL vials and stored at -80°C till needed. The concentration of the viruses was determined by a titer test. BHK21 cells were sowed into 96 well plates with DMEM with 10% FBS and 1% penicillin/streptomycin, and cells were allowed to grow to 100% confluent. Media in each well was then replaced with 270µL 1x DMEM containing 2% FBS. A series of 1:10 dilutions of the virus was made for each row of wells starting with addition of 30µL of virus-containing media. The dilutions ranged from 10^{-1} to 10^{-8}. 96

well plates were incubated at 37°C for three to four days after infection. Cells in each well were examined for signs of cytopathic effect (CPE). Titer was estimated based on how dilute a virus could be, while still causing cells to display CPE. Only viruses showing CPE at 10^{-6} to 10^{-8} dilution were used to inject *Ae. aegypti* mosquitoes.

Injection of Mosquitoes with Recombinant SINVs. 0-24 hour post-emergence *Ae. aegypti* non-bloodfed females were anesthetized by cold. Mosquitoes were microinjected with 0.5µL of virus each by injection into the abdomen.

Real-time PCR. Real-time PCR assays were designed and produced by Applied Biosystems (Foster City, CA) for the *AaRPS7*, *AaOr7*, *AaGr1*, and *AaGr2* genes. Reactions were composed of 1µL of 1:3 diluted cDNA template, Taqman Universal PCR Master Mix (Applied Biosystems), and assay mix. Reactions were done in triplicate to serve as technical replicantes. A 7300 Real-Time PCR System machine (Applied Biosystems) was used to cycle reaction temperatures and collect data. Data was analyzed on the included software. Uninfected or TE3'2JE2 infected antenna, or maxillary palp was used as a calibrator. Relative quantities of all genes was determined using the $2^{-\Delta\Delta Ct}$ method [42].

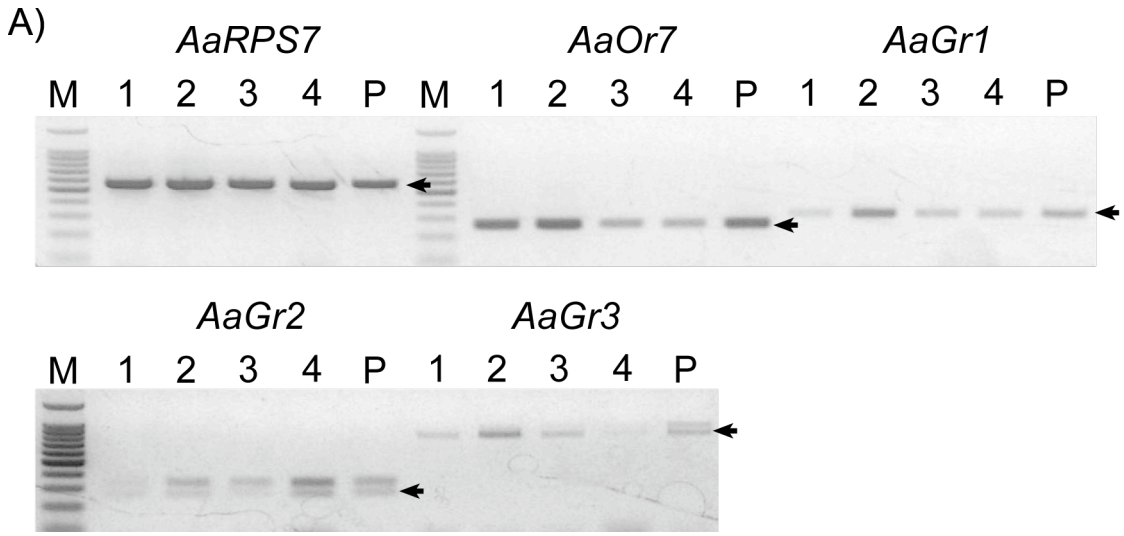
Behavior Assays. Behavior assays were conducted in the apparatus illustrated in Figure 4.6. In brief the assay apparatus consisted of a plastic container with mesh netting between container bottom and lid. Female non-bloodfed *Ae. aegypti* mosquitoes were anesthetized using cold, and placed into the assay apparatus.

Mosquitoes were then left at room temperature for 30 minutes to allow mosquitoes to recover from cold treatment. After cold treatment plastic spacers were placed above two equidistant circular openings in the lid. Odors or water were spotted onto cotton pads, one cotton pad was placed in the spacer, and then heated water bottles were placed on top of the pad. Heated water bottles were connected to a water bath kept at 42°C. Assays were conducted for five minutes allowing mosquitoes to associate with either side of the apparatus. Mosquitoes were considered attracted to a particular odor when they landed below a spacer and showed probing behavior. After counting the total number of mosquitoes on both holes we divided each number by the total number of mosquitoes tested to give the proportion of mosquitoes responding to each side.

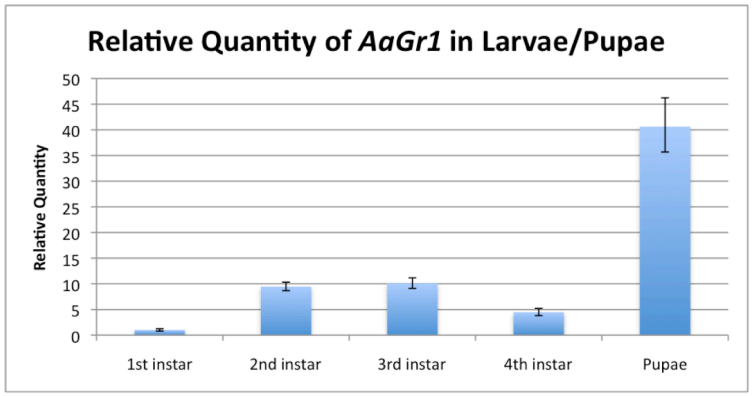
4.6: Figures

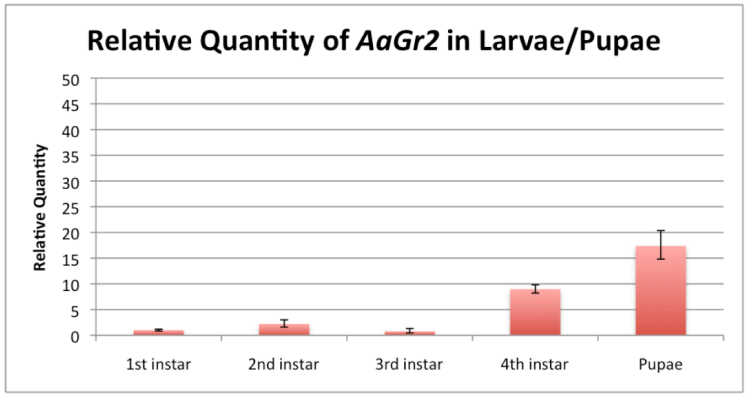
Figure 4.1: *AaOr7* and *AaGr1-3* non-quantitative and quantitative RT-PCR results in young mosquitoes. A) Gene specific primers were used in combination with cDNA template derived from 1st (1), 2nd (2), 3rd (3) , and 4th (4) instar larvae and pupae (P) to generate RT-PCR products. Expected sizes of products lacking introns are 562bp for *AaRPS7*, 261bp for AaOr7, 317bp for *AaGr1*, 260bp for *AaGr2*, and 793bp for *AaGr3*. Arrows point to the cDNA products. B) Real-time PCR results for larvae and pupae. Taqman assays were used along with cDNA to determine relative quantities of a gene

