UNIVERSITY OF THE WESTERN CAPE

RESEARCH THESIS

Name of candidate

Obiero, George Fredrick Opondo

MSc. (Maseno University), *BSc. Hons. (Egerton University)*

Student number

3180446

Degree Award

Doctor of Philosophy

Program/Department

SANBI

Title of thesis

GENOME-WIDE ANNOTATION OF CHEMOSENSORY AND GLUTAMATE-GATED RECEPTORS, AND RELATED GENES IN *Glossina morsitans morsitans* TSETSE FLY

Supervisor

Prof. Alan Christoffels,

University of the Western Cape, South Africa

Co-supervisor(s)

Dr. Paul O. Mireji, *Yale University, CT, USA*

Dr. Daniel Masiga, *icipe, Kenya*

Date

November, 2014

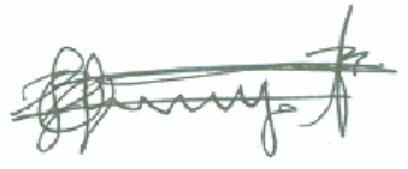
DECLARATION

I **Obiero**, George Fredrick Opondo, declare that this thesis:

(a). is my original research work;

(b). has not been presented before at this University or any other University/institution for a degree award of any kind; and

(c). does not incorporate any published work or material from another thesis.

Sign:  Date: ….....

COPYRIGHT

ABSTRACT

Tsetse flies are the sole vectors of trypanosomes that cause nagana and sleeping sickness in animals and humans respectively in tropical Africa. Tsetse are unique: both sexes adults are exclusive blood-feeders, females are mated young and give birth to a single mature larva in sheltered habitats per pregnancy. Tsetse use chemoreception to detect and respond to chemical stimuli, helping them to locate hosts, mates, larviposition and resting sites. The detection is facilitated by chemoreceptors expressed on sensory neurons to cause specific responses. Specific molecular factors that mediate these responses are poorly understood in tsetse flies. This study aimed to identify and characterize genes that potentially mediate chemoreception in *Glossina morsitans morsitans* tsetse flies. These genes included sensory odorant (OR), gustatory (GR), ionotropic (IR), and related genes for odorant-binding (OBP), chemosensory (CSP) and sensory neuron membrane (SNMP) proteins. Synaptic transmission in higher brain sites may involve ionotropic glutamate-gated (iGluR) and metabotropic glutamate-gated (mGluR) receptors. The genes were annotated in *G. m. morsitans* genome scaffold assembly GMOY1.1 Yale strain using orthologs from *D. melanogaster* as query via TBLASTX algorithm at e-value below 1e-03. Positive blast hits were seeded as gene constructs in their respective scaffolds, and used as genomic reference onto which female fly-derived RNA sequence reads were mapped using CLC Genomics workbench suite. Seeded gene models were modified using RNA-Seq reads then viewed and re-edited using Artemis genome viewer tool. The genome was iteratively searched using the *G. m. morsitans* gene model sequences to recover additional similar hit sequences. The gene models were confirmed through comparisons against the NCBI conserved domains database (CDD) and non-redundant Swiss-Prot database. Trans-membrane domains and secretory peptides were predicted using TMHMM and SignalP tools respectively. Putative functions of the genes were confirmed via Blast2GO searches against gene ontology database. Evolutionary relationships amongst and between the genes were established using maximum likelihood estimates using best fitting amino acid model test in MEGA5 suite and PhyML tool. Expression profiles of genes were estimated using the RNA-seq data via CLCGenomics RNA-sequences analysis pipeline. Overall, 46 ORs, 14 GRs, and 19 IRs were identified, of which 21, 6 and 4 were manually identified for ORs, GRs, and IRs respectively. Additionally, 15 iGluRs, 6 mGluRs, 5 CSPs, 15 CD36-like, and 32 OBPs were identified. Six copies of OR genes (*GmmOR41-46*) were homologous to DmelOr67d, a single copy *cis* vacenyl acetate (cVA) receptor . Genes whose receptor homologs are associated with responses to CO_2, *GmmGR1-4*, had higher expression profiles from amongst glossina GR genes. Known core-receptor homologs OR1, IR8a, IR25a and IR64a were conserved, and three species-specific divergent IRs (IR10a, IR56b and IR56d) were identified. Homologs of GluRIID, IR93a, and sweet taste receptors (Gr5a and Gr64a) were not identified in the genome. Homolog for LUSH protein, GmmOBP26, and sensory neuron membrane receptors SNMP1 and SNMP2 were conserved in the genome. Results indicate reduced repertoire of the chemosensory genes, and suggest reduced host range of the tsetse flies compared to other Diptera. Genes in multiple copies suggest their prioritization in chemoreception, which in turn may be tied to high specificity in host selection. Genes with high sequence conservation and expression profiles probably relate to their broad expression and utility within the fly nervous system. These results lay foundation for future comparative studies with other insects, provide opportunities for functional studies, and form the basis for re-examining new approaches for improving tsetse control tools and possible drug targets based on chemoreception.

THESIS TITLE

GENOME-WIDE ANNOTATION OF PUTATIVE CHEMOSENSORY AND GLUTAMATE-GATED RECEPTORS, AND RELATED GENES IN *Glossina morsitans morsitans* TSETSE FLY

Running title

Chemosensory and glutamate-gated receptor genes in *Glossina morsitans morsitans*

Key words

Glossina morsitans morsitans
Olfaction
Gustation
Odorant receptors
Gustatory receptors
Ionotropic receptors
Ionotropic glutamate gated receptors
Metabotropic glutamate receptors
Sensory neuron membrane proteins
Odorant binding proteins
Chemosensory proteins

DEDICATION

I dedicate this thesis to my family, who were the source of my energy every morning every hour, and my support in the distant depths of science.

ACKNOWLEDGEMENT

My sincere gratitude and honor goes to my university supervisor Professor Alan Christoffels and entire staff at SANBI, University of the Western Cape, who were supportive in my entire research work. I appreciate the financial support from the South African Medical Research Council (MRC). I also thanks to Dr. Daniel Masiga of *icipe* and Dr. Paul Mireji of Yale University for their supervision and mentorship. I recognize DAAD fellowship awarded to me via ARPPIS program at *icipe*, without which I could not have carried out my research. I sincerely appreciate *icipe* management for availing facilities to help me do my work in an enhancing environment the entire life of my PhD. thesis. In particular, my sincere thanks to Ms. Lilian Igweta of icipe CB&ID and her staff Ms. Lisa Omondi and Mrs. Margaret Ochanda.

My project group members, 'the Glossina', Dr. Steven Nyanjom, Ms. Rosaline Macharia, and Mr. Kevin Marucha I thank you for the long hours we all endured doing the 'bioinformatics' – a true testimony that we can teach ourselves the things we never learnt in school. The efforts you made in commenting and providing insights to my manuscripts and thesis drafts were immeasurable. The team spirit will make us stand, keep it up.

Much appreciations to the IT departments of icipe via Mr. Glenn, ILRI via Mr. Alan Orth, and departmental administration at SANBI through Ms Maryam Salie for mental and technical support and Ms. Esther Waweru of *icipe* Molecular Biology and Bioinformatics Unit (MBBU) department. I found a strong support group from all of you in terms of setting up computing networked platforms and operational logistics. I salute all the members of *icipe* for their immense moral and social support. I reckon the athletic spirit we enjoyed together in our marathon jogging team, led by Harrison, Nelly, Mary, Vincent, Kevin, Kabii, Davy, and other team-mates from management like Raph, Caro, Siprin, and others. Research is much like running a marathon, you get frustrated along the way and when on the brinks of giving up you realize you have actually reached the finish line! Continue the charity run!

God bless you all for your priceless support.

LIST OF ABBREVIATIONS

AAT	Animal African Trypanosomiasis
bp	base pairs
B2G	Blast2GO
CEGMA	Core Eukaryotic Gene Mapping Approach
CNS	Central Nervous System
CSPs	Chemosensory-specific proteins
cVA	(Z)-*11-cis*-vacenyl acetate
DALYs	Daily Adjusted Life Years
DNA	Deoxy-ribonucleic acid
EAAT	Excitatory Amino Acid Transporter
EST	Expressed Sequence Tag
FAO	Food and Agriculture Organization
GABA	Gamma Amino Butyric Acid
GO	Gene Ontology
GPCRs	G-protein-coupled receptors
GRs	Gustatory Receptors
HAT	Human African trypanosomiasis
HMM	Hidden Markov Models
icipe	International Center of Insect Physiology and Ecology
IGGI	International Glossina Genome Initiative
iGluRs	ionotropic Glutamate-gated ion channel Receptors
ILRI	International Livestock Research Institute
IRs	Ionotropic Receptors
LBDs	Ligand Binding Domains
Mbp	Mega base pairs
MEGA	Molecular Evolutionary Genetics Analysis
mGluRs	metabotropic Glutamate-gated Receptors

NTDs	Neglected Tropical Diseases
OBPs	Odorant Binding Proteins
ORF	Open Reading Frame
ORNs	Odorant Receptor Neurons
ORs	Odorant Receptors
PNS	Peripheral Nervous System
RNA	Ribonucleic acid
RPKM	reads per kilobase of mapped exons per million reads
SIT	Sterile Insect Techniques
SNMPs	Sensory Neuron Membrane Proteins
SOG	sub-esophageal ganglion
VSG	Variant Surface Glycoproteins
WHO	World Health Organization

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF PUBLICATIONS

1. Daniel Masiga, **George Obiero**, Rosaline Macharia, Paul Mireji, Alan Christoffels (2014). Chemosensory receptors in tsetse flies provide link between chemical and behavioural ecology. *Trends in Parasitology*. DOI:10.1016/j.pt.2014.06.007. Available http://www.sciencedirect.com/science/article/pii/S147149221400110X.

2. **Obiero GFO**, Mireji PO, Nyanjom SRG, Christoffels A, Robertson HM, & Masiga DK (2014). Odorant and Gustatory Receptors in tsetse fly *Glossina morsitans morsitans. PLoS Negl Trop Dis* 8: e2663. DOI:10.1371/journal.pntd.0002663

3. International Glossina Genome Initiative (IGGI) (2014). Genome Sequence of the Tsetse Fly (*Glossina morsitans*): Vector of African Trypanosomiasis. *Science* 344 (6282): 380-386. DOI: 10.1126/science.1249656. (**Obiero GFO** co-authored chemosensory section).

4. **Obiero GFO**, Masiga DK, Mireji PO, Alan Christoffels (xxxx). Conserved glutamate-gated ion receptors in tsetse fly *Glossina morsitans morsitans* – (In preparation)

5. **Obiero GFO**, Mireji PO, Masiga DK, Alan Christoffels, (xxxx). Conserved chemosensory soluble proteins in tsetse fly *Glossina morsitans morsitans* – (In preparation)

TABLE OF CONTENTS

UNIVERSITY of the
WESTERN CAPE

CHAPTER ONE

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Chemoreception is an evolutionary survival endowment present in prokaryotes and eukaryotes (Ishimoto & Tanimura 2004), and mediates an organism's interaction with its immediate environment through olfaction and gustation. Olfaction detects and responds to aerial volatiles (smell molecules), while gustation detects and responds to semi-soluble and solid chemicals including sweet tastes, bitter tastes, and soluble CO_2 gas. These molecules (odorant and gustatory) are also described as semiochemicals when they originate from other organisms and perceived by others (Reinhard, 2004). Detection of semiochemicals form the basis of insect behavioral responses while searching for important resources (reviewed in Hilldebrand, 1995). Olfaction and gustation relate to the sensation of the semiochemicals by the peripheral nervous system (PNS) nerve endings exposed to external environment according to each organism's morphological forms. The semiochemicals interact with sensory nerve-ending surface proteins (receptors) such as odorant receptors (ORs), ionotropic receptors (IRs), and gustatory receptors (GRs). These receptors are expressed in peripheral organs, mainly antennae, maxillary palps, legs, ovipositors, and wings. In insects, these receptors are re-classified as ionotropic receptors since they function as ligand-gated ion channels, probably regulated by downstream second messenger cascades, different from non-chemoreceptor channels downstream of sensory neurons such as ionotropic glutmate-gated receptors (iGluRs), and metabotropic glutamate receptors (mGluRs) (Wicher, 2013). The detection starts with direct or indirect binding (via odorant binding proteins (OBPs) or pheromone binding proteins (PBPs)) of semiochemicals onto receptors, then conversion of chemical message to electrical signals. The electrical signals are relayed to higher brain sites in the central nervous system (CNS) for integration, perception and response via the motor system, a process that is markedly different from prokaryotes to lower and to higher eukaryotes (Hansson, 1995). The basic process is, however, fundamentally the same, and mediates such important behaviors as mate finding and feeding.

Mating is an important progenation process in the life of any organism, and insects detect potentially mature mates via semiochemicals (pheromones) emanating from the opposite sexes using a combination of many proteins. For instance in *Drosophila melanogaster*, detection of male specific pheromone, (Z)-11-*cis* vaccenyl acetate (cVA), requires Or67d receptor, an extracellular Obp76a, (also known as LUSH protein), and a trans-membrane CD36-like receptor, SNMP (Xin *et al.*, 2008). The LUSH protein (a member of odorant binding protein, OBP, subfamily) is necessary for binding and

presenting cVA to the Or67d (Benton *et al*., 2007; Jin *et al*., 2008; Dahanukar & Ray *et al*., 2011). The SNMP is thought to make the union of Or67d and LUSH reach an optimal sensitivity threshold to detect cVA, hence its role in pheromone signal perception on the antennal trichoid neurons (Benton *et al.*, 2007; Xin *et al.*, 2008; Vogt, 2009). Additionally, SNMP displays dimorphic attractive mating behavior in the presence of cVA, and also inhibitory phenotypes in the absence of the ligand (Xin *et al*., 2008; Vogt, 2009). Other proteins, chemosensory proteins (CSPs), are associated with chemosensation because some of them are expressed in the antennae, an organ dedicated to chemoreception. This thesis concentrated on the identification, annotation and analysis of the chemosensory receptor related genes encoded in the genome of a blood-feeding Diptera, *Glossina morsitans morsitans* Westwood tsetse fly, with focus on ORs, GRs, IRs, iGluRs, mGluRs, OBPs, CSPs, and SNMPs. The outcome will serve as precedent information relating to molecular chemoreception in the genus Glossina. In addition, the information will be useful in enabling any future neurogical, functional and comparative analyses with other species and related members of diptera.

1.2 Tsetse flies physiology and ecology

Tsetse flies (*Glossina* sp) are exclusive blood feeding insects, endemic to tropical Africa and are the sole vectors of sleeping sickness in humans and nagana in livestock (FAO, 1992; WHO, 2010). The flies are unique among insect vectors since both sexes are hematophagous, strict blood feeders. Females are mated once at teneral stage and are adenoviviparous i.e. give birth to live larvae – one per pregnancy resulting in low progeny (Rogers *et al*., 1994; Attardo *et al.*, 2010). They have a pair of antennae, which are distinctively long with prominent aristae, connected beneath to olfactory pits. Their stylet-like proboscis is long and narrow, pointing directly in-front, while their wings fold back atop each other seamlessly. They have ommatidia-loaded compound eyes sensitive to near object movement, and ocelli with three simple eyes sensitive to long distance acuity and cryptic host distinction at high light intensities during flight (FAO, 1992; Rogers *et al.*, 1994). These features (Figure 1.1) enable the fly to efficiently seek and obtain both blood-meal and mates, and to enhance survivorship of their progeny by selecting a suitable larviposition site.