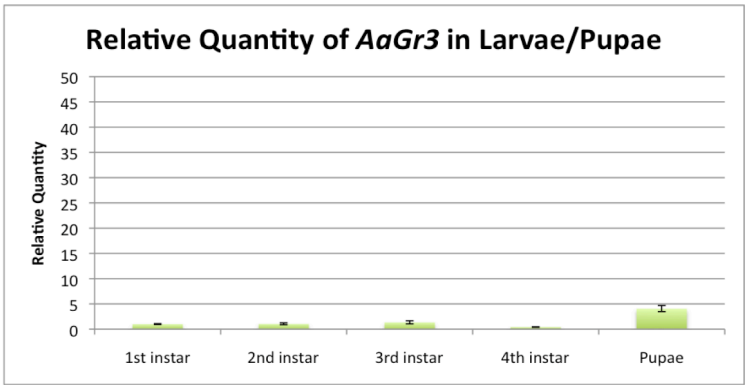
in one tissue is compared to another. 1st instar larvae were used as the calibrator in all graphs.


A)
AaRPS7
AaOr7
AaGr1
M 1 2 3 4 P M 1 2 3 4 P 1 2 3 4 P
AaGr2
AaGr3
M 1 2 3 4 P 1 2 3 4 P

B)


Relative Quantity of AaGr1 in Larvae/Pupae
Relative Quantity
50
45
40
35
30
25
20
15
10
5
0
1st instar
2nd instar
3rd instar
4th instar
Pupae


Relative Quantity of AaGr2 in Larvae/Pupae
Relative Quantity
50
45
40
35
30
25
20
15
10
5
0
1st instar
2nd instar
3rd instar
4th instar
Pupae


Relative Quantity of AaGr3 in Larvae/Pupae
Relative Quantity
50
45
40
35
30
25
20
15
10
5
0
1st instar
2nd instar
3rd instar
4th instar
Pupae

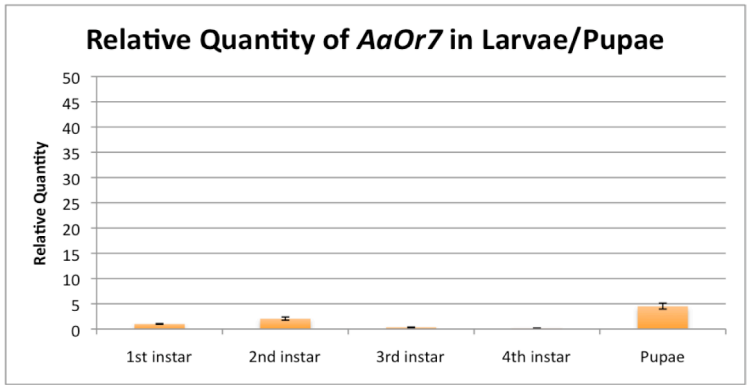
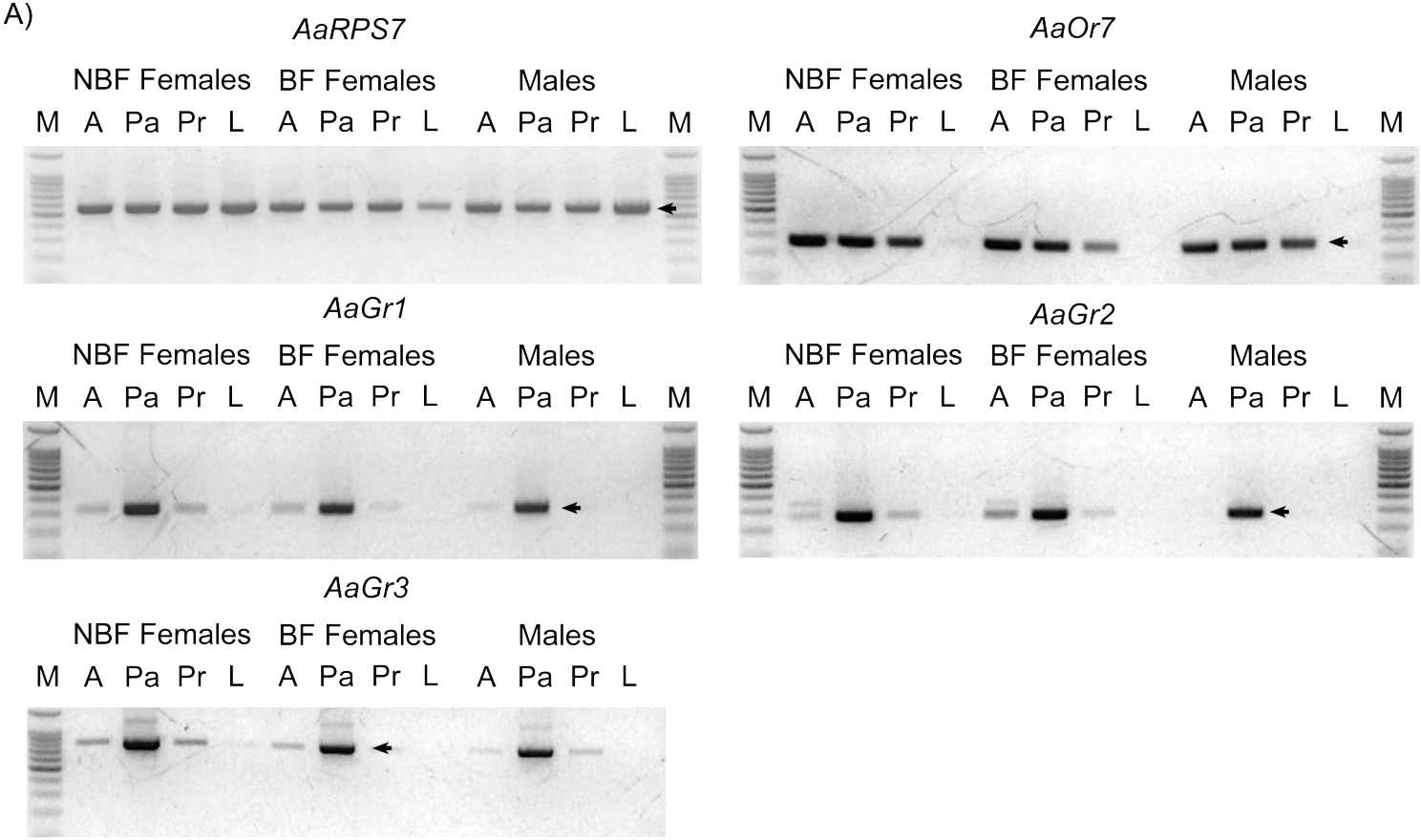
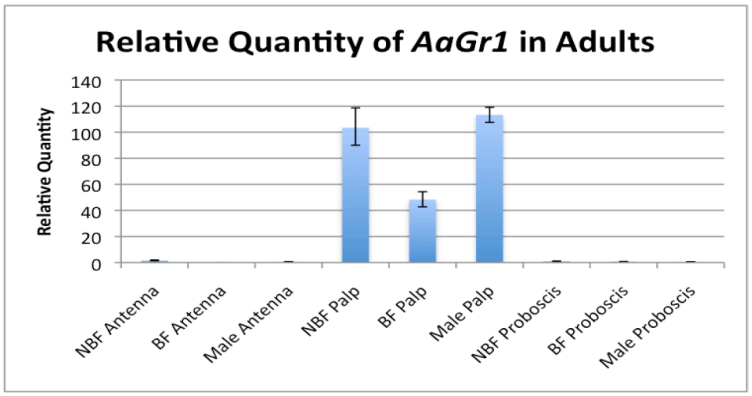

Relative Quantity of AaOr7 in Larvae/Pupae
Relative Quantity
50
45
40
35
30
25
20
15
10
5
0
1st instar
2nd instar
3rd instar
4th instar
Pupae

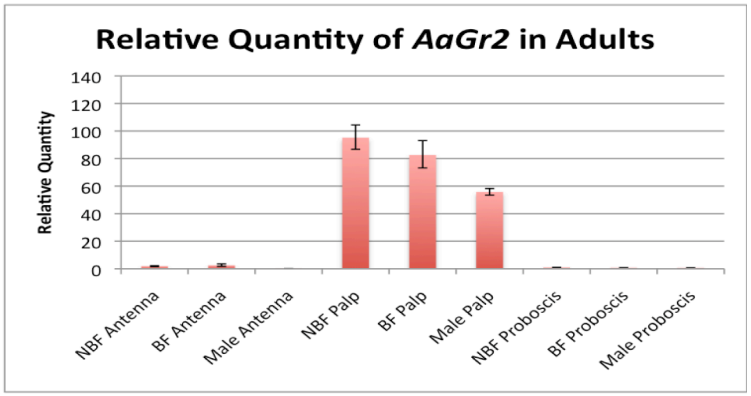
Figure 4.2: *AaOr7* and *AaGr1-3* non-quantitative and quantitative RT-PCR results in adult chemosensory tissues. A) *Ae. aegypti* 5-6 day old non-bloodfed (NBF) female, 24 hour post-bloodfed female (BF), and male chemosensory tissues were dissected. The tissues were antenna (A), maxillary palp (Pa), proboscis (Pr), and legs (L). Total RNA extracted from these tissues was used to generate cDNA. See Figure 4.1 for expected product sizes. B) Real-time PCR results for adult chemosensory tissues. NBF Proboscis was used as the calibrator sample in all graphs.

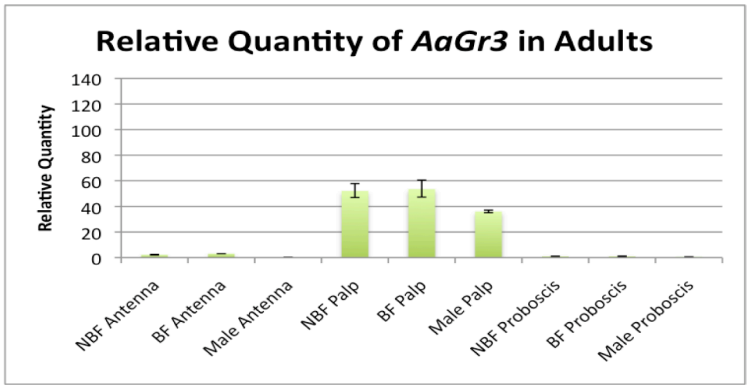
A)


AaRPS7
NBF Females
BF Females
Males
M A Pa Pr L A Pa Pr L A Pa Pr L M
AaOr7
NBF Females
BF Females
Males
M A Pa Pr L A Pa Pr L A Pa Pr L M
AaGr1
NBF Females
BF Females
Males
M A Pa Pr L A Pa Pr L A Pa Pr L M
AaGr2
NBF Females
BF Females
Males
M A Pa Pr L A Pa Pr L A Pa Pr L M
AaGr3
NBF Females
BF Females
Males
M A Pa Pr L A Pa Pr L A Pa Pr L

B)


Relative Quantity of AaGr1 in Adults
Relative Quantity
140
120
100
80
60
40
20
0
NBF Antenna
BF Antenna
Male Antenna
NBF Palp
BF Palp
Male Palp
NBF Proboscis
BF Proboscis
Male Proboscis


Relative Quantity of AaGr2 in Adults
Relative Quantity
140
120
100
80
60
40
20
0
NBF Antenna
BF Antenna
Male Antenna
NBF Palp
BF Palp
Male Palp
NBF Proboscis
BF Proboscis
Male Proboscis


Relative Quantity of AaGr3 in Adults
Relative Quantity
140
120
100
80
60
40
20
0
NBF Antenna
BF Antenna
Male Antenna
NBF Palp
BF Palp
Male Palp
NBF Proboscis
BF Proboscis
Male Proboscis

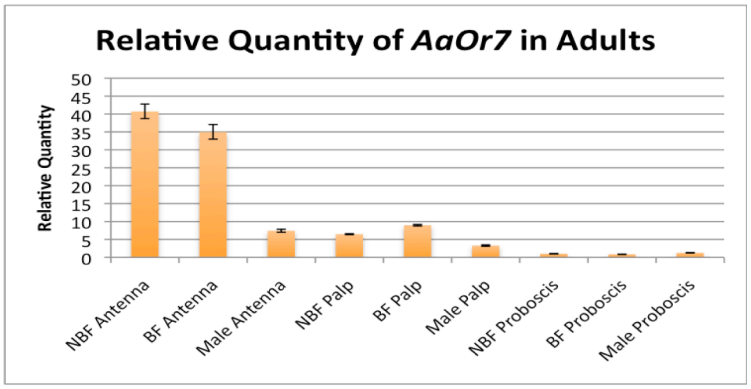

Relative Quantity of AaOr7 in Adults
Relative Quantity
50
45
40
35
30
25
20
15
10
5
0
NBF Antenna
BF Antenna
Male Antenna
NBF Palp
BF Palp
Male Palp
NBF Proboscis
BF Proboscis
Male Proboscis

Figure 4.3: Uninfected and GFP-SINV infected *Ae. aegypti* females. Uninfected mosquitoes show little auto-fluorescence, while those infected with GFP-SINV Sindbis virus show fluorescence in many tissues. GFP-SINV infected mosquitoes display flourescence between antennal segments (pink arrows), on the maxillary palp tips (blue arrows), and at the tip of the proboscis (orange arrows). This suggests SINV infection can spread from the abdomen to olfactory tissues and potentially infect olfactory sensory neurons.