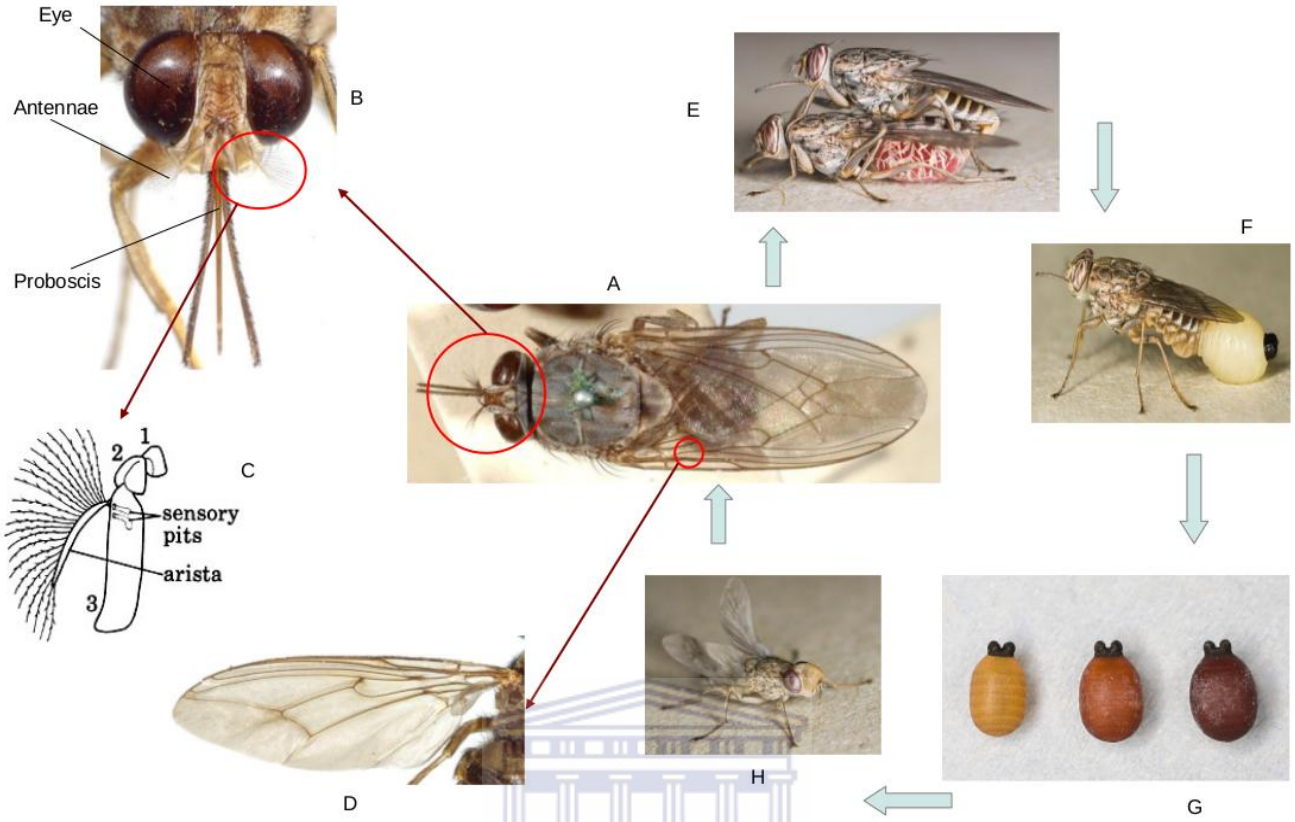


Figure 1.1 Distinctive tsetse fly head features (A – D) and life cycle (A, E-H).
A – a mature adult fly aerial view with folded wings; B – head region showing compound eyes, proboscis and antenna; C – fly antenna showing segments 1 – 3, sensory pits, and arista plumes; D – hatchet shape of fly wing; E – young flies (teneral) mating after female's first blood-meal; F – parturition of mature larva; G – pupae at different stages; H – an emergent adult (from eclosion) with developing wings. (Adapted from – www.raywilsonbirdphotography.co.uk/).

1.2.1 Tsetse fly taxonomy

Tsetse flies belong to the order Diptera, family Glossinidae, and all species are of the genus *Glossina.* There are 34 species and subspecies described, spread across 38 African countries (Gooding and Krafsur, 2005; FAO, 2008; Krafsur, 2009). The species are split into three groups based on a combination of distributional, behavioral, molecular and morphological characteristics. The groups are savannah flies, (also sub genus *Morsitans*) including *Glossina morsitans morsitans* (Westwood 1850), *G. m. submorsitans, G. pallidipes* (Austen 1903), and *G. austeni* (Newstead 1912)*;* forest flies (sub-genus *Fusca*) e.g. *G. fusca fusca* (Walker 1849), *G. fuscipleuris* (Austen 1911), and the newest species *G. fresili* (Gouteux 1987); and riverine flies (sub-genus *Palpalis*) e.g. *G. fuscipes fuscipes* (Newstead 1911), *G. caliginea* (Austen 1911), and *G. fuscipes martini* (Zumpt 1935) (reviewed in WHO, 2004).

1.2.2 Pathogenesis of African Trypanosomiasis

Tsetse flies are the sole vectors of the kinetoplastid protozoan, trypanosome, which causes Human African Trypanosomiasis (HAT) and Animal African Trypanosomiasis (AAT) (WHO, 2010). Though all tsetse flies can potentially transmit all trypanosomes, different species are specialized in transmitting specific trypanosomes to different vertebrates, resulting in induced chronic and/or acute forms of the disease in humans, livestock, and vertebrate wildlife (FAO, 2008). Chronic forms of HAT experienced in west and central Africa are caused by *T. brucei gambiense* transmitted by *G. palpalis, G. fuscipes,* and *G. tachinoides,* while the acute forms in east and southern Africa are caused by *T. brucei rhodesiense* are vectored mainly by *G. m. morsitans, G. swynnertoni,* and *G. pallidipes* (Gooding & Krafsur, 2005; FAO, 2008; Krafsur 2009). The acute nagana in cattle, camels, antelopes, domestic pigs, and horses is caused by *T. brucei brucei, T. vivax* and *T. simiae* while the chronic form is mainly inoculated by *T. congolense* in cattle, camels, horses, and *T. suis* in domestic pigs and warthogs (Mugasa *et al.*, 2008). The *G. m. morsitans* is the most important vector for AAT.

The strict blood-sucking lifestyle of both sexes of tsetse enhances their ability to spread trypanosomes to a wider vertebrate host range (Aksoy *et al.*, 2005; Solano *et al.*, 2010). Tsetse transmit trypanosomes in two ways: mechanically by temporarily feeding on an infected host, and then shifting to another healthy host to satisfy their appetite thereby infecting the healthy host; and biologically whereby the trypanosomes undergo incubative sexual phase in the tsetse after first ingestion before inoculating a healthy host. The infected hosts, if not treated early, often die (Masiga *et al.*, 2002). However, before the hosts die,, new vectors easily get their first infected blood-meal, thereby re-starting the cycle of infection. Apparently, the many infected hosts (especially wildlife) act aspathogen reservoirs, without suffering disease (WHO, 2004). Infected flies remain infective for the rest of their life (Steverding, 2008). Unlike other insect vector related infections, tsetse flies and their vertebrate hosts develop no immunity against trypanosoma infection. However, the flies initiate a late immune response through pathogen clearance mechanism in the gut, but only after the pathogen have sufficiently established and gained numerical advantage of being able to be transmitted with any subsequent host bites (Aksoy, 2003). Unfortunately, in both animal and human hosts, the trypanosomes have developed drug resistance resulting from the prolonged use of decades-old only veterinary and medical drugs available (Stich *et al.*, 2002). Of all species, the Morsitans group of tsetse flies is often the most efficient vectors of the trypanosomes causing acute forms in east and southern African animals, perhaps due to its good diurnal behavior – being active early in the morning and late afternoon

when the weather is calm, and host plumes are stable (WHO, 2004). This partly informed the choice of this thesis topic, targeting the identification of chemosensory receptors in the *G. m. morsitans* genome.

1.2.3 Tsetse infestation and Trypanosomiasis economic impact

Tsetse flies inhabit much of Africa between the Sahara and Kalahari desert latitudes 14° N and 29° S, and the highland water-towers of Indian and Atlantic oceans, occupying an area of over 10 million square kilometers in close to 40 countries (Kamuanga, 2003). However, the individual species are concentrated in specific foci within the regional zones they inhabit (Figure 1.2), thus defining the truncated spread of the trypanosomiasis across the continent (Rogers *et al.*, 1996; Steverding *et al*., 2004). Perhaps the most dreaded and devastating plague that ever visited Africa are the tsetse flies, which have gained extensive study because of their importance in transmission of trypanosomiasis, resulting in huge losses in both veterinary and human health in the entire tropical sub-Saharan Africa (Barrett *et al.*, 2003; Simarro *et al.*, 2008). An estimated 60 million people in 38 sub-Saharan African countries are at risk of HAT, with direct reported total economic loses at over two billion US dollars (Kioy *et al*., 2004; WHO, 2010). Cattle deaths per year reach over three million causing about 1.5 million Daily Adjusted Life Years (DALYs) (Stich *et al*., 2002; FAO, 2008; WHO, 2011). The human livelihoods in the infested regions are negatively affected: people and their animals lie sick or die, agricultural lands remain fallow and uninhabited, while scientific research for potential drug targets are frustrated due to low funding and difficulty in establishing laboratory colonies of the flies and pathogens. Attempts to develop cost-effective drugs and vaccines against both blood-stages and nervous system stages of the pathogen have largely failed, and the current drugs in use have undesirable toxicity fatalities (Stich *et al.*, 2002; WHO, 2004).

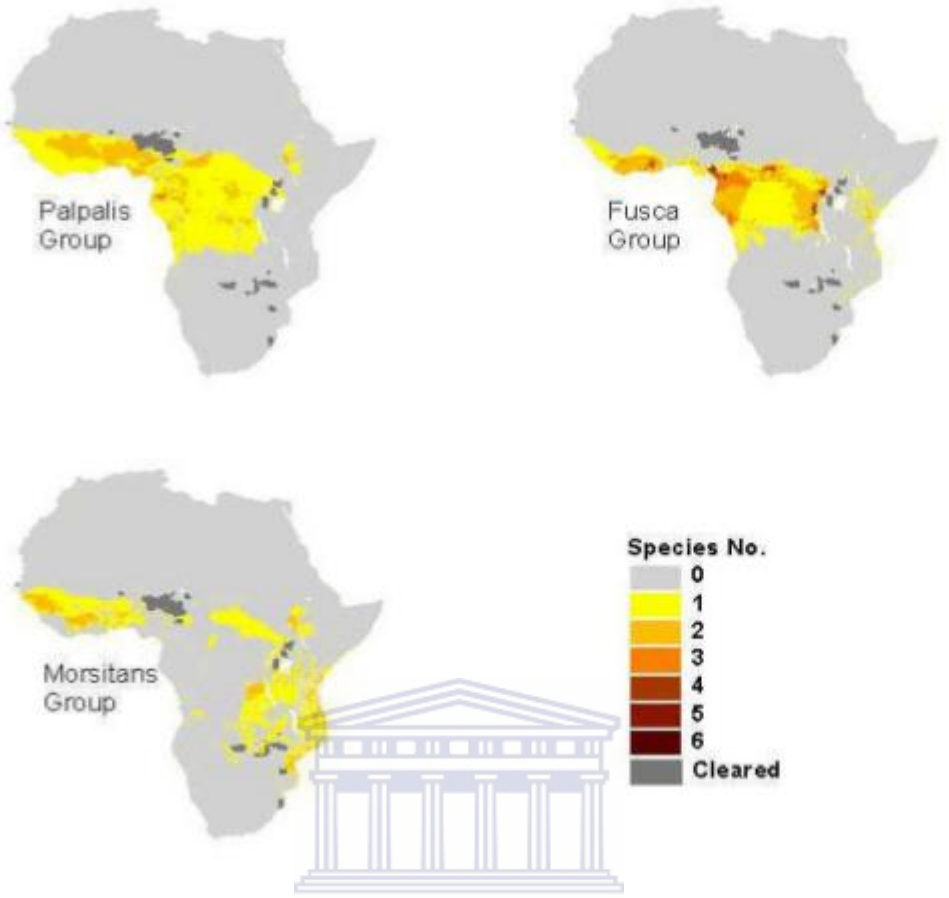


Figure 1.2 Distributions of three main tsetse fly sub-species.
The palpalis and fusca groups overlap within humid and semid-humid sub-Saharan riverines and forests of central and west Africa. The morsitans are mainly found in the arid and semi-arid lands of east and southern Africa, but also found in west Africa. FAO web site: http://ergodd.zoo.ox.ac.uk/livatl2/tsetse.html/).

1.2.4 Approaches for Controlling Tsetse flies and Trypanosomiases

Two main alternative strategies have been undertaken to reduce the incidences of African trypanosomiasis. One is aimed at combating the trypanosomes' establishment in the hosts via detection and treatment, and the other on the trypanosome life-cycle interruption to stop vector-host contact via vector population reduction.

1.2.4.1 Medical/Veterinary Strategies

This strategy targets the disease pathogens directly using monitoring and surveillance for disease occurrence and prevalence, and pathogenic detection and treatment of victims, so as to reduce the chances of disease transmission (FAO, 1992; WHO, 2004). Estimating the overall presence of both the tsetse vectors and the diseases (the acute and chronic forms

of both nagana and sleeping sickness) has been difficult, as the available literature points out (Stich *et al.*, 2002). Neither are the statistics clear about the pathogen reservoirs especially for the eastern and southern Africa trypanosomiases, because very diverse non-human and non-livestock vertebrates (that apparently do not suffer the disease) harbor high pathogen load, and subsequently act as sources of re-infection (WHO, 2004; ILRI, 2011). However, all previous biogeographic mapping of main infestation regions only focused on detection and treatment using unfortunately, toxic arsenic melarsoprol and pentamindine (Stich *et al.*, 2002; WHO, 2004). The challenge remains though: there are no vaccines against the trypanosoma, whose antigenic variant surface glycoproteins (VSGs) vary widely (FAO, 1992; WHO, 2004). This approach has also been hampered by low research incentive towards development of potential new drug targets against neglected tropical diseases (NTDs) in general (Aksoy, 2003; Alves-Silva *et al.*, 2010). An over three decades old partnership between WHO and major pharmaceutical companies to develop some front-line drugs (WHO, 2004), brought short-lived relief and hope occasioned by numerous reasons: poor infrastructural development in most resource-poor African communities that frustrate transportation, storage and administration of the drugs (Barrett *et al.*, 2003; Baker *et al.*, 2011), situations that are worsened by perennial civil wars common in sub-Saharan Africa (Stich *et al.*, 2002). Further, the strategy has been plagued by not only the lengthy regime of drug administration to both humans and animals, but also high cost of the drugs and high fatalities resulting from their toxicity (FAO, 1992; Stich *et al.*, 2002; WHO, 2010). The resultant drug lethargic and fatal states in both human and animal victims aggravate and complicate the already devastating tryapanosomiasis negative economic status of the communities.

1.2.4.2 Entomological strategy

This strategy is aimed at disrupting the cycle of transmission by reducing tsetse fly population by total eradication (Barrett *et al.*, 2003). This strategy is more promising and informs current research efforts towards management of trypanosomes in general. Initially, very rudimentary methods of vegetation clearing and killing of suspected reservoir hosts were used (WHO, 2004). Though the resultant environmental impacts were not desirable, these methods significantly reduced vector populations and trypanosome loads in regions where they were applied. The methods were then coupled with an approach of treating alternative hosts, but the sheer number of wild hosts and the level of re-infection could not be matched for its long term use (Rogers *et al.*, 1994). The use of insecticides sprayed in/directly onto insects, habitats, and target hosts were most successful and a reliable method for many years – across the major continental and oceanic-island tsetse infestation

belts (FAO, 1992). As the overall tsetse population drastically dropped, large-scale insecticide application, especially aerial as was applied in Botswana (FAO, 1992), unfortunately became uneconomical. Furthermore, re-infestation that emanated from the neighboring non-sprayed game reservoirs and residually surviving flies, led to a sudden upsurge of the flies. This demanded continuous localized surveillance and treatment. Negatively, the chemical sprays led to ecological poisoning due to unchecked usage, and caused indiscriminate killing and/or cumulative effects in non-target organisms. Environmentally safe sterile insect technique (SIT) was then successfully attempted in eradicating tsetse flies in Zanzibar Island (Kamuanga, 2003): it involved laboratory breeding and irradiating of male tsetse flies to make them sterile then releasing them to mate female flies in nature (Kamuanga, 2003). Success with SIT was likely achieved because of isolated population such as in the island. However, SIT was limited by the cost of integrating it with other methods such as use of targets, traps, pour-ons, and chemotherapy (Aksoy, 2003; Gooding & Krafsur, 2005).

1.2.4.3 Olfactory-based Bait-traps and Target lures

The frustrations from searching for a unified control approach applicable for all tsetse species and across the continent, led to re-designing of fly traps and target lures from the initial innovative trap of Harris in 1930s, (reviewed in WHO, 2004). Previous tsetse control attempts focused more on the flies visual acuity and smell towards attractions to traps and targets, which were later laced with insecticides and odors to improve efficiency (Jordan, 1995). Indeed, these early works established the differential attractions and landings onto targets by the different groups of tsetse, where the Morsitans were more attracted to horizontal block shapes than the Palpalis that preferred vertical shapes. These observations corresponded to the feeding preferences by majority of each species members i.e. the Morsitans preferred feeding on mobile vertebrates while the palpalis preferred humans (Jordan, 1995). The flies were also observed to be attracted to blue-black stripped target backgrounds: basically the blue color attracts them from a distance but the black color enhances their landing tendency, and an upward white-netted cloth caging to lure them to attempt escaping upwards whereupon they get trapped and either die of desiccation or can be collected for sampling purposes (FAO, 2008; WHO, 2010; see also Figure 1.3). Alternatively, such attractive background visual lures and targets were laced with insecticides, so that upon contact the flies picked lethal doses that knocked them down or killed them elsewhere on escape (FAO, 2008). Generally, the tsetse flies were observed to be sensitive to shape, color, movement, and smell; and accordingly different types of traps were designed targeting different species in different regions.

Further, new small-scale user-friendly methods based on odor-baited traps and target lures (also called bait technology) were advocated at varied scales (see inset top, Figure 1.3). This involved use of ecological volatiles (naturally from hosts or artificially synthesized) to either attract to trap and kill, or to repel the insects from the hosts. Most of these chemicals are natural emanations from various animals such as metabolic wastes or excreta (Omolo *et al.*, 2009), coupled with pyrethroids which were used as dip and/or pour-on on live animals (as mobile targets) (Kamuanga, 2003), the bait technology was relatively cost effective and technically simple (WHO, 2004). However, it became unsustainable for the resource poor rural farmers who lacked sustenance means, and also suffered field losses due to theft, damages from humans and wild animals, fire or loss from being washed away in the rainy season (Solano *et al.*, 2010). Not all tsetse species in a given locality respond the same way to odor-baited traps and targets.