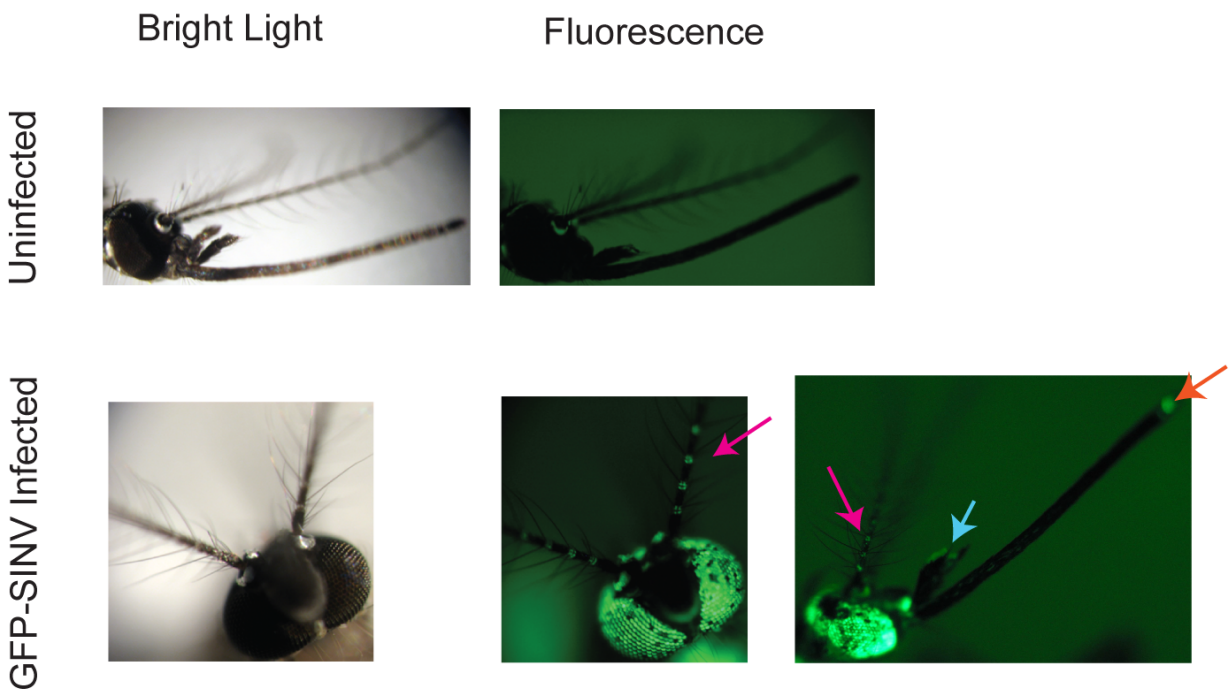

Bright Light
Fluorescence
Uninfected
GFP-SINV Infected

Figure 4.4: Development of Recombinant Sindbis Viruses. A) A diagram of the *AaOr7* gene where boxes represent exons and wedges represent introns. Dashed wedges indicate intronic regions not drawn to scale. Colored lines below the gene indicate regions amplified by gene specific primers and the names of these products. Intronic regions were not amplified as cDNA was used as template. B) Diagrams of recombinant SINV encoding plasmids. Lines represent vector backbone, while boxes represent receptor gene fragment inserts. Names of the recombinant SINVs are to the left of the diagrams.

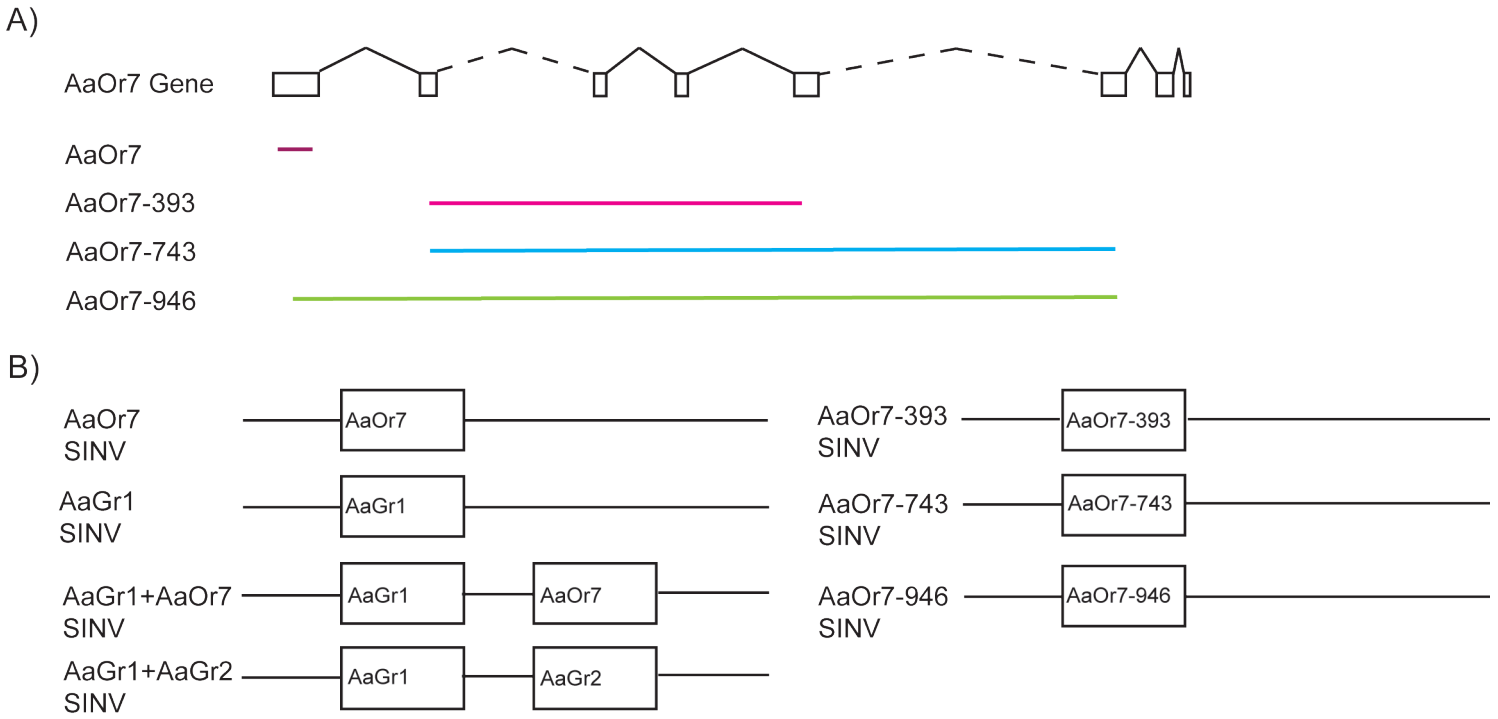
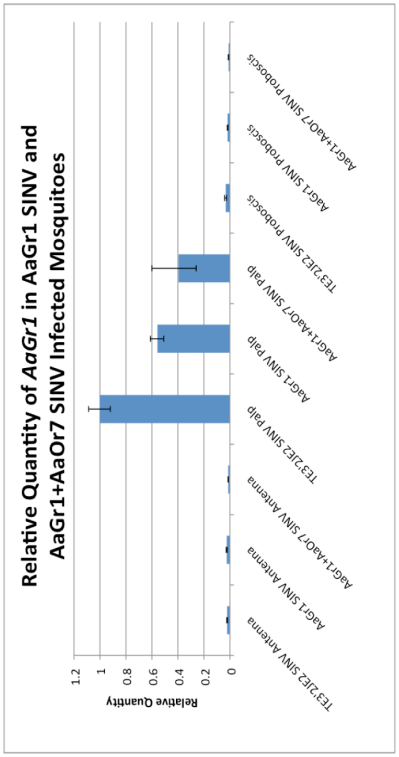
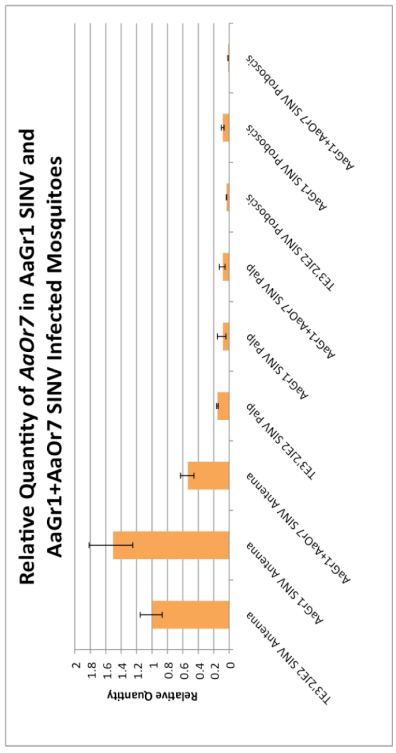

A)
AaOr7 Gene
AaOr7
AaOr7-393
AaOr7-743
AaOr7-946
B)
AaOr7 SINV
AaOr7
AaGr1 SINV
AaGr1
AaGr1+AaOr7 SINV
AaGr1
AaOr7
AaGr1+AaGr2 SINV
AaGr1
AaGr2
AaOr7-393 SINV
AaOr7-393
AaOr7-743 SINV
AaOr7-743
AaOr7-946 SINV
AaOr7-946

Figure 4.5: Real-time PCR results for infected mosquitoes. Each graph represents relative quantities of *AaOr7*, *AaGr1*, and/or *AaGr2* in uninfected and infected female chemosensory tissues. Calibrator samples are those set to a relative quantity of one.


Relative Quantity of AaGr1 in AaGr1 SINV and AaGr1+AaOr7 SINV Infected Mosquitoes
Relative Quantity
1.2
1
0.8
0.6
0.4
0.2
0
TE3'2J2 SINV Antenna
AaGr1 SINV Antenna
AaGr1+AaOr7 SINV Antenna
TE3'2J2 SINV Palp
AaGr1 SINV Palp
AaGr1+AaOr7 SINV Palp
TE3'2J2 SINV Proboscis
AaGr1 SINV Proboscis
AaGr1+AaOr7 SINV Proboscis


Relative Quantity of AaOr7 in AaGr1 SINV and AaGr1+AaOr7 SINV Infected Mosquitoes
Relative Quantity
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
TE3'2J2 SINV Antenna
AaGr1 SINV Antenna
AaGr1+AaOr7 SINV Antenna
TE3'2J2 SINV Palp
AaGr1 SINV Palp
AaGr1+AaOr7 SINV Palp
TE3'2J2 SINV Proboscis
AaGr1 SINV Proboscis
AaGr1+AaOr7 SINV Proboscis

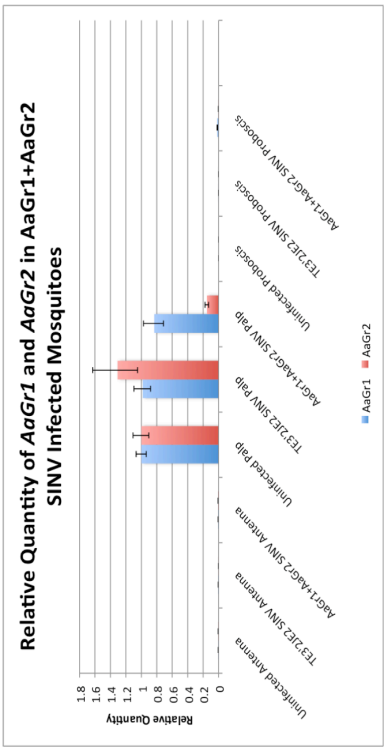
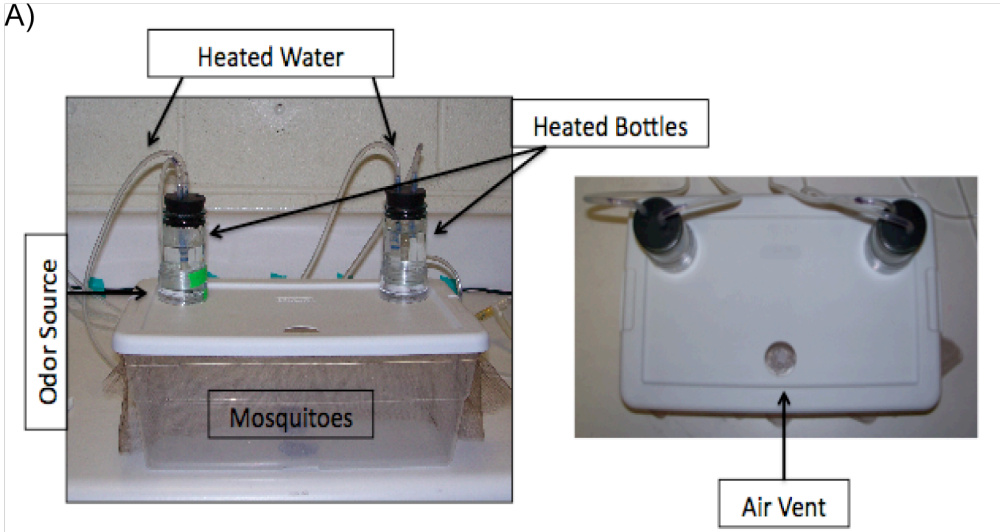
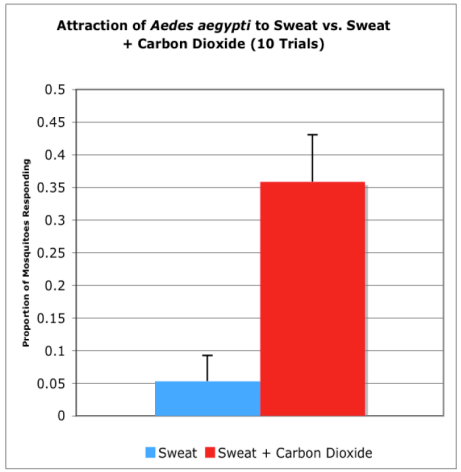