The above technical failures led to a deeper look into what could possibly be making the different tsetse fly species respond differently to similar baited trap and target designs. Different species and sub-species have attraction preference for particular hosts than to others, necessitating search for molecular understanding of differences in tsetse species' behavioral responses to host, mate-seeking, and oviposition sites. The unity of all these behaviors is olfaction - detection of semio-chemicals (smell molecules) emanating from both hosts and non-hosts vertebrates via insects' peripheral olfactory receptors. Previous studies showed the use of cattle *(Bos taurus)* and buffalo *(Syncerus caffer)* odor attractants used in baited traps and targets against the Morsitans sub-group of *G. m. morsitans* (Owaga *et al.*, 1988; Gikonyo, *et al.*, 2003; Logan & Bickett, 2007). Recently, kairomones from monitor lizards were used to increase landing and trapping rates of the Palpalis sub-group species, *G. fuscipes fuscipes* (Omolo *et al.*, 2009), the most important HAT vector across sub-Saharan Africa. These findings suggested that some mammalian hosts' emanations possibly contained repellent odors similar to the lizard's natural emanations. Indeed, natural emanations from a non-host waterbuck, *Kobus defassa* (Bovidae) (see inset bottom, Figure 1.3) were repellent to the Morsitans sub-group (Gikonyo *et al.*, 2002; 2003), and have recently been incorporated into, perhaps, the most novel tool ever devised as a simple collar strap that is hanged around the farms or on cattle necks to slowly release the laced water-buck repellent stench by researchers at International Center of Insect Physiology and Ecology, (icipe), Kenya (see inset middle, Figure 1.3; Saini & Hassanali, 2007). So far, the pilot field trials with the 'cow collar' amongst pastoral communities in major tsetse-prone areas (foci) have been promising success in repelling the vectors from the hosts (Saini & Hassanali, 2007). Clearly, all these previous reports have instrumentally, contributed to the understanding of the chemical ecology underlying odorant detection and behavioral responses in the tsetse vectors. However, these strategies are

yet to be applied uniformly across the three tsetse groups and sub-species in various focal sites. This emanates from the reports that while the lures and targets can be effective in trapping and repelling, for instance the savanna flies, they often fail when applied for the riverine and forest groups of flies (FAO, 2008). Consequently, there has been a resurgence of the trypanosomiasis in the past decades in specific geographical foci, making the disease to be re-classified as one of the re-emerging NTDs that urgently require new control approaches and tools (Aksoy, 2003; WHO, 2011).

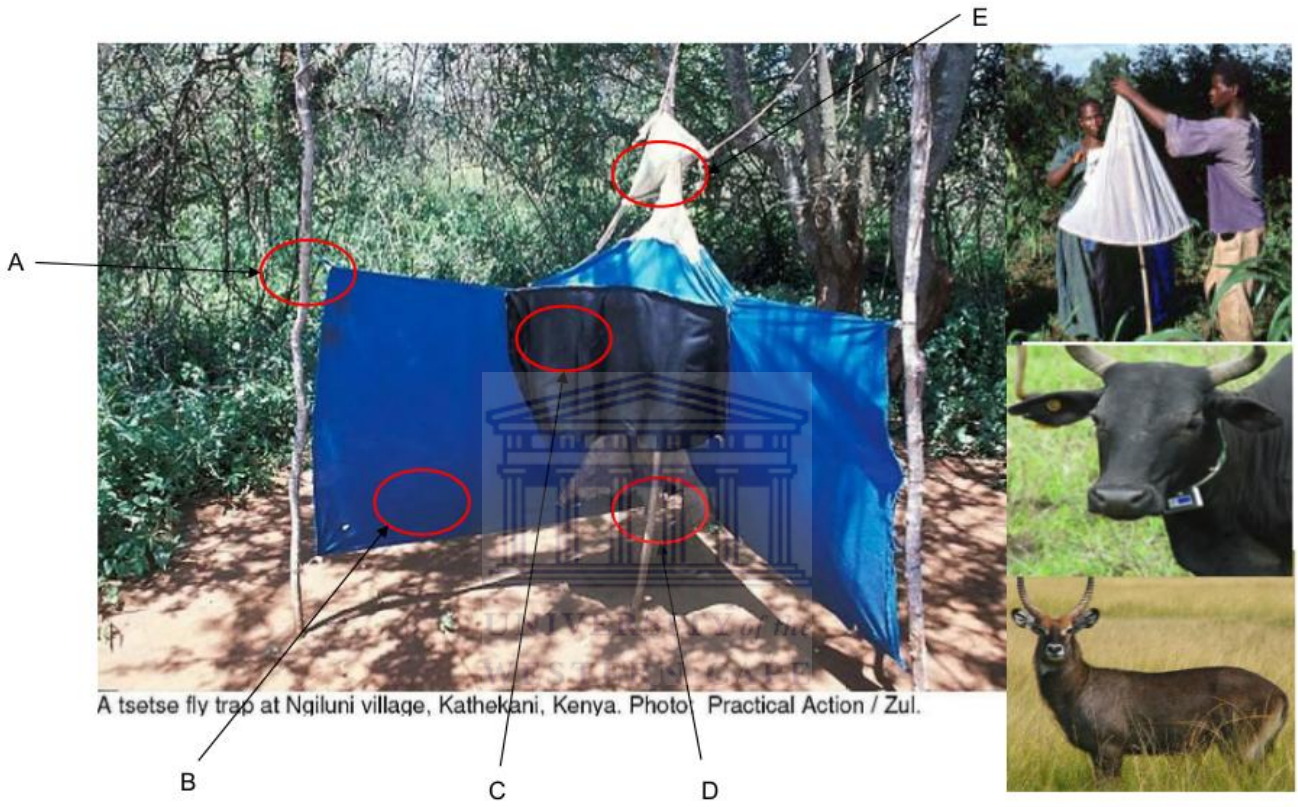


Figure 1.3 Tsetse lure traps and targets laced with odor attractants.
A – straight wooden sticks used to erect the trap-target; B – blue cloth target background that attracts flies from a distance; C – black cloth simulating host under flanks where flies prefer to suck from, enhance landing; D – odor attractant (natural host urine or artificial attractive odors); E – netted clothing trap to which the fly is lured by light. **Insets**: **to**p – an alternative trap; **middl**e - cattle collar trap laced with repellent smell extract from waterbuck, **botto**m - the waterbuck, *Kobus defass*a. (Photos sourced from http://en.howtopedia.org/wiki/File:Tsetse_fly_management_img_26.jpg; and www.one.org/International/blog/battling-the-tsetse-fly-in-Kenya/).

1.2.5 Chemosensory receptors and related soluble proteins

As in vertebrates, insect ion channel trans-membrane proteins can be classified into two broad categories: ionotropic receptors and metabotropic receptors. Ionotropic receptors are those that open ion (anions or cations) channels directly

upon some extraneous activation, resulting into rapid distortions of membrane protein conformations. On the other hand, metabotropic receptors are those that upon initial extracellular activation, trigger underlying intracellular surface bound proteins in succession to open specific types of channels, causing a time lapse in membrane protein conformational changes (Lodish *et al.*, 2000). Recently, the insect chemosensory receptors expressed in peripheral organs (wings, legs, antennae, proboscis, maxillary palps, cerci) interacting directly with the environment, were collectively renamed as ionotropic ion channel receptors, which include non-glutamate odorant (ORs) and gustatory (GRs) receptors, and glutamate-gated ionotropic receptors, IRs (Wicher, 2013). The IRs are variants of a highly conserved family, ionotropic glutamate-gated receptors (iGluRs) that are expressed in CNS (Benton *et al.*, 2009). Some ionotropic ion channel receptors require co-expression of sensory neuron membrane proteins (SNMPs), which are variants of CD36-like scavenger receptors (Benton *et al.*, 2007; Xin *et al.*, 2008; Vogt, 2009). The SNMPs are implicated in mating odor (pheromone-like) detections. To function, some ionotropic receptors need mediation by soluble proteins, amongst them odorant binding proteins, OBPs (some are also called pheromone binding proteins, PBPs), and chemosensory-specific proteins, CSPs.

The metabotropic receptors are found along the diverse sensory pathways, expressed in the insect CNS. These include the largest superfamily of G-protein coupled receptors (GPCRs), which have diverse ligands such as light, peptides, hormones, and neurotransmitters. In insects, fast neural transmission in CNS is mediated by iGluRs that use amino acid glutamate as its neurotransmitter ligand. Also, class C or subfamily 3 GPCRs contain neurotransmitter gated members, amongst them metabotropic glutamate receptors, mGluRs that modulate neurotoxicity of iGluRs action (Niswender & Conn, 2010).

1.2.5.1 Insect Non-glutamate chemoreceptors - ORs and GRs

Amongst insects, the Diptera have a simple but robust chemoreceptor system with higher specificity and sensitivity for detecting a complex mix of semiochemicals at varied concentrations and for varied lengths of time compared to vertebrates (Buck & Axel, 1991; Hallem *et al.*, 2006; Hallem & Carlson, 2006; Ha & Smith, 2008; Fuss & Ray, 2009). Presently, the insect non-glutamate ion channels, ORs and GRs, have been well described as distinct peripheral chemoreceptor super-family, highly conserved at functional and structural levels but highly divergent at nucleotide sequence level (Ishimoto & Tanimura, 2004; Dahanukar *et al.*, 2005; Ghaninia *et al.*, 2007; de Bruyne & Baker, 2008; Ha & Smith, 2008; Sanchez-Gracia *et al.*, 2009; Galizia & Rössler, 2010).

The insect ORs play crucial roles in blood-meal host and mate finding. In particular, the drosophila Or67d has been established as a male-pheromone (Z)-11-*cis*-vaccenyl acetate (cVA) binder, in conjunction with other proteins like SNMPs and LUSH proteins (discussed below). Another drosophila receptor, Or65a, has also been implicated in male mating responses; though the response is inhibitory, its ligand is yet unclear (Dahanukar & Ray, 2011). The OR gene subfamily is enormous and highly divergent in both DNA sequence similarity, residue number, and identity across eukarya genomes (Benton, 2006; Benton *et al.*, 2006; Fuss & Ray, 2009). Comparatively, humans have about 800 OR gene loci of which 400 are pseudo-genes, while mice encode about 1200 genes (Niimura & Nei, 2005). In insects, *D. melanongaster* has 60 OR gene loci encoding 62 proteins by alternative splicing (Clyne *et al.*, 1999; Vosshall *et al.*, 1999; Benton, 2006; Ha *et al.*, 2008; Sanchez-Gracia *et al.*, 2009), the malaria vector *An. gambiae* possess 79 OR proteins from 52 loci (Fox *et al.*, 2001; Hill *et al.*, 2002), the yellow fever and dengue virus vector *Ae. aegypti* has 131 OR genes (Bohbot *et al.*, 2007), while the filaria worm and West Nile encephalitis vector, *Culex quinquefasciatus*, encode 180 candidate OR loci (Arensburger *et al.*, 2010). These genes encode proteins with very low sequence similarity (Benton, 2006), and are structurally clustered in groups of 2-9, scattered in the genomes (Hallem *et al.*, 2006; Sanchez-Gracia *et al.*, 2009). Comparatively within and between insect lineages, OR sequences show low sequence conservation, except for the relatively conserved intron/exon boundary motifs, suggesting they probably arose from a common ancestor (Robertson *et al.*, 2006). In addition, the OR gene family have many gene duplication events in a species-specific trend, generating expanded lineages in specific insect lineages (Sanchez-Gracia *et al.*, 2009). Despite high sequence divergence, insect ORs share conserved sequence domains. Such domains include the insect annotated seven trans-membrane, 7tm6_olfct_rcpt (PF02949) (Buck & Axel, 1991), and a set of three C-terminal structural domains whose function are not yet established (Miller & Tu, 2008).

The insect ORs are a rapidly evolving gene subfamily, perhaps due to the diverse habitats insects inhabit. Consequently, identification of these ORs becomes a daunting task because of very low sequence similarity (ranging from 14% to 70%) between any pair of sequences, be they paralogs or orthologs. In addition, this makes it harder to identify novel genes in new genomes for which there are no cloned receptor gene sequences to be used as probes (Gao & Chess, 1999). Furthermore, it is difficult to incorporate this diverse set of receptors into designing possible control tools applicable to a wide range of disease-vector species and sub-species (Carey & Carlson, 2011).

Yet again, mammals encode higher numbers of GRs than insects (Nozawa & Nei, 2007). The fruit fly, *D. melanogaster*, has about 73 GRs encoded by alternative splicing of 60 genes, while *An. gambiae* reportedly encodes 76 GR-

like loci for sugars, water and salts (Ishimoto & Tanimura, 2004; Sanchez-Gracia *et al.*, 2009). The insect GRs are expressed on sensory dendrites found in proboscis sensillia (bristles) and labial palps (Stocker, 1994), maxillary palps (Hansson, 1995), the legs, wings, and antennal dendrites (Silbering & Benton, 2010; de Brito-Chanchez & Giurfa, 2011). Most GRs are expressed on true gustatory sensory neurons (GSNs) linked directly to higher brain site, sub-esophageal ganglion (SOG) (Sanchez-Gracia *et al.*, 2009; Galizia *et al.*, 2010; Montell, 2010; Silbering & Benton, 2010). The *D. melanogaster* have exceptionally expanded species-specific GR receptors (Gr5a and Gr64f) sensitive to yeast sugar, trehalose (a disaccharide, glucose+glucose) (Ishimoto & Tanimura, 2004; Ebbs & Amrein, 2007). These are useful for sampling suitable fruit juices (de Brito-Chanchez & Giurfa, 2011). Some gustatory bristle receptors (e.g. *D. melanogaster* Gr21a and Gr63a) specialize in detecting CO_2, eliciting aversive response, but are expressed on the antennae (Montell, 2010; de Brito-Chanchez & Giurfa, 2011). Further, some GRs act as contact sensors for con-specific cuticular hydrocarbons that also regulate reproductive behaviors, e.g. Gr68a, Gr32a, Gr33a and Gr39a in drosophila (Gardiner *et al.*, 2007). For instance, activation of Gr68a elicits appetitive mating behavior while activation of Gr32a and Gr33a elicits refractory behaviors (Dahanukar & Ray, 2011). In many insect orders, the expression, chemical-electrical signal conversions, and processing of chemosensory stimuli in higher brain sites take sexually dimorphic patterns (Hansson, 1995; de Brito-Chanchez & Giurfa, 2011). Many GR gene loci in diptera genomes not only lack homologs in a wide number of other insects, but also encode gene-lineage expanded repertoires, corresponding to the diverse diet tastes in insects (Montell, 2010). In tsetse fly *G. m. morsitans*, a diptera, the repertoires of ORs and GRs should stand out peculiarly in line with their discordant biology compared to other dipteral relatives.

1.2.5.2 Molecular mechanisms of Diptera ORs and GRs

Based on reports from the model fly, *D. melanogaster*, the ORs and GRs are related members of a multi-gene family of about 370 - 400 amino acids, and are spatially expressed on OSNs and GSNs dendrites respectively that extend into the sensillar lymph (Zwiebel & Takken, 2004; Dahanukar *et al.*, 2005; Hallem *et al.*, 2006; Ghaninia *et al.*, 2007; de Bruyne & Baker, 2008; Fuss & Ray, 2009; Galizia *et al.*, 2010). The receptors are predicted with heptatopic membrane domains, but share no sequence similarity except at the C-terminus that coincides with the seventh trans-membrane domain (Robertson *et al.*, 2003). At molecular level, the GRs are more conserved in sequence and structure, and appear to be more ancient than ORs (McBride *et al.*, 2007; Gardiner *et al.*, 2008), meaning the ORs have higher sequence diversity within and between

species (Benton, 2006; Vosshal & Stocker, 2007). Both ORs and GRs are thought to have evolved as parallel chemoreceptors across diverse organisms, accommodated by distinct neurons that send their perturbations to distinct separate brain sites (Ishimoto & Tanimura, 2004). Despite being related, the insect ORs and GRs in general share no obvious primary sequence identity to neither their mammalian counterparts nor to known GPCRs, thus revealing that they are a completely different family operating via an atypical molecular mechanism (Benton *et al.*, 2006; de Bruyne & Baker, 2008; Ha *et al.*, 2008; Nakagawa *et al.*, 2009; Fan *et al.*, 2010). However, unlike the GRs, insect ORs are clearly known to function as heterodimers, and their N-termini are unusually located on the cytoplasmic side of the cell membranes (Fan *et al.*, 2010; Galizia *et al.*, 2010). This contrasts the GPCR-type hetero-multimeric complexing of OR mediating odor responses in vertebrates, whose N-termini are extracellular (Nakagawa *et al.*, 2009; Pellegrino & Nakagawa, 2009).