Relative Quantity of AaGr1 and AaGr2 in AaGr1+AaGr2 SINV Infected Mosquitoes
Relative Quantity
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
AaGr1
AaGr2
Uninfected Antenna
TE3'2J2 SINV Antenna
AaGr1+AaGr2 SINV Antenna
Uninfected Palp
TE3'2J2 SINV Palp
AaGr1+AaGr2 SINV Palp
Uninfected Proboscis
TE3'2J2 SINV Proboscis
AaGr1+AaGr2 SINV Proboscis

Figure 4.6: Behavior Assay Setup and Preliminary Results. A) Side and top views of behavior assay setup. Mosquitoes are placed into container and container is sealed by mesh netting. Odor-soaked cotton pads are placed on spacers located above two openings in the lid. Constantly heated water bottles are then placed on top of the spacers. Mosquitoes are able to choice which side of the container to associate with depending on their attraction or repulsion to odors tested. B) Preliminary results of *Ae. aegypti* non-bloodfed female mosquitoes to palm sweat and CO_2 versus palm sweat alone. Results reflect 10 trials involving 250 individual mosquitoes.

A)


Heated Water
Heated Bottles
Odor Source
Mosquitoes
Air Vent

B)


Attraction of Aedes aegypti to Sweat vs. Sweat + Carbon Dioxide (10 Trials)
Proportion of Mosquitoes Responding
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Sweat
Sweat + Carbon Dioxide

4.7: Tables

Table 4.1: Relative quantities of *AaGr1* mRNA in *Ae. aegypti* larvae and pupae

Sample	ΔC_T Standard Error	$\Delta\Delta C_T$	Relative Quantity
1st instar	0.309	0	1
2nd instar	0.128	-3.241	9.454492375
3rd instar	0.147	-3.334	10.08402713
4th instar	0.225	-2.156	4.456774603
Pupae	0.187	-5.344	40.61666835

Table 4.2: Relative quantities of *AaGr1* mRNA in *Ae. aegypti* adult chemosensory tissue

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
NBF Antenna	0.263	-0.668	1.5888688
BF Antenna	N/A	Und.	0
Male Antenna	0.266	0.751	0.594191553
NBF Palp	0.2	-6.691	103.3217375
BF Palp	0.173	-5.593	48.26816245
Male Palp	0.074	-6.823	113.2211757
NBF Proboscis	0.25	0	1
BF Proboscis	0.133	0.417	0.748980467
Male Proboscis	0.204	0.89	0.539614118

Table 4.3: Relative quantities of *AaGr2* mRNA in *Ae. aegypti* larvae and pupae

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
1st instar	0.246	0	1
2nd instar	0.459	-1.135	2.196185628
3rd instar	0.78	0.384	0.76630998
4th instar	0.13	-3.167	8.981771377
Pupae	0.23	-4.118	17.36366995

Table 4.4: Relative quantities of *AaGr2* mRNA in *Ae. aegypti* adult chemosensory tissue

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
NBF Antenna	0.236	-0.921	1.893427262
BF Antenna	0.498	-1.323	2.50185816
Male Antenna	0.324	2.616	0.163119368
NBF Palp	0.134	-6.572	95.141311
BF Palp	0.173	-6.367	82.53876804
Male Palp	0.062	-5.802	55.79252733
NBF Proboscis	0.054	0	1
BF Proboscis	0.124	0.405	0.755236293
Male Proboscis	0.064	0.494	0.710053679

Table 4.5: Relative quantities of *AaGr3* mRNA in *Ae. aegypti* larvae and pupae

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
1st instar	0.127	0	1
2nd instar	0.25	-0.061	1.043188594
3rd instar	0.323	-0.404	1.32317144
4th instar	0.192	1.285	0.410370804
Pupae	0.213	-2.015	4.041805786

Table 4.6: Relative quantities of *AaGr3* mRNA in *Ae. aegypti* adult chemosensory tissue

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
NBF Antenna	0.159	-1.123	2.177994031
BF Antenna	0	-1.612	3.056753042
Male Antenna	0.254	2.374	0.192910022
NBF Palp	0.15	-5.703	52.09236384
BF Palp	0.179	-5.742	53.51976959
Male Palp	0.041	-5.172	36.05181531
NBF Proboscis	0.116	0	1
BF Proboscis	0.257	-4.60E-04	1.000319107
Male Proboscis	0.154	0.81	0.570381858

Table 4.7: Relative quantities of *AaOr7* mRNA in *Ae. aegypti* larvae and pupae

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
1st instar	0.118	0	1
2nd instar	0.222	-1.029	2.040609318
3rd instar	0.329	1.66	0.316439148
4th instar	0.395	2.775	0.146097156
Pupae	0.188	-2.172	4.506476914

Table 4.8: Relative quantities of *AaOr7* mRNA in *Ae. aegypti* adult chemosensory tissue

Sample	**ΔC_T Standard Error**	**$\Delta\Delta C_T$**	**Relative Quantity**
NBF Antenna	0.072	-5.347	40.70121622
BF Antenna	0.084	-5.128	34.96888765
Male Antenna	0.077	-2.892	7.422987812
NBF Palp	0.032	-2.694	6.471050787
BF Palp	0.033	-3.164	8.963113714
Male Palp	0.084	-1.715	3.282966435
NBF Proboscis	0.122	0	1
BF Proboscis	0.095	0.238	0.847919965
Male Proboscis	0.091	-0.344	1.269270886