Unlike GRs, the co-expression of ORs by forming a hetero-dimeric union between a specific OR and a non-canonical partner, Or83b (Orco), has been unique in insects. The Orco gene is the most conserved across divergent insect species and when jointly expressed, it enhances sensitivity of specific OR partners to odors. The Orco also chaperons the specific ORs to their various odorant sensory neuron (OSN) dendrites (Krieger *et al.*, 2002; Ha *et al.*, 2008; Sato *et al.*, 2008; Wicher *et al.*, 2008; Nakagawa *et al.*, 2009; Liu *et al.*, 2010). In fact, the individual Orco-mutant OSNs or their homologs in other species where ORs are expressed singly, have shown widespread defects in both odorant detection and trafficking or tuning of other ORs to the OSNs dendrites (Nakagawa *et al.*, 2009), despite retaining ability to induce spontaneous odor sensitivity (Dahanukar *et al.*, 2005; Ha *et al.*, 2008). Separate research groups of Sato and Wicher re-classified them as ligand-gated ion channels (peripheral ionotropic receptors) distinct from other 7-trans-membrane GPCRs (Sato *et al.*, 2008; Wicher *et al.*, 2008), but Wicher's group observed that the Orco partner may utilize a downstream metabotropic cascades to regulate the ORx partners (Wicher *et al.*, 2008; 2013). Nevertheless, the almost universal co-expression of Orco/ORx confers non-selective cation channels permeability to Na^+, K^+, and Ca^{2+} (Ha *et al.*, 2008), thus inducing membrane depolarization, for fast activation speed that the insects need. In all, the Orco orthologs are always found because of its high sequence similarity across the insect lineages when searched in public genome databases and EST libraries (Nakagawa *et al.*, 2009; Liu *et al.*, 2010; Carey & Carlson, 2011; Wicher *et al.*, 2013).

1.2.5.3 Insect Glutamate-gated ion channels

1.2.5.3.1 Neurotransmitter glutamate

Glutamate is a ubiquitous amino acid in all living systems, and participates in crucial biochemical processes in two main ways: one, nutritionally providing energy to various cells; and two, neurally acting as an excitatory post-synaptic membrane neurotransmitter. After their neuronal role discovery over half a century ago (Julio-Pieper *et al.*, 2011; and articles therein), the glutamate is established as the most abundant excitatory neurotransmitter, binding onto both the ionotropic and metabotropic synaptic neurotransmitter receptors (Danysz *et al.*, 1995; Kugaya & Sanacora, 2005). By this they help in transmitting peripheral chemical signals to the higher brain sites. Though the glutamate may cyclically move through different pools, the ones engaged in direct neuronal signal transmission across the synapse are the synaptic vesicle pool (Danysz *et al.*, 1995; Gegelashvili *et al.*, 2001; Kugaya & Sanacora, 2005). From the synaptic vesicles, the glutamate is released at the synapse when an incoming impulse (in our case may be the olfactory information) depolarizes the synaptic knob terminals via Ca^{2+} influx and subsequently fuses the synaptic vesicle membrane to the neuron membrane (Broman *et al.* 2000; Gegelashvili *et al.*, 2001). Once released, the glutamate shuttle across the gap to bind post-synaptic membrane excitatory neurotransmitter iGluRs and/or modulatory mGluRs to depolarize and initiate a new impulse. Within split second, the synaptic glutamate is cleared-off via re-uptake by pre-synaptic membrane Na+/K+ ion channels, excitatory amino acid transporter (EAAT), and ultimately into the vesicles again via proton-pump ATP-Mg^{2+} dependent synaptic vesicle transporter. The other clearance route is the glutaminase biosynthetic pathway via ion channels of the astroglial cells (Gegelashvili *et al.*, 2001; Kugaya & Sanacora, 2005; Julio-Pieper *et al.*, 2011).

1.2.5.3.2 Glutamate-gated ion receptors – IRs, iGluRs and mGluRs

In general, glutamate-gated ion channels are of two kinds: (1) directly gated IRs and iGluRs, and (2) indirectly gated mGluRs. Unlike the vertebrate IRs that are expressed in the CNS post-synaptic membranes, the insect IRs are expressed on the PNS dendrites (Hansson & Stensmyr, 2011). After they were discovered to be expressed on the antennal sensillae and confirmed to participate in detection of odors (Benton *et al.*, 2009), insect IRs have continued to attract much research attention. They are mainly expressed on the coeloconia sensilla sensory neurons, where other receptors are rarely found, but participate in detecting odorants as ORs do. Unlike ORs, the IRs belong to the super gene family of glutamate-gated ion

channels, and are divergent variants of iGluRs (Benton *et al.*, 2009). In addition, the insect IRs seem to be more evolutionary conserved than ORs, and potentially more sensitive to detecting general odorants (e.g. abundant amine odors) implicated in common physiological functions across many insect species (Benton *et al.*, 2009; Croset *et al.*, 2010; Hansson & Stensmyr, 2011). While the ORs form membrane channel by a union of two subunits with a mandatory canonical partner, Orco, the IRs form hetero-multimeric ion channels of at least three partners with different members acting as co-receptors (Benton *et al.*, 2009). The iGluRs are expressed on the CNS synaptic membranes; they are similar to those found in vertebrates, and are conserved across diverse Protostome lineages (Benton *et al.*, 2009; Croset *et al.*, 2009; Hansson & Stensmyr, 2011).

Both IRs and iGluRs are characterized by glutamate ligand-gated ion channels with conserved domains (Benton *et al.*, 2009; Croset *et al.*, 2010; Olivier *et al.*, 2010). The domains comprise extracellular N terminus ligand binding domains (LBDs) lobes S1 and S2, trans-membrane ion channel domains, and a cytoplasmic C-terminal. The ion channel is made up of four trans-membrane domains – three complete trans-membrane domains and an intervening loop re-entrant domain between trans-membrane one and two (Benton *et al.*, 2009). The LBD S1 lobe has an arginine (R) residue that binds glutamate α-carboxyl group, the first half of LBD S2 lobe has threonine (T) residue that binds glutamate γ-carboxyl group and in the second half a conserved aspartate (D) or glutamate (E) that interacts with glutamate α-amino group. These ligand interacting residues are highly conserved in the iGluRs than in the IRs, except in IR8a and IR25a (Naur *et al.*, 2007). Despite the iGluRs and IRs being similar in terms of 4-transmembrane channel domains, they differ with the former having an extensive N-terminal with an amine-terminal domain (ATD) and venus fly-trap-like metabotropic ligand-binding domain sites, LBDs, S1 and S2 (Traynelis *et al.*, 2010). With exception of IR8a and IR25a, the IRs exhibit shorter N-termini, ~200 amino acids, and often lack at least one of the ligand-interacting residues (Benton *et al.*, 2009; Tikhonov & Magazanik, 2009).

At the sensory dendrites, signal transduction takes place after odors bind onto IRs, thereby directly gating the trans-membrane cationic channels. As the impulse travels to the antennal lobe and onto the mushroom bodies (Masse *et al.*, 2009; Martin *et al.*, 2011), the synaptic knob glutamate-containing vesicles are triggered to move towards the pre-synaptic membrane, whence the contents are released into the synapse, and are picked by the post-synaptic membrane receptors. At the synaptic gaps and neuromuscular junctions, there are various ionotropic and metabotropic receptor sub-types of iGluRs, excitatory mGluRs, and inhibitory gamma aminobutyric acid (GABA) (Danysz *et al.*, 1995; Kugaya & Saracona, 2005;

Lehmann *et al.*, 2012). Therefore, glutamate receptors, IRs, iGluRs and mGluRs, may jointly mediate excitatory neuronal activities, and knowing their repertoires may inform how the odor plumes are detected, sensed, processed, perceived and pursued by the tsetse fly to locate their obligate vertebrate hosts.

1.2.5.4 Chemosensory-related proteins – OBPs CSPs and SNMPs

The insect ORs bind hydrophilic odors directly and hydrophobic odors indirectly. In the latter, they require mediation of soluble proteins to bind and transport the hydrophobic odors across the hydrophilic sensilla fluid. For instance, detection of the only known male fruit fly specific pheromone cVA, requires Or67d, extracellular odorant binding protein (also LUSH protein), and a trans-membrane, SNMP (Xin *et al.*, 2008). For the Or67d to reach an endogenous optimal sensitivity to detect the cVA, it requires co-expression of both LUSH protein and SNMP on Or67d-expressing neurons on trichoid sensilli. LUSH, a soluble protein, is necessary in binding and presenting cVA to the Or67d (Xu *et al.*, 2005; Benton *et al.*, 2007; Jin *et al.*, 2008; Dahanukar & Ray *et al.*, 2011). By homology to the vertebrate scavenger protein, the SNMP is thought to be involved in lipoprotein transportation and neuronal signal transduction – the latter process links it to perception of pheromone signals on the antennal trichoid neurons (Benton *et al.*, 2007; Smith, 2007; Xin *et al.*, 2008; Vogt, 2009). Detection of cVA display sexually dimorphic behaviors amongst fly males and females, e.g. it is appetitive for female fruit flies, stimulating mating responses, while it also yields inhibitory responses to courting mated females in male flies. The SNMP in itself enhances mating in the presence of cVA, but also inhibits mating behavior in the absence of cVA (Xin *et al.*, 2008; Vogt, 2009). In fruit flies, mosquitoes and honey bees, the number of annotated SNMPs stands at two. In this respect, the identification of SNMP homologs in *G. m. morsitans* genome gives a clue to the likely underlying genes whose products interact to regulate mating behavior in the tsetse fly.

The OBPs are implicated in binding and solubilizing hydrophobic odorants such as pheromones, and other general odorants, and presenting them to specific chemosensory receptors. They are also critical in defining the molecular interaction between the chemoreceptors and different odors (Sanchez-Gracia *et al.*, 2009). However, the OBPs also function in non-chemosensory physiologies, evidenced by their expression in other tissues that are not linked to olfaction and gustation (Pelosi *et al.*, 2006). There are 52 and 82 OBPs annotated in *D. melanogaster,* and *An. gambiae* respectively, and the numbers vary in other insect species as well (Foret & Maleszka, 2006; McBride & Arguello, 2007). The insect OBPs have been subdivided into four groups depending on the primary amino acid sequence conservation of cysteine

residues. Group one, classical OBPs with an average molecular weight 14-20kDa, contain exactly 6 conserved cysteine sites; group two, plus-C OBPs, have2 or 3 additional cysteine sites and at least one highly conserved proline site; group three, minus C OBPs have less than 6 cysteine sites; and group four, atypical OBPs have at least 6 cysteine conserved sites with a longer C-terminus (Liu *et al*., 2010; Zhou *et al* 2010). Recently, analysis of *G. m. morsitans* EST transcriptome library revealed presence of 20 OBPs, of which eight were highly expressed in the antennae, but with reduced expression level 72hours post blood-meal. The important finding was that all the predicted OBPs were either of the Classic group or the Plus-C group, and are expressed in chemosensory organ, antennae (Liu *et al*., 2010).

Drosophila melanogaster and *An. gambiae* encode four and eight copies of CSP genes respectively, and are known to be expressed in the chemosensory sensillar fluid (Foret & Maleszka, 2006; McBride & Arguello, 2007; Gardiner *et al*., 2008). Like OBPs, CSPs also have conserved disulfide bridges. While the OBPs form three bridges from six neighboring conserved cysteine sites, the CSPs form two alpha-helical disulfide bridges from four conserved cysteine sites (Angeli *et al*., 1999). Five CSPs (*GmmCSP1-5)* were recently reported in *G. m. morsitans* EST library, with q-PCR quantification revealing *GmmCSP1* and *GmmCSP3* expressed in tissues other than antennae (Liu *et al*., 2012). The analysis also compared expression in males and females post blood meals, and revealed that *GmmCSP2* get down-regulated hours after blood meal in females, meaning it could be critical in blood-meal host finding in starving females (Liu *et al*., 2012). Because OBPs are highly diverse compared to CSPs across different insect species, the genomic annotation approach in this report, presented a unique opportunity to recover any putative OBP and CSP in the *G. m. morsitans* genome assembly GMOY1.1.

In summary, the repertoires and functional characterization of chemosensory receptors and related proteins have been studied in a number of blood-sucking, fruit juice seeping, and social insects (Robertson *et al*., 2003; 2006; McBride *et al*., 2007; Benton *et al*., 2009; Olivier *et al*., 2010; Mentel, 2010; Hansson *et al*., 2013), but not in any tsetse fly species. The ability of tsetse fly, a dipteran blood feeder, to locate hosts, mates and avoid predators are well established (Krafsur, 2009; Lehane, 2005). However, the molecular description of the tsetse fly chemosensory system remains less understood. This thesis provides foundational information based on genomic description of chemosensory associated and receptor genes involved in olfaction and gustation – pivotal biological processes that can elucidate the peculiarities surrounding the tsetse life.

1.2.6 *Glossina m. morsitans* genome characterization: sequencing and assembly

Genome sequencing is the determination of an organism's linear pattern of DNA sequence – a process that yields enormous biological data. Basically, the genomic DNA sample is sheared into pieces, which are sequenced by amplification using new generation sequencing techniques to yield many redundant copies called short sequence contigs. The contigs are then assembled into supercontigs (scaffolds), which are indexed for convenience of gene identification process (Figure 1.4). This has been the noble task in the field of insect science ever since the completion of fruit fly *D. melanogaster* draft genome in the year 2000 (Adams *et al.*, 2000). Since then, a number of blood-feeding vectors genomes have been sequenced, including *An. gambiae* (Holt *et al.*, 2002), *An. gambiae* S, *An. gambiae* M (Lawniczak *et al.*, 2010), *Ae. aegypti* (Nene *et al.*, 2007), and *Culex quinguiefasciatus* (Arensburger *et al.*, 2010). These genomes facilitated the study ofinsect chemosensory related genes through identification of ORs, GRs, IRs, SNMPs, OBPs and CSPs. Other genomes whose sequencing have also been useful include social insects like honey bee, *Apis mellifera* (Robertson *et al.*, 2003); ants (Zhou *et al.*, 2010); and noctuid moth, *Spodoptera littoralis* (Olivier *et al.*, 2010).

The genome sequencing of tsetse fly *G. m. morsitans* was launched over ten years ago (Aksoy *et al.*, 2003), and its first genome assembly GMOY1.1 Yale strain annotations subsequently published (IGGI, 2014). Along with the publication of the genome, different cooperating laboratories round the globe released lots of related data to the International Glossina Genome initiative (IGGI) community, including whole transcriptome expression RNA-sequence data, derived from pregnant female (Benoit *et al.*, 2014), courtesy of Aksoy's lab, Yale University. These opened opportunities for active annotation of the genome with different laboratories spread across the globe tackling specific gene families part of the work reported in this thesis was part and subsequently published (IGGI, 2014). In addition, sequencing and assembly of other selected species of the genus *Glossina* are on-going (Aksoy *et al.*, 2005).

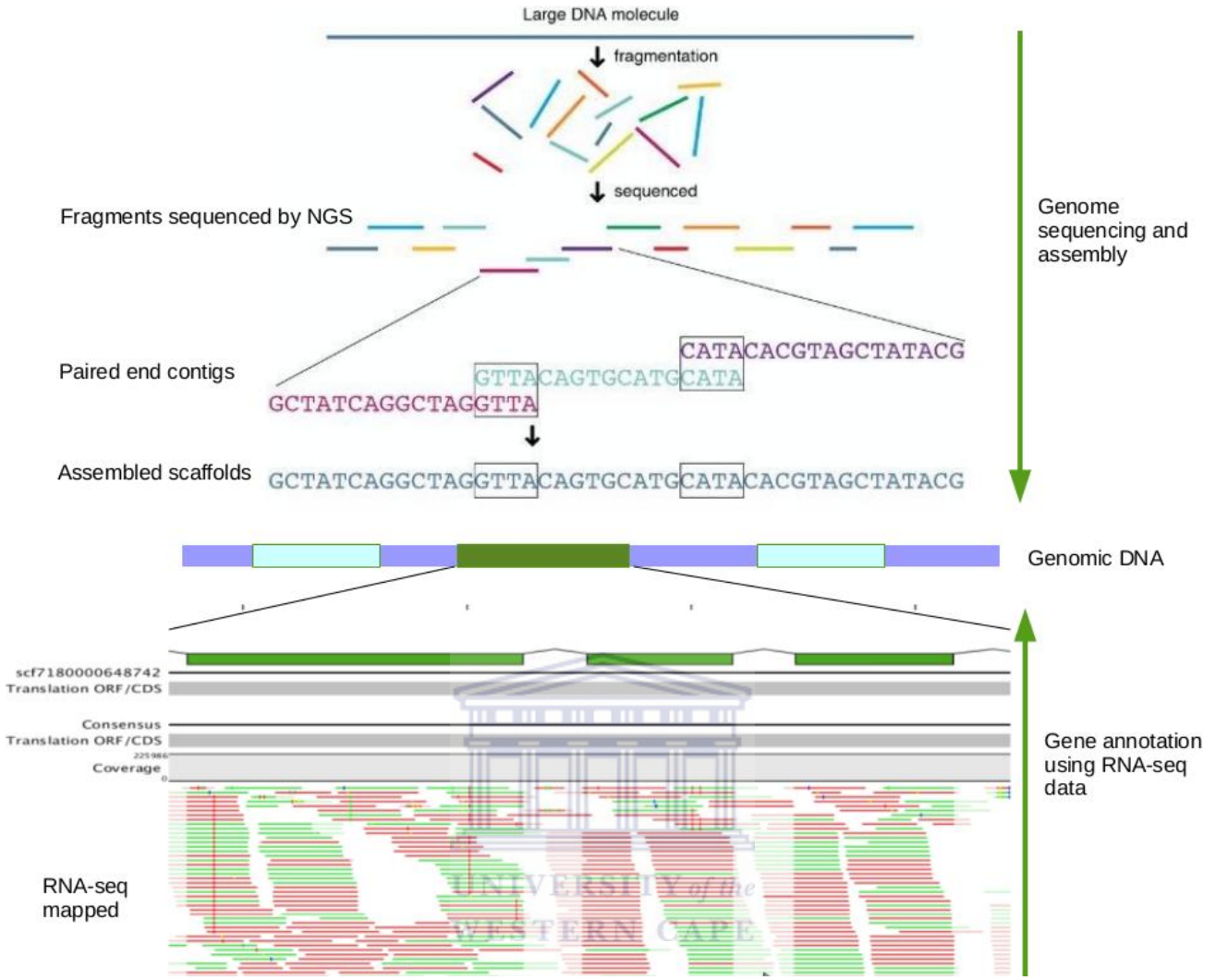


Figure 1.4 Summary of genome sequencing, assembly and annotation processes.
The extracted genomic DNA molecule is subjected to differential fragmentation by specific digestive enzymes that cut the sequence at precisely known sites, and then the fragments are sequenced by next generation sequencing (NGS) techniques. The sequenced products (contigs) are then assembled by end-matching sequences to form longer supercontigs that are again end-matched to form scaffolds. The process direction is shown by the downward-pointing green arrow. The annotation of gene regions (two pale and green blocks on light blue background block representing genomic DNA) is aided by mapping RNA-seq onto the DNA. RNA-seq junction read alignments (at bottom - red and green bars of mapped RNA-seq) help define exon/intron boundaries of the gene of interest (three consecutive green bars with linked lines)

1.2.7 Gene finding and annotation

Gene identification entails determination of gene constructs, i.e. protein coding regions - exons, open reading frames (ORFs) and non-coding sequences (introns and regulatory regions) in an assembled DNA sequence. Gene identification includes characterization of gene coding start codons, stop codons, splice sites and frame-shift constructs; to defining

specific location of the gene within the genome and assigning it a specific feature coordinate relative to all other genes (Birney *et al.*, 2004; Curwen *et al.*, 2004). Overall, there are three approaches in gene finding: (1) *content-based* relying on inherent properties of the sequence itself e.g. codon usage, coding potential, repeat periodicity, and compositional complexities; (2) *site-based* (or inherent signals) focusing on presence or absence specific sequences, patterns or consensus such as donor and acceptor sites, poly-adenylation (poly-A) tracts, transcriptional factor binding sites (TFBS), translational initiation codons, and start and stop codons; and (3) *comparative-based* strategy that searches for homology to previously characterized protein coding regions that matches the query sequence (Baxevanis & Quellette, 2001). The former two strategies are intrinsic, also referred to as *de novo* gene search, while the latter is described as an extrinsic approach. Gene finding remains a difficult task, as no single strategy can be used independently and different genomes require application of different identification pipelines. In addition, these strategies as implemented in various bioinformatic platforms are often twixt to use either nucleotides or amino acid sequences and trained for specific genome datasets.

Gene annotation on the other hand, involves determination of conserved functional domain sequences within an identified gene sequence, and making an attempt to attach a putative function to the gene. Characterizing the domains again follow two main approaches: first, comparative annotation where the query genome is annotated against a reference template genome whose domain features are known; and second, an *ab initio* approach in which the gene structures are computationally predicted and characterized from statistical and artificial intelligence methods (Baxevanis & Quellette, 2001). Many *ab initio* gene predictors have been developed that use purely genomic sequence for their gene structure prediction. Such gene prediction tools include Genie (Reese *et al.*, 2000), Genscan (Burge and Karlin, 1997), FGENEH/FGENES (Salamov & Solovyev, 2000), Geneid (Parra *et al.*, 2000), Augustus (Stanke *et al.*, 2008) among others; and genome viewers and analysis tools like Artemis genome viewer tool (Carver *et al.*, 1997). Results of these sequence homology searches and various *ab initio* gene prediction methods are crucial in manually exacting and supporting gene feature constructions (Curwen *et al.*, 2004). The process also rely so much on data retrieval browsers such as the University of California Santa Cruz (UCSC) browser; Gbrowser which is implemented in Vectorbase and Flybase browser sites; the NCBI Map Viewer; and the Ensembl (Birney *et al.*, 2004).

For the present work, the *G. m. morsitans*chemoreceptome including odorant, gustatory and glutamate-gated ion receptome annotation was done comparatively using mainly the *D. melanogaster* known annotated orthologs as queries (McQuilton *et al.*, 2012). Computational techniques included *de novo* approach via Core Eukaryotic Genes Mapping

Approach (CEGMA) (Parra *et al.*, 2007) and the MAKER approach (Cantarel *et al.*, 2008), jointly implemented by International Glossina Genome Initiative (IGGI) community (IGGI, 2014). These were supplemented by various standalone and web-based tools. The choice of the query genome was informed by the notion that Drosophila is evolutionarily closer to Glossina, and also as a model insect genome that has been relatively well annotated and most gene models experimentally validated.

1.2.8 Importance of understanding the molecular basis of tsetse fly chemosensory system

The insect heads play very crucial roles towards their survival instincts, which include detection of visual cues using a pair of compound and simple eyes, detection of volatiles, soluble and non-soluble chemicals using a pair of antennae and maxillary palps, and feeding on various types of food resources using specially adapted mouth parts. The visual sense help them to pick directions with respect to position of the sun, perceive light intensity and colors, decode both animate and inanimate shapes and their movements, and estimate their relative position with respect to horizon – a specialty of oceli (FAO, 2008). The pairs of antennae and maxillary palps represent the vertebrate noses and tongues respectively. The antennae help in detection of plumes of smell molecules including pheromones, which either excite or inhibit their physical behaviors. On the other hand, the maxillary palps help them to sample soluble/semi-soluble food substances as well as detecting of con-specific cuticular hydrocarbons by contact (FAO, 2008). These senses basically dictate how the tsetse flies communicate with each other – both at intra- and inter-species level, and with the environment as they navigate to find food, mates and resting sites.

The tsetse fly pair of antennae stands distinct amongst the hexapods with their bushy arista and distinct sensory hair pits at base of the arista. Each antenna bear sensory neuron hairs called sensillia (singular, sensillium) running underneath on the third segment (see Figure 1.1). On the surface of these neurons, as established in fruit flies, there are numerous chemoreceptor proteins expressed for detecting and decoding ecological odor volatiles and pheromones at varied sensitivities and specificities (Benton *et al.*, 2009; Carey & Carlson, 2011). The receptors include ORs and IRs. Because no study has clearly described the tsetse sensilla fine structures, it is assumed here that each tsetse sensory pit and aristal nerve hairs contain the receptors that make the interaction points between the tsetse PNS and the external turbulent odorous environment. On the maxillary palps and other organs, the sensory hair neurons express mainly GRs that help insects in detecting and perceiving soluble sweet tastes, and bitter salts (Robertson & Wanner, 2006; Sánchez-Gracia *et al.*, 2009).

Given that the earlier mentioned tsetse control approaches fail either *in toto* or in-part in different focal areas and the threats of nagana and sleeping sickness remain real, improvement of odor-baited technologies via understanding of the tsetse chemosensory system is an important approach (Serap, 2005). The molecular chemosensory approaches target two ends. Firstly, to provide the molecular basis of understanding the chemosensory biology of the tsetse flies as vectors of infectious agents, trypanosomes. Secondly, to give drive towards development of novel alternative tools for suppressing population of vectors and blocking trypanosome transmission. Knowledge of chemoreceptome will improve understanding of the unusual biology of tsetse, including strict blood-feeding, viviparous style of reproduction, and the observed low physical activity within restricted habitats. In particular, it will be interesting to link the diversity and complexity that tsetse flies display in detecting and selecting specific host odor cues (Kaufman *et al.*, 2002; Liu *et al.*, 2010) to their ability to transmit different trypanosomes.

1.3 Aims of the thesis research

This thesis aimed at identifying and characterizing putative chemosensory-related genes in *G. m. morsitans*. The olfactory, gustatory, glutamate-gated receptors and soluble chemosensory proteins were targeted because they are predicted to play a key role in host-seeking, blood-meal feeding and mate-finding in the *G. m. morsitans* by decoding odors from the mammalian hosts, vegetation, and members of same species respectively. Therefore, the identification, and analysis of molecular diversity of the whole range of chemosensory related genes in this obligate hematophagous vector is the subject of this thesis.

1.3.1 Objectives

1.3.1.1 Main Objective

The study was to characterize the repertoire of chemoreceptor, glutamate-gated receptor, and chemosensory associated genes in *G. m. morsitans* genome using computational approaches. The chemoreceptors are odorant receptors (ORs) and gustatory receptors (GRs); the glutamate-gated ion receptors include ionotropic receptors (IRs), ionotropic glutamate receptors (iGluRs) and metabotropic glutamate receptors (mGluRs); and chemosensory associated genes are sensory neuron membrane proteins (SNMPs), soluble odorant binding proteins (OBPs) and chemosensory specific proteins (CSPs).

1.3.1.2 Specific Objectives

(i) To annotate putative non-glutamate gated chemoreceptors (ORs and GRs), glutamate-gated receptors (IRs, iGluRs and mGluRs), and related chemosensory genes (SNMP, OBP and CSP) in *G. m. morsitans* genome.

(ii) To determine the expression profiles of the putative OR, GR, IR, iGluR, mGluR, SNMP, OBP and CSP genes in *G. m. morsitans.*

(iii) To reconstruct the phylogenetic relationships between and among OR, GR, IR, iGluR, mGluR, SNMP, OBP and CSP in and between G. *m. morsitans* and selected Diptera.

1.3.2 Research Rationale

About 60 million people in 38 countries in tropical sub-Saharan Africa reportedly live at risk of human African trypanosomiasis (HAT), with over 2 million direct livestock loss (Kioy *et al.*, 2004; FAO, 2008; WHO, 2000; 2010) and 1.5 million daily-adjusted life years (DALYs) (WHO, 2002). Animal African trypanosomiasis, AAT, is a devastating disease in livestock, rendering huge economic losses to the livestock industry in Africa (Gooding & Krafsur, 2005; Krafsur, 2009). The principal causal agent is *Trypanosoma brucei* transmitted by *G. m. morsitans* tsetse fly (Aksoy, 2003; Pittendrigh *et al.*, 2006). Numerous efforts to control tsetse vector population and incidences of trypanosomiasis require multi-faced approaches applicable across different tsetse species in Africa. Recent upsurge of HAT and AAT in the past one and a half decade occasioned by lapse in surveillance and treatment regimes, calls for urgent alternative approaches to manage the tsetse flies. Tsetse flies use olfaction to decode diverse semiochemicals to find blood-meal hosts, mates, breeding sites, predators and facilitate their social-communication and coordination (Fuss & Ray, 2009). Fully integrating odor-baited technologies into vector traps and target lures require clear understanding of underlying genomic molecular basis of decoding odors and tastes and processing of the signals within nervous system to initiate behavioral response in the fly.

There has been increased research interest in insect chemoreception, particularly the detection mechanisms of semiochemicals involving the ionotropic receptors (inclusive of ORs, GRs and IRs), and modulating proteins such as SNMPs, soluble OBPs/PBPs, and CSPs amongst the blood-feeding vectors of infectious diseases. Electro-physiological and *in vitro* experiments on peripheral chemosensory systems have revealed that most insect chemoreceptors are expressed on maxillary palpi and antennae (de Bruyne & Baker, 2008; Fuss & Ray, 2009), and labial palps, proboscis, wings, legs and ovipositor (Hansson, 1995; de Brito-Sanchez & Giurla, 2011). Majority of these studies have targeted specific genes or

gene families involved in both olfaction and gustation across diverse insect lineages, but molecular description of chemoreceptors in tsetse have been lacking. This study focused on *G. m. morsitans* Westwood 1850, the major vector of *T. b. rhodesiense*, *G. m. morsitans* Westwood 1850. Based on the newly available *G. m. morsitans* genome data, this thesis determined the molecular characteristics of the chemoreceptor and related genes involved in olfaction and gustation. Unlike other blood-feeding genera, the molecular diversities of the genes that encode chemosensory receptors were unknown in tsetse flies until now. The findings will provide a better understanding of molecular diversity in chemosensory genes mediating host and mate selection in the fly, and potentially find application in improving existing technologies, increase drive towards options for designing new control tools to trap and/or target the tsetse field populations for elimination. Additionally, they will provide foundation for future comparative and functional studies.

CHAPTER TWO

MATERIALS AND METHODS

This chapter describes the materials and methods used to undertake the study. The methods described can be supplemented by the adjoining publication manuscript (Obiero *et al.,* 2014) that emanated from the same research, given in list of publications above. The description applies to all the gene classes reported here (the *G. m. morsitans* ORs, GRs, SNMPs, CSPs, OBPs, IRs, iGluRs and mGluRs), unless otherwise stated for a specific class of genes. It covers the procedures for gene identification, gene naming confirmatory bioinformatics analyses, ontology functional classification and phylogenetic analyses, and expression profile analyses.

2.1 Identification of chemoreceptors, glutamate-gated receptors, and chemosensory associated genes in *G. m. morsitans*

Annotated protein coding sequences (CDS) of ORs, GRs, SNMPs, CSPs, OBPs, IRs, iGluRs and mGluRs of *D. melanogaster* (Adams *et al.,* 2000) from FlyBase5.13 (McQuilton *et al.*, 2012), supplemented with those of *An. gambiae* PEST strain (Holt *et al.,* 2002), and *Ae. aegypti* (Nene *et al.,* 2007) from vector-base (Lawson *et al.*, 2009; Megy *et al.,* 2012), were used to search for orthologs in the Yale strain of *G. m. morsitans* genome and transcriptome (GMOY1.1) assembly. The genome assembly was searched using the TBLASTX algorithm (Altschul & Lipman, 1990). The positive blast hits from the three homologous searches (*Drosophila*, *Anopheles* and *Aedes*) were merged. Scaffolds containing best blast hits at cut-off e-value below 1e-03 were retrieved, and the hit regions curated as putative seed gene constructs in their respective scaffolds using CLC Genomics workbench suite version 4.8 build 48014 (CLC bio, 2012). These scaffolds with seeded gene constructs were used as genomic references, onto which 84 million illumina RNA sequence reads (derived from pooled samples of female flies) were mapped using the next generation sequence (NGS) analysis pipeline in CLC workbench. Complete gene models were then manually edited by searching for intron/exon boundaries, and detecting any possible alternative splice variants within the curated gene models. The gene models were then viewed and edited to fit into correct reading frames using Artemis genome viewer tool v13.2.12 (Carver *et al.,* 2012), confirming the exon/intron boundaries using motifs GT to mark 5' end donor site and AG marking 3' end acceptor site of introns. In all gene models, start codon ATG was fixed at 5' end and terminated at any of the three stop codons (TAA, TGA, or TAG). The putative *Glossina* gene model sequences were used to iteratively blast (e-value 1e-03) the same *Glossina* genome assembly at

vector-base to exhaustively recover any similar hit sequences. The gene models were matched with the *Glossina* automated gene feature predictions that were availed for the *Glossina* community annotation use (Megy *et al.*, 2012). The automated gene feature identities were used to assign the recovered chemosensory genes in *Glossina* identities. The recovered gene models that did not match any automated feature were assigned in-house temporary annotator identities, and their features manually created using artemis tool gene builder (Carver *et al.*, 2008). Both nucleotide and amino acid sequences of the models were retrieved for downstream analysis.