Table 4.9: Primers used to synthesize SINV inserts

Primer Name	Primer Sequence 5' to 3'	Primer Name	Primer Sequence 5' to 3'	SINV Insert Produced
Gr1-XbaI	CCCCTCTAGACATAAACTACAAAACCCGTTC	Gr1-AscI	AAAAGGCGCGCCGTGTCAACATGCCCCGAACTA	AaGr1
Gr2-Not I	TTTTGCGGCCGCCTGGTTGAACCTCAGCGAGCT	Gr2-PacI	AAAATTAATTAACGCCGTTAGCAACTCCCG	AaGr2
Or7-PacI	AAAATTAATTAAGCAGTTTTCGCATCTTAGCCA	Or7-NotI	AAAAGCGGCCGCGACAAAGTACCATGGGCTGGT	AaOr7
AaOr7-743-393F	TTAATTAACGATAACATTCTTCGGCGATA	AaOr7-393R	TCTAGAAATCTTTCGTGTCGGGATCTTT	AaOr7-393
AaOr7-743-393F	TTAATTAACGATAACATTCTTCGGCGATA	AaOr7-743-946R	TCTAGACACCTGAGCCAGAGCATAGAC	AaOr7-743
AaOr7-946F	TTAATTAACACCCATTCTGTGACCAAGTT	AaOr7-743-946R	TCTAGACACCTGAGCCAGAGCATAGAC	AaOr7-946

Table 4.10: Relative quantities of *AaOr7*, *AaGr1*, and *AaGr2* in infected samples

AaGr1 Real-Time Assay

Sample	ΔC_T Standard Error	$\Delta\Delta C_T$	Relative Quantity
Uninfected Antenna	0.407	7.707	0.00478588
TE3'2JE2 SINV Antenna	0.149	7.459	0.005683518
AaGr1+AaGr2 SINV Antenna	0.163	7.201	0.006796464
Uninfected Palp	0.094	0	1
TE3'2JE2 SINV Palp	0.157	0.026	0.982139595
AaGr1+AaGr2 SINV Palp	0.224	0.264	0.832775771
Uninfected Proboscis	0.902	10.925	0.000514336
TE3'2JE2 SINV Proboscis	0.173	8.586	0.002602295
AaGr1+AaGr2 SINV Proboscis	0.268	5.843	0.017421347

Sample	ΔC_T Standard Error	$\Delta\Delta C_T$	Relative Quantity
TE3'2JE2 SINV Antenna	0.205	5.542	0.021463066
AaGr1 SINV Antenna	0.244	5.358	0.024382671
AaGr1+AaOr7 SINV Antenna	0.09	6.183	0.013763584
TE3'2JE2 SINV Palp	0.12	0	1
AaGr1 SINV Palp	0.133	0.842	0.557869661
AaGr1+AaOr7 SINV Palp	0.604	1.341	0.394746943
TE3'2JE2 SINV Proboscis	0.313	4.944	0.032486857
AaGr1 SINV Proboscis	0.191	5.815	0.017762765
AaGr1+AaOr7 SINV Proboscis	0.228	6.527	0.010843693

AaGr2 Real-Time Assay

Sample	ΔC_T Standard Error	$\Delta\Delta C_T$	Relative Quantity
Uninfected Antenna	0.333	7.512	0.005478513
TE3'2JE2 SINV Antenna	0.095	8.027	0.003833824
AaGr1+AaGr2 SINV Antenna	0.157	7.089	0.007345111
Uninfected Palp	0.146	0	1
TE3'2JE2 Palp	0.317	-0.387	1.307671349
AaGr1+AaGr2 Palp	0.222	2.733	0.150412878
Uninfected Proboscis	0.867	10.951	0.00050515
TE3'2JE2 Proboscis	0.514	9.465	0.001414983
AaGr1+AaGr2 Proboscis	0.23	7.84	0.004364403

AaOr7 Real-Time Assay

Sample	ΔC_T Standard Error	$\Delta\Delta C_T$	Relative Quantity
TE3'2JE2 SINV Antenna	0.204	0	1
AaGr1 SINV Antenna	0.27	-0.59	1.505246747
AaGr1+AaOr7 SINV Antenna	0.234	0.9	0.535886731
TE3'2JE2 SINV Palp	0.091	2.72	0.151774361
AaGr1 SINV Palp	0.931	3.634	0.080548414
AaGr1+AaOr7 SINV Palp	0.637	3.609	0.081956376
TE3'2JE2 SINV Proboscis	0.135	4.886	0.03381952
AaGr1 SINV Proboscis	0.309	3.599	0.082526428
AaGr1+AaOr7 SINV Proboscis	0.191	6.05	0.015092755

Abreviations

24 hours post-bloodfed female mosquito, BF; green fluorescent protein, GFP; gustatory receptor, GR; non-bloodfed female mosquito, NBF; odorant receptor, OR; Sindbis virus, SINV

Acknowledgements

Thanks to Dr. Jim Beidler for the kind gift of his *AaRPS7* primers. Thanks to Randy Saunders for his maintaining the *Ae. aegypti* mosquito stocks and providing all the mosquitoes needed for these experiments. Thanks to both Drs. Zach Adelman and Kevin Myles for all their help with the sindbis virus, including providing us with the TE3'2Jmcs, TE3'2JE2, and TE3'2JGFP plasmids.