2.2 Bioinformatic validation analyses

The annotated *G. m. morsitans* gene models were confirmed as complete genes by searching for homologous sequences from non-redundant Swiss-Prot databases (Bairoch & Apweiler, 2000) using amino acid sequences of the models via DELTA BLAST (Boratyn *et al.*, 2012). The *G. m. morsitans* gene models were then assigned related homolog gene names accordingly. Completeness of each gene model (OR, GR, IR, iGluR, mGluR, OBP, CSP and SNMP) was assessed by searching for specific conserved domains from the ncbi conserved domains databases (CDD) (Marchler-Bauer *et al.*, 2011) and Pfam database (Bateman *et al.*, 2004. For instance, 7tm-6olf-recpt and 7tm-7olf-recpt domains were searched to confirm *G. m. morsitans* ORs and GRs respectively; ligand-binding domains (LBD), ligand-gated ion channel (combined Pfam PF00060 and PF10613) domain and amino terminal domain (ATD) (Pfam PF01094) were confirmatory domains for *G. m. morsitans* IRs and iGluRs; and GPCR family 3 (IPR0000337) and/or predicted metabotropic glutamate receptor (IPR000162) domains were confirmations for *G. m. morsitans* mGluRs. In addition to conserved amino acid cysteine signatures, the *G. m. morsitans* OBPs were checked for the presence of the general odorant binding protein domain (GOBP, pfam01395), which closely associates with insect pheromone binding specific signature (smart00708), while in CSPs, the insect olfactory specific binding family signature, OS-D domain (pfam03392) was confirmed. The OBPs and CSPs were also examined for the presence of secretory peptide signal via web tool SignaP v4.1, with default settings (Petersen *et al.*, 2011). The *G. m. morsitans* CD36-like proteins were validated using the CD36 (pfam01130) domain signature, a highly conserved functional domain found in lipoproteins - scavenger receptor proteins of class B across many lineages. The sequences that lacked detectable or poorly conserved functional domains, and seemed incomplete were set as pseudo-genes. Presence of alpha helix trans-membrane domains for the receptor genes were predicted using TMHMM server v2.0 (Krogh *et al.*, 2001). Analysis of atypically conserved C-terminal motifs in *Glossina* OR and GR peptides was done using a

Multiple Em for Motif Elicitation (MEME) program v4.8.1 at MEME suite web server (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi). The MEME suite parameters were set to find one motif per sequence for a maximum of five different motif types, optimum width for each motif range 10-30, and program allowed to shuffle the sequence at 50 sites (Bailey and Elkan, 1994; Bailey, 2008; Bailey *et al.*, 2009). The conserved Gene Ontology (GO) database (Conesa *et al.*, 2005) was used to evaluate ontology functional classification in these gene models. Analyses were done via Blast2GO b2g_aug12 database web-server (http://www.geneontology.org).

2.3 Nomenclature of chemoreceptor and chemosensory related genes in *G. m. morsitans*

The *G. m. morsitans* genes and proteins were named and assigned symbols following the gene and protein nomenclature for *D. melanogaster*. For instance, GmmOR* and GmmGR* were adopted for naming *G. m. morsitans* odorant receptor and gustatory receptor genes respectively, but were assigned numerical identities in ascending order, represented by '*'. Already published names and symbols were adopted for naming *G. m. morsitans* OBPs (GmmOBP1-29), and CSPs (GmmCSP1-5). *Glossina m. morsitans* CD36-like genes were identified and named using their non-redundant database homolog names. GmmSNMP1 and SNMP2 were renamed from CD36-1 and CD36-2 respectively, while the rest were named as CD36-3 through to CD36-15. The *D. melanogaster* ortholog nomenclature for glutamate-gated ion channels was adapted to name the *G. m. morsitans* IRs, iGluRs, and mGluRs genes. In all sub-families, the names for genes were italicized (i.e. *GmmOR*, GmmGR*, GmmOBP*, GmmCSP*, GmmIR*, GmmiGluR*,* and *GmmmGluR*)* and those for peptides were not italicized (GmmOR*, GmmGR*, GmmOBP*, GmmCSP*, GmmIR*, GmmiGluR*, and GmmmGluR*). Names for iGluRs - NMDA receptors, AMPA receptors, and Kainate receptors were retained, and '*' used to represent specific identities of the different loci recovered. Alternative splice variants were denoted with gene name followed by a period (.) and a digit number, e.g *GmmiGluRIIA.1* and *GmmiGluRIIA.2*. Putative duplicates were represented by gene name, hyphen then a lower case letter, e.g. *GmmiGluR2-a* and *GmmiGluR2-b*. Community annotation identifiers were represented by temporary indices given as GMOYxxxxx, the 'xs' being numerical indices generated within each scaffold giving the order of the gene loci from automated predictions. However, identities for new gene models without automated identifiers were manually generated using Artemis tool gene builder window (Carver *et al.*, 2008) and assigned a temporary identity (eg TMP_OR* or TMP_GR*). The *G. m. morsitans* gene models as annotated were subsequently submitted to the vector-base community annotation portal for *G. m. morsitans* for integration into genome database (Megy *et al.*, 2012). Lists of annotated amino

acid coding sequences are provided in supplementary material as Dataset S1, S2 and S3.

2.4 Expression profiles of chemoreceptor, glutamate-gated receptor and chemosensory associated genes in *G. m. morsitans* using female fly genome-wide RNA-sequence reads

Expression levels of annotated genes were estimated using whole transcriptome RNA-sequence reads that were generated and pooled from laboratory reared female flies utilized at different pregnancy stages, kindly availed by the IGGI. The reads were processed as detailed in Benoit *et al.* (2014) and IGGI (2014). In total, 84 million illumina reads averaging 75 base pairs each were used to determine expression profiles of the individual genes annotated in *G. m. morsitans ORs, -GRs, -IRs, -iGluRs,-mGluRs, -SNMPs, -OBPs,* and *-CSPs*. The nucleotide sequences of each annotated gene set were inserted into CLC Genomics Workbench suite and aligned with the 84 million RNA-seq reads using the RNA-seq analysis pipeline with default settings (CLC bio, 2012). Since tsetse flies are eukaryotes having spliced coding sequences, the results were normalized counts of density of reads corresponding to whole coding sequence (exonic regions only) and quantified as reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi *et al.*, 2008).

2.5 Phylogenetic analysis of chemoreceptors, glutamate-gated receptors, and chemosensory associated proteins in *G. m. morsitans*

Phylogenetic relationships between *G. m. morsitans* chemosensory proteins and their homologs from *D. melanogaster* and *An. gambiae* were analyzed for each gene family separately. For instance, ORs were analyzed separately from GRs; the glutamate-gated receptors (IRs, iGluRs, and mGluRs) were analyzed together; and SNMPs, OBPs and CSPs were analyzed individually. The analysis for ORs comprised 43, 62 and 73 ORs from *G. m. morsitans, D. melanogaster,* and *An. gambiae* respectively. The GRs analysis comprised 14, 59 and 94 GRs from *G. m. morsitans, D. melanogaster,* and *An. gambiae* respectively. Glutamate-gated family analysis had 15 iGluRs, 19 IRs and 6 mGluRs from *G. m. morsitans;* 58 iGluRs, 12 IRs and 5 mGluR-like from *D. melanogaster;* and 12 iGluRs-like, 41 IRs and 14 GPRMGL from *An. gambiae*. The analysis for OBPs comprised 32, 52 and 55 from *G. m. morsitans, D. melanogaster,* and *An. gambiae* respectively; CSP analysis composed of 5, 4 and 6 members from *G. m. morsitans, D. melanogaster,* and *An. gambiae* respectively. Finally, SNMP analysis was combined to include all related CD36-like sequences from the three species – that is, there were 15, 14

and 15 CD36-like sequences from *G. m. morsitans, D. melanogaster,* and *An. gambiae* respectively. The SNMP is a highly conserved extracellular receptor involved uniquely in mating responses across many insects. To assess the relationships of homologs *G. m. morsitans, D. melanogaster,* and *An. gambiae, A. mellifera* homologs were used as out-groups in the analysis for CD36-like sequences.

Each category of sequences was aligned using MUSCLE (MUltiple Sequence Comparison by Log-Expectation) tool (Edgar, 2004) and edited using Jalview web-server (Waterhouse *et al.*, 2009). Phylogenetic trees were reconstructed using Maximum Likelihood (ML) approaches and tree topologies inferred by Whelan and Goldman + Freq. model (Whelan & Goldman, 2001) as best fitting model (chosen from a panel of prior model test for each protein sub-family). Initial tree constructs were generated using neighbor joining heuristic method, and consensus trees bootstrapped 500 times and taken to represent the evolutionary history of the taxa analyzed. Branches with less than 50% bootstrap replicates were collapsed (Felsenstein, 1985). The evolutionary rate difference among sites was modeled using a discrete Gamma distribution with five categories, allowing some sites to be evolutionarily invariable. All positions with less than 95% site coverage were set to be eliminated during analysis. The trees for ORs, GRs, IRs, iGluRs and mGluRs were reconstructed using MEGA5 suite (Tamura *et al.*, 2011), and those of OBPs, CSPs and CD36-like were reconstructed using an ML improved amino acid replacement matrix model, LG (Le & Gascuel, 2008), in PhyML program version 3.0 (Quindon *et al.*, 2010) and edited in FigTree tool (http://tree.bio.ed.ac.uk/software/figtree/). Evolutionary distance principal component analysis (PCA) with respect to chemoreceptor genes against other selected diptera species was done via GenAlEx version 6.5 tool (Peakall & Smouse, 2012). Independent analysis between selected chemoreceptors (ORs, GRs and IRs) in *G. m. morsitans* and annotated homologs from *D. melanogaster, An. gambiae, Ae. aegypti, Culex quinquefasciatus, A. mellifera* and *T. casteneum* was done via Glossinia phylome database 182 (Huerta-Cepas *et al.*, 2011). In Overall, glossinia database contained 223 different phylomes representing different proteomes (some sub-species and clones). The database was searched using individual *G. m. morsitans* annotation identifiers input in the pipeline as seed queries to retrieve the related trees using Maximum Likelihood method based on amino acid models, either JTT matrix-based (Jones *et al.*, 1992) or WAG model (Whelan & Goldman, 2001). The glossinia database can be accessed at www.phylomedb.org/?q=user/28, with user credentials.

2.6 Statistical Analysis

All the data presented here incorporates individual statistical analyses as per the bioinformatic tools that have been used to analyze them. The analyzed data were presented using appropriate tables and graphs.

CHAPTER THREE

RESULTS

The International Glossina Genome Initiative (IGGI) recently sequenced the *G. m. morsitans* genome, and assembled its 366 Mb size into approximately 13,807 scaffolds (IGGI, 2014). The assembly was computationally annotated by the VectorBase using different tools and pipelines (Megy *et al.*, 2012). A total of 12,308 protein coding genes were predicted of which approximately 3,000 were manually curated by researchers within the IGGI consortium (IGGI, 2014). RNA-seq data was used to seed the manual curation of chemosensory genes in the Glossina genome to produce a gold standard of genes that were subsequently functionally annotated. The number of protein coding sequences reported by the IGGI consortium included 75 genes computationally predicted that were chemosensory related. This figure was improved via an in-house manual curation, in which 43 extra genes were identified. Overall, the work presented in this chapter contributed to the manually curated *G. m. morsitans* chemoreceptome already integrated into the VectorBase database.

The gene identification, annotation and analyses thereof, were done using standard bioinformatic tools either as standalone tools downloaded and installed locally or as online tools available from public databases or from private servers with permissions. The gene identification and annotation process was to determine the total number of each category of chemosensory-related genes present in the *G. m. morsitans* genome, their genomic location, genome structure and individual gene structure characteristics. The RNA-seq data were used for two purposes: (i) to facilitate gene identification and alternative splice variants detection processes – this helped in assessing the completeness of the coding regions of the genes; and (ii) to estimate the relative expression levels of all the identified genes. Knowing the expression levels of the genes was to help determine which genes could probably be targeted for future functional analyses. The phylogenetic analyses were performed to assess the likely evolutionary relationships with homologs found in other related insects. The results are presented in three categories namely: (a) Non-glutamate chemosensory receptors - ORs and GRs; (b) Glutamate-gated ion receptors - IRs, iGluRs, and mGluRs; and (c) Chemosensory soluble proteins - OBPs, CSPs, and CD36-like sensory neuron membrane proteins (SNMPs). Table 1 presents a comparative analysis of these gene repertoires in selected insects.

Table 1. Comparison of *Glossina m. morsitans* ORs, GRs, GluRs, OBPs, CSPs and CD36 with other selected insect species

Insect species	ORs	GRs	IRs	OBPs	CSP	CD36	References
D. melanogaster	60 (2)*	60 (13)*	61(1)	51	5	14	1; 2; 3; 4; 5; 6
G. m. morsitans	**46 (3)**	**14**	**19**	**32**	**5**	**15**	**This study.**
An. gambiae	79	76	41	67	8	13	5; 7; 8; 9; 10
Ae. aegypti	100(31)	91(23)	...	111	...	13	9; 10; 11; 12
Apis mellifera	163 (11)	10 (3)	...	21	6	8	5; 13; 14; 15
Tribolium casteneum	265 (76)	220 (25)	...	49	20	15	5; 16; 17

Figures in parentheses are numbers of incomplete genes and or pseudogenes of the receptors. *****- in parentheses are alternatively spliced forms. The gaps (…) indicate no literature found with details. References: **1** - Clyne *et al*., 1999; 2000; **2** - Robertson *et al*., 2003; **3** - Benton *et al*., 2009; **4**- Hekmat-Scafe, 2002; **5** - Nichols & Vogt, 2008; **6** - Vieira *et al*., 2007; **7** - Fox *et al*., 2001; **8** - Hill *et al*., 2002; **9** - Manoharan *et al*., 2013; **10** - Vieira & Rozas, 2011; **11** - Bohbot *et al*., 2007; **12** - Kent *et al*., 2008; **13** - Robertson & Wanner, 2006; **14** - Foret & Maleszka, 2006; **15** - Maleszka & Stange, 1997; **16** - Richards *et al*., 2008; **17** - Engsontia *et al*., 2008.

3.1 The annotated *G. m. morsitans* odorant receptor (OR) genes

3.1.1 Repertoires of *G. m. morsitans* ORs

Overall, 46 OR genes with 1-12 exons were identified in the Glossina genome. The length of the complete mature OR proteins ranged from 284-514 amino acids (Table 2). The *G. m. morsitans* OR genes were numerically named as *GmmOR1-46* and their proteins denoted as GmmOR1-46. 25 of the 46 OR genes were computationally predicted by the VectorBase team, and assigned computational identifiers. However, the accuracy of these computationally identified genes was improved by correcting spurious truncations and exons, and also identifying alternative splice variants for some of the genes. The remaining 21 OR genes were recovered through manual curation in this study, thus setting the total at 46. The manually curated ORs were assigned temporary gene identities designated as TMP_Or* (Table 2).

Most *G. m. morsitans* ORs were scattered amongst the scaffolds, 50% of them encoded as single-copies on their respective scaffolds while the rest were encoded as pairs or triplets per scaffold. Five sets of the ORs (*GmmOR6/OR7/OR8*, *GmmOR18/OR19* (scf7180000648792), *GmmOR22/OR25* (scf7180000649009), *GmmOR27/OR28* (scf7180000650866) and *GmmOR41/OR42*) (scf7180000649048) were encoded in tandem on their respective scaffolds (Appendix 4). *GmmOR28* was truncated on scaffold (scf7180000650866) and together with *GmmOR20* and *GmmOR22* were set as pseudo-genes. Though the stop codon for *GmmOR43* could not be determined as it seemed truncated on its scaffold (scf7180000651490) at the 3' end, it was sufficiently long (389 amino acids) and was considered as complete gene. Four alternative splice variants were detected in *GmmOR5* (GMOY012018), of which only one (*GmmOR5a*) that looked complete was analyzed.

Comparatively, the *G. m. morsitans* annotated OR genes were 14 fewer than those known in *D. melanogaster* (60 OR genes encoding 62 proteins by splice variants), and also 33 fewer than those reported in *An. gambiae* (73 genes encoding 79 ORs). Homology searches revealed three likely gene lineage expansions: first, a cluster of six related genes, GmmOR41-46 homologous to a drosophila Or67d gene known to encode a male specific mating pheromone receptor (see Table 1) - this is a classic expansion that may relate to functional importance of the mating pheromone in tsetse, deterring mated females from re-mating; second, GmmOR6-9 were homologs of drosophila Or42b, a gene that encodes a receptor with high affinity for low ethyl acetate concentrations; and third, GmmOR14-16 homologs to drosophila Or45a that encodes a larval deterrent receptor protein. Five *G. m. morsitans* OR genes namely GmmOR15, OR16, OR18, OR37, and OR40

were homologous to *Drosophila* but not in mosquitoes. Fifty nine percent (27 out of 46) of the *G. m. morsitans* OR genes had reciprocal BlastP hits to drosophila homologs; the rest were homologous to Mediterranean fruit fly (medfly, *Ceratitis capitata*) OR-like ESTs (14), *Stomoxys calcitrans* putative OR ESTs (3), *Chrysomya ruffacies* OR-like sequence (1) and to *An. gambiae* annotated OR (1).

Table 2. Annotations of odorant receptor genes in *G m. morsitans* and their query homologs from *D. melanogaster* and *An. gambiae*

G. m. morsitans	Length (AA)	Exons	TMMs	Identities	*D. melanogaster* (E-value)	*An. gambiaae* (E-value)
GmmOR1	521	8	6	GMOY005610	DmOr83b(0.00)	AgOr7(6.26E-101)
GmmOR2	394	3	7	GMOY005796	DmOr2a(2.67E-067)	AgOr51(3.33E-006)
GmmOR3	387	3	3	GMOY004772	DmOr19a(2.31E-038)	AgOr45(0.034)
GmmOR4	384	2	7	**TMP_Or4**	DmOr59a(2.70E-074)	AgOr1(3.64E-012)
GmmOR5*	442	4	5	GMOY012018	DmOr33b(4.07E-054)	...
GmmOR6	387	4	5	GMOY009475	DmOr42b(1.24E-054)	AgOr56(9.63E-003)
GmmOR7	406	3	6	**TMP_Or7**	DmOr42b(6.30E-048)	AgOr56(7.94E-007)
GmmOR8	389	4	6	**TMP_Or8**	DmOr42b(4.68E-045)	AgOr56(9.55E-003)
GmmOR9	409	3	6	**TMP_Or9**	DmOr42b(1.98E-087)	AgOr47(0.01)
GmmOR10	444	3	6	**TMP_Or10**	DmOr46a(1.04E-165)	AgOr12A(0.12)
GmmOR11	341	3	6	GMOY010761	DmOr46a(3.69E-048)	AgOr65(6.04E-008)
GmmOR12	340	3	3	GMOY009271	DmOr94b(2.81E-060)	AgOr34(2.07E-018)
GmmOR13	391	6	6	GMOY003312	DmOr82a(1.15E-045)	...
GmmOR14	341	3	6	GMOY001365	DmOr45a(5.06E-024)	AgOr35(9.31E-008)
GmmOR15	446	4	7	**TMP_Or15**	DmOr45a(6.42E-020)	...
GmmOR16	387	4	6	**TMP_Or16**	DmOr45a(1.85E-009)	...
GmmOR17	541	12	8	GMOY005386	DmOr69a(3.38E-004)	AgOr48(0.01)
GmmOR18	420	8	6	**TMP_Or18**	DmOr63a(2.98E-065)	...
GmmOR19	385	8	7	GMOY012322	DmOr63a(1.32E-026)	...
GmmOR20#	269	7	6	**TMP_Or20**	DmOr85b(1.89E-077)	...
GmmOR21	465	5	2	GMOY011399	DmOr83a(2.50E-030)	AgOr40(1.96E-021)
GmmOR22#	296	4	5	**TMP_Or22**	DmOr49a(4.26E-025)	...
GmmOR23	331	4	5	**TMP_Or23**	DmOr85b(4.35E-005)	AgOr40(0.22)
GmmOR24	388	3	6	GMOY010839	DmOr85c(3.05E-086)	AgOr8(2.39E-029)
GmmOR25	385	3	6	GMOY012357	DmOr56a(1.33E-063)	...
GmmOR26	418	4	5	**TMP_Or26**	DmOr85b(9.02E-013)	AgOr40(7.90E-003)
GmmOR27	415	3	6	GMOY008038	DmOr67c(2.132E-14)	...
GmmOR28#	260	2	7	**TMP_Or28**	DmOr92a(2.64E-038)	...
GmmOR29	438	3	4	**TMP_Or29**	DmOr67a(1.59E-006)	AgOr69(0.01)

GmmOR30	361	3	6	**TMP_Or30**	DmOr67a(8.42E-007)	AgOr69(6.99E-003)
GmmOR31	435	7	5	**TMP_Or31**	DmOr24a(6.39E-104)	AgOr54(0.685)
GmmOR32	450	5	7	GMOY005084	DmOr13a(4.71E-124)	AgOr39(9.29E-007)
GmmOR33	353	6	5	GMOY005479	DmOr49b(2.88E-116)	...
GmmOR34	360	7	4	GMOY011902	DmOr30a(2.06E-097)	...
GmmOR35	392	5	6	**TMP_Or35**	DmOr43a(2.41E-090)	AgOr10(4.47E-026)
GmmOR36	343	7	6	**TMP_Or36**	DmOr43a(4.50E-009)	AgOr2(1.12E-023)
GmmOR37	430	4	4	**TMP_Or37**	DmOr74a(4.12E-079)	...
GmmOR38	371	5	6	**TMP_Or38**	DmOr47b(2.99E-023)	...
GmmOR39	403	3	6	GMOY004392	DmOr88a(1.15E-023)	...
GmmOR40	284	5	6	GMOY012356	DmOr56a(3.21E-027)	...
GmmOR41	386	4	6	GMOY006480	DmOr67d(1.13E-037)	AgOr44(3.12E-007)
GmmOR42	386	4	5	GMOY006479	DmOr67d(5.15E-039)	AgOr41(4.04E-006)
GmmOR43	389	4	5	**TMP_Or43**	DmOr67d(2.77E-031)	...
GmmOR44	390	4	6	GMOY006265	DmOr67d(9.65E-057)	AgOr73(8.34E-013)
GmmOR45	385	4	7	GMOY007896	DmOr67d(4.30E-031)	AgOr43(3.22E-006)
GmmOR46	348	4	3	GMOY003305	DmOr67d(1.30E-048)	AgOr23(0.019)

GmmOR – *Glossina morsitans morsitans* ordorant receptor; TMM- Trans-membrane helices; GMOY – *Glossina morsitans* Yale strain; TMP_Or – Provisional odorant receptor ID; DmOr- *Drosophila melanogaster* odorant receptor;*- longest splice variant in gene GmmOR5; # - pseudogene; Gene identities in **bold face** are novel genes manually curated. The gaps (…) indicate no matching query sequence found in *An. gmambiae.*

Analyses of heptatopic helices in *G. m. morsitans* ORs are summarized in Appendix 14. Eight *G. m. morsitans* OR proteins, namely GmmOR3, OR11, OR32, OR33, OR36, OR37, OR45, and OR46, had known reversed extracellular N-termini topology for insect ORs. Four *G. m. morsitans* ORs namely, GmmOR8, OR12, OR18, OR23 had no alpha helix within their first 20 codons counts from their N-termini that correspond to probable site for insert signal for membrane proteins. The average predicted alpha helix length of *G. m. morsitans* ORs was 22 amino acids, with 19 OR sequences having between 17 and 19 amino acids. However, some genes had disproportionate alpha helix sizes e.g. GmmOR14 had its 6th helix predicted size of 34 amino acids, and similarly, the 4th helices of both GmmOR33 and GmmOR34 was 31 amino acids (Appendix 15 and 16). Definitive conserved seven trans-membrane domains for insect odorant receptor, 7tm-6, were detected in all the *G. m. morsitans* ORs.

Three different motifs of unknown functions were detected in at least each of the 42 complete *G. m. morsitans* ORs (Figure 3.1). The motifs were randomly distributed along the OR sequences. Common residues in the motifs were F, I, L, A, M, W, V (all hydrophobic in nature), S, Q, G (polar non-charged), D, E (acidic residues), and K, R (positively charged). The regular expression for the three *G. m. morsitans* OR motifs were FA[ML]Y[AQ]x[RV]L[EL]x (motif 1), SSxIx[EQ]GGLK (motif 2), and I[GS]AA[YF]SS[IL]LS (motif 3) (Appendix 5).

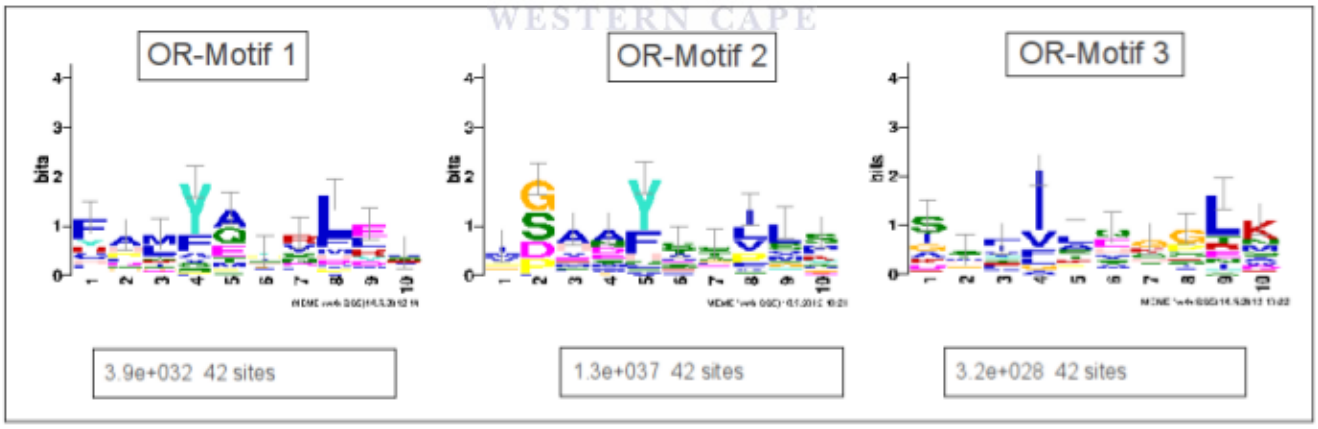


Figure 3.1 Atypically conserved motifs in GmmORs.
The motifs with unknown functions predicted using MEME suite (Bailey *et al.*, 2009). The logos indicate the consensus amino acid residues across the 42 complete OR sequences.

3.1.2 Assessment of functional classification of ORs in *G. m. morsitans*

Cellular classifications confirmed *G. m. morsitans OR* genes were associated with neuronal membrane proteins (Figure 3.2). Ten percent of the *G. m. morsitans ORs* (six genes, *GmmOR1, OR19, OR23, OR32, OR33,* and *OR37)* had motifs for GPCR molecular activity with significant GPCR signaling pathway class, in support of earlier observations by Wicher et al. (2008). The motif for chemical stimulus detection was present (37%) and sensory signal transduction processes (24%). The *G. m. morsitans* ORs with olfactory receptor activities were not associated with sensory signal transduction, suggesting stimulus perception is separated from signal transduction. Molecular sensory function was confirmed by the presence of receptor activity function including olfactory (for ORs), and odorant binding activities, constituting classification for sense stimuli detection at 51%, and signal transduction activity at 42% (Figure 3.2).

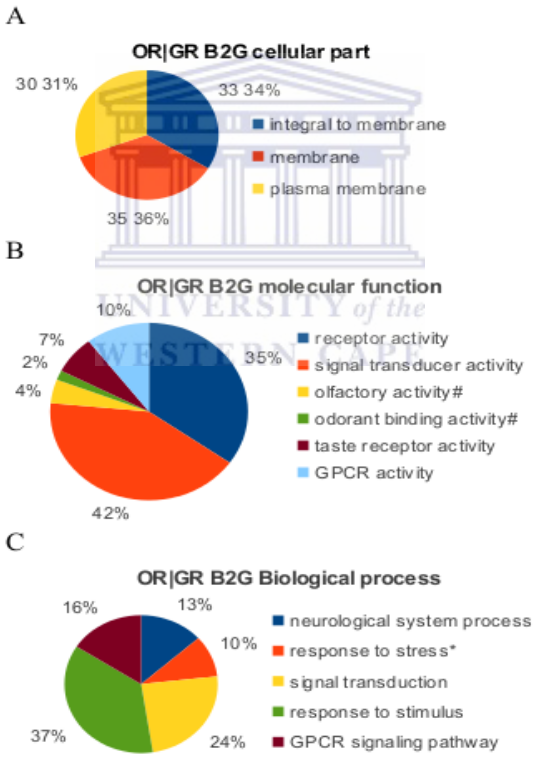


Figure 3.2 Functional classifications of ORs and GRs in *G. m. morsitans*.
A – Cellular localizations indicating the genes are membrane proteins; B – molecular function classification showing olfactory and signal transducer activities; C – biological process classification indicate response to stimuli as the main process in which these genes are involved. # - functions specific to *GmmORs* only; * - functions specific to *GmmGR* only. The functional classifications were estimated via B2G web server (Conesa *et al*., 2005).

3.1.3 Putative expression levels of OR genes in *G. m. morsitans*

The RNA-seq expression profiles for *G. m. morsitans ORs* are presented in (Figure 3.3). The expression levels of *GmmOR15* (684810.26 RPKM), *OR2* (21764.01 RPKM), *OR1* (8145.92 RPKM), *OR43* (6891.30 RPKM) and *OR9* (2712.13 RPKM) genes were relatively higher than the rest, and accounted for 96.26% of total RNA-seq reads (n = 752,853.89 RPKM) supporting *G. m. morsitans* OR gene expression. Of this, the expression level of *GmmOR15* (see inset in Figure 3.3), a homolog of Or45a in *D. melanogaster,* accounted for 90.75% of the supporting data. The second most expressed OR gene was *GmmOR2* (longest bar graph from bottom). There were no sufficient data to quantify expression levels of five *G. m. morsitans* OR genes – *GmmOR8, OR11, OR25, OR31,* and *OR39*. This does not necessarily mean that the genes are dormant and not expressed.

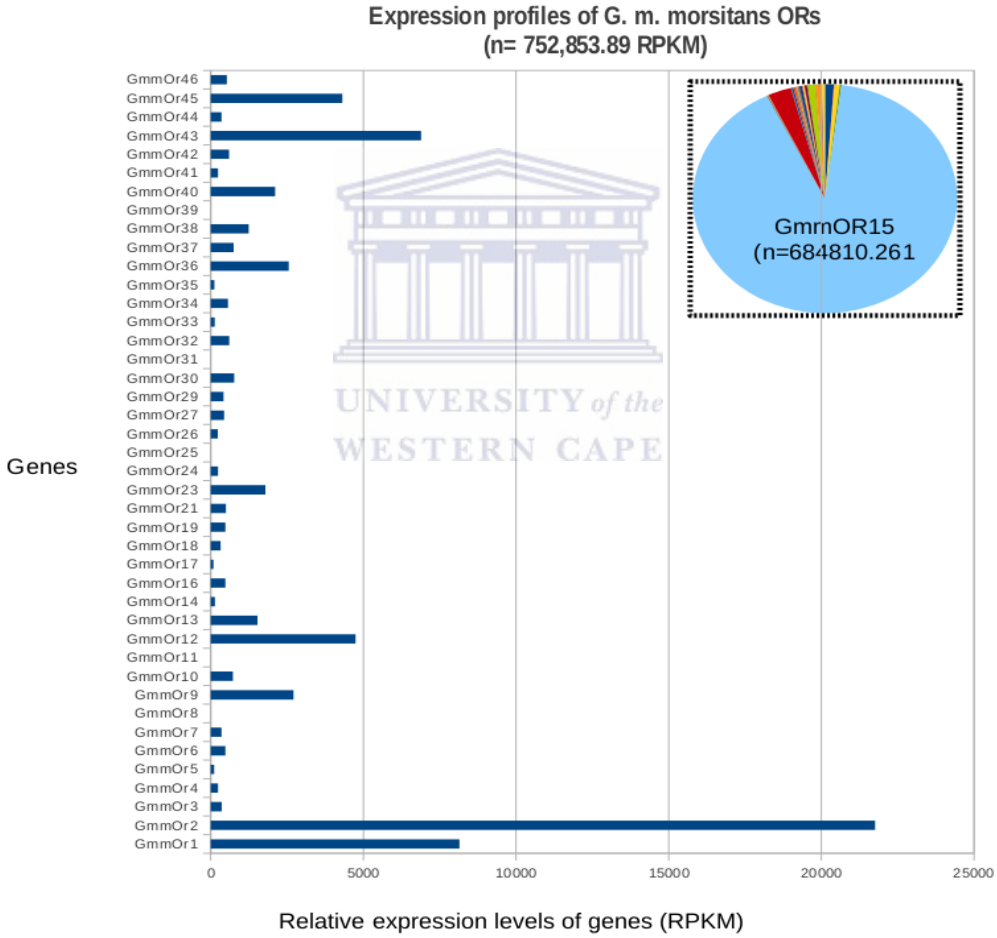


Figure 3.3 Expression levels of *G. m. morsitans* ORs.
The gene names are numerically ordered. *GmmOR2, OR1* (Orco) and *OR43* are highly expressed. **Inset:** The pie chart depicting expression levels of all GmmORs including *GmmOR15* (light blue area of chart), this represented over 90.75% of supportive data (n = 684,810.261). No sufficient data to quantify expression of *OR8, OR11, OR25, OR31* and *OR39.*

3.1.4 Phylogenetic relationships between *G. m. morsitans* odorant receptors and those in *D. melanogaster* and *An. gambiae*

Maximum likelihood tree rooted at mid-point comprising 46 *G. m. morsitans*, 62 *D. melanogaster* and 73 *An gambiae* ORs revealed five sub/clusters (Figure 3.4). Many *G. m. morsitans* ORs lineages clustered amongst themselves, suggesting they could be related by close ancestry. For instance – cluster A comprised GmmOR23-30, OR38 and OR39 that are not directly related to homologs in the other insects; cluster B has GmmOR13-16, OR18, OR19, OR31, OR32 and OR37 in a heterogeneous group with a fair number of membership from other two species. These receptors likely regulate pregnancy and/or larval survival, and larviposition site selection. The cluster also contained paralogs GmmOR14-16 which were homologous to DmelOr45a, a larval deterrent receptor. Cluster C contained GmmOR41-46 clustering with the drosophila receptor DmelOr67d (a cVA receptor) and DmelOr83c, in a sub-tree dominated by *An. gambiae* specific receptors. Non-canonical Orco homolog GmmOR1 (counterpart of DmelOr83b and AgOr7) and distantly related to GmmOR17, rooted a cluster that comprised also GmmOR2-OR5, OR9 (cluster D). Finally, cluster E, included GmmOR33-OR36, OR40, and distantly OR21. Cluster E occupied the base branching of the tree with expanded membership from the *Anopheles*. Additionally, five sets of Glossina OR residues - GmmOR6-OR8, OR10-OR11, OR14-OR16, OR18-OR19, and OR26-OR30 clustered uniquely suggesting they could be duplicates.