4.8: Bibliography

1. Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR: **A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila.** *Neuron* 1999, **22**:327-338.
2. Gao Q, Chess A: **Identification of candidate Drosophila olfactory receptors from genomic DNA sequence.** *Genomics* 1999, **60**:31-39.
3. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R: **A spatial map of olfactory receptor expression in the Drosophila antenna.** *Cell* 1999, **96**:725-736.
4. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ: **G protein-coupled receptors in Anopheles gambiae.** *Science* 2002, **298**:176-178.
5. Bohbot J, Pitts RJ, Kwon H-W, Ruťzler M, Robertson HM, Zwiebel LJ: **Molecular characterization of the Aedes aegypti odorant receptor gene family.** *Insect Mol Biol* 2007, **16**:525-537.
6. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family.** *Genome Res* 2006, **16**:1395-1403.
7. Wanner KW, Anderson AR, Trowell SC, Theilmann DA, Robertson HM, Newcomb RD: **Female-biased expression of odourant receptor genes in the adult antennae of the silkworm, Bombyx mori.** *Insect Mol Biol* 2007, **16**:107-119.
8. Engsontia P, Sanderson AP, Cobb M, Walden KKO, Robertson HM, Brown S: **The red flour beetle's large nose: An expanded odorant receptor gene family in Tribolium castaneum.** *Insect Biochem Mol Biol* 2008, **38**:387-397.
9. Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM: **A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid.** *Proc Natl Acad Sci U S A* 2007, **104**:14383-14388.
10. Hallem EA, Fox AN, Zwiebel LJ, Carlson JR: **Olfaction: mosquito receptor for human-sweat odorant.** *Nature* 2004, **427**:212-213.
11. Wetzel CH, Behrendt HJ, Gisselmann G, Störtkuhl KF, Hovemann B, Hatt H: **Functional expression and characterization of a Drosophila odorant receptor in a heterologous cell system.** *Proc Natl Acad Sci U S A* 2001, **98**:9377-9380.
12. Nakagawa T, Sakurai T, Nishioka T, Touhara K: **Insect sex-pheromone signals mediated by specific combinations of olfactory receptors.** *Science* 2005, **307**:1638-1642.
13. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB: **Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction.** *Neuron* 2004, **43**:703-714.
14. Störtkuhl K, Kettler R: **Functional analysis of an olfactory receptor in Drosophila melanogaster.** *Proc Natl Acad Sci U S A* 2001, **98**:9381-9385.
15. Neuhaus EM, Gisselmann G, Zhang W, Dooley R, Störtkuhl K, Hatt H: **Odorant receptor heterodimerization in the olfactory system of Drosophila melanogaster.** *Nat Neurosci* 2005, **8**:15-17.
16. Kiely A, Authier A, Kralicek AV, Warr CG, Newcomb RD: **Functional analysis of a Drosophila melanogaster olfactory receptor expressed in Sf9 cells.** *J Neurosci Methods* 2007, **159**:189-194.
17. Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K: **Insect**

olfactory receptors are heteromeric ligand-gated ion channels. *Nature* 2008, **452**:1002-1006.
18. Wicher D, Schäfer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS: **Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels.** *Nature* 2008, **452**:1007-1011.
19. Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR: **Integrating the molecular and cellular basis of odor coding in the Drosophila antenna.** *Neuron* 2003, **37**:827-841.
20. Hallem EA, Ho MG, Carlson JR: **The molecular basis of odor coding in the Drosophila antenna.** *Cell* 2004, **117**:965-979.
21. Kreher SA, Kwon JY, Carlson JR: **The molecular basis of odor coding in the Drosophila larva.** *Neuron* 2005, **46**:445-456.
22. Goldman AL, van der Goes van Naters W, Lessing D, Warr CG, Carlson JR: **Coexpression of two functional odor receptors in one neuron.** *Neuron* 2005, **45**:661-666.
23. Hallem EA, Carlson JR: **Coding of odors by a receptor repertoire.** *Cell* 2006, **125**:143-160.
24. Clyne PJ, Warr CG, Carlson JR: **Candidate taste receptors in Drosophila.** *Science* 2000, **287**:1830-1834.
25. Dunipace L, Meister S, McNealy C, Amrein H: **Spatially restricted expression of candidate taste receptors in the Drosophila gustatory system.** *Curr Biol* 2001, **11**:822-835.
26. Scott K, Brady R, Cravchik A, Morozov PS, Rzhetsky A, Zuker C, Axel R: **A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila.** *Cell* 2001, **104**:661-673.
27. Suh GSB, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ: **A single population of olfactory sensory neurons mediates an innate avoidance behaviour in Drosophila.** *Nature* 2004, **431**:854-859.
28. Dahanukar A, Foster K, van der Goes van Naters W, Carlson JR: **A Gr receptor is required for response to the sugar trehalose in taste neurons of Drosophila.** *Nat Neurosci* 2001, **4**:1182-1186.
29. Ueno K, Ohta M, Morita H, Mikuni Y, Nakajima S, Yamamoto K, Isono K: **Trehalose sensitivity in Drosophila correlates with mutations in and expression of the gustatory receptor gene Gr5a.** *Curr Biol* 2001, **11**:1451-1455.
30. Slone J, Daniels J, Amrein H: **Sugar receptors in Drosophila.** *Curr Biol* 2007, **17**:1809-1816.
31. Lu T, Qiu Wang Ruťzler M, Kwon H-W, Pitts RJ, Loon v, Takken W, Carlson JR, Zwiebel LJ: **Odor Coding in the Maxillary Palp of the Malaria Vector Mosquito Anopheles gambiae.** *Curr Biol* 2007, **17**:1533-1544.
32. Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB: **Two chemosensory receptors together mediate carbon dioxide detection in Drosophila.** *Nature* 2007, **445**:86-90.
33. Kwon JY, Dahanukar A, Weiss LA, Carlson JR: **The molecular basis of CO2 reception in Drosophila.** *Proc Natl Acad Sci U S A* 2007, **104**:3574-3578.
34. Takken W, Knols BGJ: **Odor-mediated behavior of Afrotropical malaria mosquitoes.** *Annu Rev Entomol* 1999, **44**:131-157.
35. Johnson BW, Olson KE, Allen-Miura T, Rayms-Keller A, Carlson JO, Coates CJ, Jasinskiene N, James AA, Beaty BJ, Higgs S: **Inhibition of luciferase expression in transgenic Aedes aegypti mosquitoes by Sindbis virus expression of antisense**

luciferase RNA. *Proc Natl Acad Sci U S A* 1999, **96**:13399-13403.
36. Adelman ZN, Blair CD, Carlson JO, Beaty BJ, Olson KE: **Sindbis virus-induced silencing of dengue viruses in mosquitoes.** *Insect Mol Biol* 2001, **10**:265-273.
37. Travanty EA, Adelman ZN, Franz AWE, Keene KM, Beaty BJ, Blair CD, James AA, Olson KE: **Using RNA interference to develop dengue virus resistance in genetically modified Aedes aegypti.** *Insect Biochem Mol Biol* 2004, **34**:607-613.
38. Kent LB, Walden KKO, Robertson HM: **The Gr Family of Candidate Gustatory and Olfactory Receptors in the Yellow-Fever Mosquito Aedes aegypti.** *Chem senses* 2007.
39. Melo ACA, Ruťzler M, Pitts RJ, Zwiebel LJ: **Identification of a chemosensory receptor from the yellow fever mosquito, Aedes aegypti, that is highly conserved and expressed in olfactory and gustatory organs.** *Chem Senses* 2004, **29**:403-410.
40. Uhlirova M, Foy BD, Beaty BJ, Olson KE, Riddiford LM, Jindra M: **Use of Sindbis virus-mediated RNA interference to demonstrate a conserved role of Broad-Complex in insect metamorphosis.** *Proc Natl Acad Sci U S A* 2003, **100**:15607-15612.
41. Faucher C, Forstreuter M, Hilker M, de Bruyne M: **Behavioral responses of Drosophila to biogenic levels of carbon dioxide depend on life-stage, sex and olfactory context.** *The Journal of experimental biology* 2006, **209**:2739-2748.
42. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**:402-408.