Additional evaluation of phylogenetic relationships between *G. m. morsitans* ORs and other annotated OR proteomes across selected eukarya genomes (for our relevance, *Homo sapiens*, *Tribolium casteneum*, and *Apis mellifera* outgroups) in PhylomeDB, revealed one GmmOR lineage cluster of five in-paralogs (GmmOR41, OR42, OR44, OR45, OR46), all homologous to DmelOr67d, probably arose from three speciation and two duplication events after their most recent common ancestor, MRCA (Appendix 6). The cluster also had expanded co-orthologs in *An. gambiae*, *Cx. quinquefasciatus*, *Ae. aegypti*, and had a copy each in *T. casteneum*, and *A. mellifera* outgroups. Another cluster comprised GmmOR2 -OR12, predictably arose from several duplications and speciation events from a common ancestor. Majority of the *Glossina* ORs were closely related to the drosophila homologs than to the mosquito homologs.

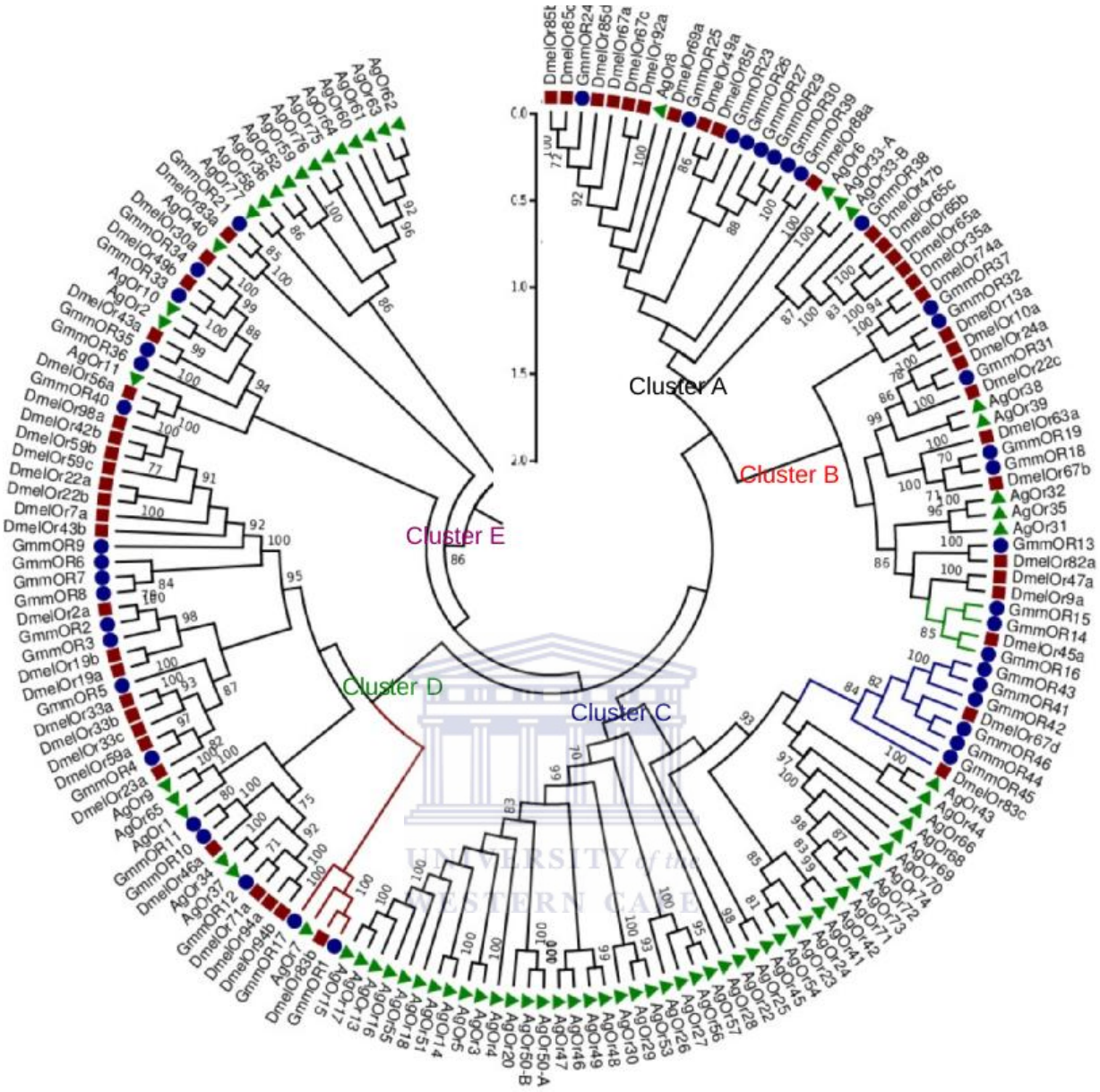


Figure 3.4 Maximum likelihood tree for *G. m. morsitans* ORs with homologs from *D. melanogaster* and *An. gambiae*.
The terminals in blue dots are Glossina ORs, red squares are Drosophila ORs, while green triangles are Anopheles ORs. Leaf branches in green indicate homologs of DmelOr45a, branches annotated blue show homologs of DmelOr67d, while branches in red represent the Orco members. The bootstrap values are indicated on the tree nodes, at least 70% bootstrap value supports are shown. Multiple sequence alignment was done using MUSCLE tool (Edgar, 2004) and edited via Jalview tool (Waterhouse *et al.*, 2009). The tree was reconstructed using MEGA5 tool (Tamura *et al.*, 2011).

3.2 The annotated gustatory receptor (GR) genes in *G. m. morsitans*

3.2.1 Repertoires of *G. m. morsitans* GRs

Fourteen GR genes were identified and annotated in *G. m. morsitans* genome, of which eight genes corresponded to those computationally predicted by the IGGI VectorBase team; while six novel genes were manually curated and assigned temporary identities as TMP_Gr* in this study. The *G. m. morsitans* GR genes were named *GmmGR1-14* and their proteins as GmmGR1-14. The GmmGRs exons ranged from 2-8 exons and protein lengths ranged 309-541 amino acids (Table 3). The numbers of recovered *G. m. morsitans* GR genes were fewer by 54 relative to *D. melanogaster* (68 GRs) and fewer by 76 relative to *An. gambiae* (considering a total of 90 GRs so far reported). All the *G. m. morsitans* GRs had a representative homolog in both the drosophila and the mosquito, except *GmmGR8* that was not represented in the mosquito. Reciprocal DELTA Blast using *G. m. morsitans* GRs against the non-redundant protein database revealed homologs with similarity to CO_2 sensitive receptors, appetitive (attractive) tastes, and bitter (aversive) tastes. The *G. m. morsitans* GRs (GR1-4) that were similar to related CO_2 sensors were conserved as in other diptera.

Four *G. m. morsitans* GRs were homologous to *An. gambiae* GR proteins (GmmGR2 and GR3 to AgGr23 (AGAP003098), GmmGR11 to AgGr26 (AGAP006716), and GmmGR13 to AgGr33 (AGAP006717)), and two (GmmGR1 and GR4) were homologous to *Chrysomya megacephala* predicted gustatory receptor 1 and 2. The remaining eight Glossina GRs had homologs amongst drosophila species (see Appendix 14). The predicted trans-membrane domains ranged 4 to 8. There were no sweet taste receptor genes recovered in the tsetse genome consistent with their obligate blood feeding, in stark contrast to highly conserved homologs in other insects including drosophila and mosquitoes that feed on sugar-based saps and nectaries. Five *G. m. morsitans* GRs namely GmmGR2, GR3, GR6, GR9, and GR14 clustered on a single scaffold scf7180000652170, the largest scaffold in the genome, while the rest of the GRs were encoded as single-copies on their respective scaffolds. All *G. m. morsitans* GR genes were annotated as complete without any pseudo-gene or a splice variant detected.

Table 3. Annotation of gustatory receptor genes in *G m. morsitans* and their query homologs from *D. melanogaster* and *An. gambiae*

G. m. morsitans	**Length (AA)**	**Exons**	**TMMs**	**Identities**	***D. melanogaster*(E-value)**	***An. gambiae* (E-value)**
GmmGR1	425	3	6	GMOY007472	DmGr21a(2.00E-088)	AgGr22(0.00)
GmmGR2	514	7	6	GMOY011510	DmGr22b(2.00E-048)	AgGr23(0.00)
GmmGR3	425	6	6	**TMP_Gr5**	DmGr21a(3.00E-036)	AgGr23(1.00E-157)
GmmGR4	496	8	6	GMOY008001	DmGr63a(5.00E-066)	AgGr24(2.00E-148)
GmmGR5	467	5	7	GMOY004207	DmGr66a(2.00E-045)	AgGr2(4.00E069)
GmmGR6	443	4	8	GMOY011615	DmGr28b(1.00E-059)	AgGr33(2.00E077)
GmmGR7	402	3	7	GMOY006209	DmGr28b(1.00E-039)	AgGr33(9.00E-055)
GmmGR8	407	2	6	**TMP_Gr4**	DmGr22e(9.00E-025)	...
GmmGR9	348	5	4	GMOY011903	DmGr2a(2.00E-031)	AgGr2(4.00E018)
GmmGR10	458	4	7	GMOY003231	DmGr33a(5.00E-031)	AgGr43(2.00E-006)
GmmGR11	450	3	6	**TMP_Gr3**	DmGr22b(2.00E-021)	AgGr26(0.026)
GmmGR12	375	2	8	**TMP_Gr2**	DmGr32a(3.00E-022)	AgGr26(2.00E-004)
GmmGR13	457	2	6	**TMP_Gr1**	DmGr22b(3.00E-025)	AgGr33(4.00E002)
GmmGR14	309	3	6	**TMP_Gr6**	DmGr22b(2.00E-014)	AgGr33(3.00E-026)

GmmGR – *Glossina morsitans morsitans* gustatory receptor; TMM- Trans-membrane helices; GMOY – *Glossina morsitans* Yale strain; TMP_Gr – Provisional gustatory receptor ID (bold face); DmGR- *D. melanogaster* gustatory receptor. There was no ortholog of GmmGR8 in *An. gambiae* (see gap, '…').

The *G. m. morsitans* GRs had known conserved domain for gustatory receptors, 7tm-7. Four *G. m. morsitans* GRs namely GmmGR6, GR7, GR9, GR10, and GR14, had reversed extracellular N-termini topology. Another five *G. m. morsitans* GRs namely GmmGR1-4, and GR11, had no alpha helix within their first 20 codon counts from their N-termini. Similar to ORs, *G. m. morsitans* GRs had three conserved secondary structure motifs spread across each of the 14 sequences (Figure 3.5). The *G. m. morsitans* GR regular expressions for motifs were [RK][AQ][LK][WY]LD[VLM][KSD]E[LY][LT][QK]Q[LF][GN] (motif 1, e-value 8.1e+000), QFY[RE]AL[KQ]PLLI[LI][LSF]xI[LY]G[VLC][TM]PIx[RIL][SQ]xPK (motif 2, e-value 1.4e-002), and [NT][LA][AK]G[FYL]FN[IV][ND]REL[YL]F[GLT] (motif 3, e-value 3.8e+006). Motif 2 of *G. m. morsitans* GRs showed a somewhat bias for N-termini except in GR8 and 9, while motifs 1 and 3 tended towards the C-termini half, with 50% of motif 3 being located at the extreme end (Appendix 7).

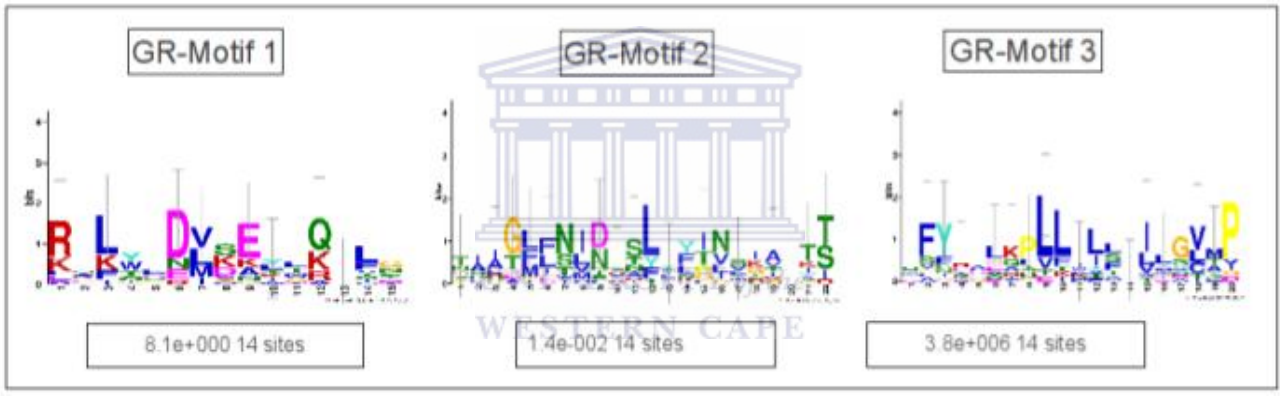


Figure 3.5 Atypically conserved motifs in GmmGRs.
The motifs with unknown functions in GRs predicted using MEME suite (Bailey *et al*., 2009). The logos indicate the consensus amino acid residues across the14 sites respectively.

3.2.2 Functional classification of GRs in *G. m. morsitans*

The distribution of gene ontology functions in *Glossina GR*s confirmed they are neurological receptor genes with molecular taste receptor activity and participate in neurological and stimulus detection processes. Chemical taste receptor activity was detected in all the GRs, majority of which also had GPCR molecular activities, proof that they function via GPCR second messenger pathway. Seven of 14 (50%) *G. m. morsitans* GRs had enriched motifs for stress responses. No ontology functions for sweet taste were detected, revealing none of the Glossina GRs are involved in detection of sweet sugars (see above Figure 3.2).

3.2.3 Expression levels of *G. m. morsitans* GR genes

The *G. m. morsitans GR1-GR4* had high RNA-seq expression levels, which also corresponded to known insect olfactory receptors for carbon dioxide (CO_2). The remaining *G. m. morsitans* GR genes had lower expression levels. There were insufficient data to quantify the expression levels for GmmGR10, GR13 and GR14 (Figure 3.6).

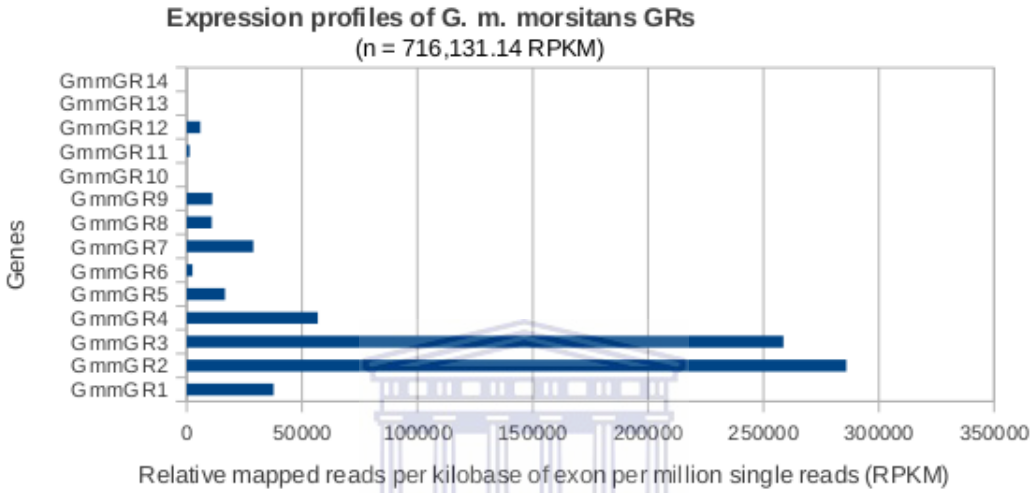


Figure 3.6 Expression levels of *Glossina m. morsitans* GRs.
GmmGR2 and GR3 are highly expressed corresponding to known receptors sensitive to CO_2. The expression profiles were estimated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

3.2.4 Phylogenetic relationships of *G. m. morsitans* gustatory receptors with orthologs from *D. melanogaster* and *An. gambiae*

Four *G. m. morsitans* GR receptors, GmmGR1-4, whose homologs (DmelGr21a, Gr63a and AgGr22-24) are linked to known CO_2 sensitive receptors clustered, revealing they are highly conserved (branches in blue Figure 3.7). However, an extended analysis with other insect homologs suggests the *G. m. morsitans* GR1-4 (the CO_2 receptors) seemed much closer to housefly *Musca domestica* and Mediterranean fruit fly *Ceratitis capitata* than to the facultative blood-feeding mosquitoes and vinegar fly *D. melanogaster* (Figure 3.8). Another cluster had GmmGR6, GR7 (homologous to AgGr33 and DmelGr28b, a splice variant of DmelGr28a), and GmmGR10 and GR14. In general, the *G. m. morsitans* GRs segregated into five clusters with their fruit fly and mosquito counter-parts. However, six *G. m. morsitans* GRs GmmGR2, GR6, GR8,

GR11, GR12, and GR14 were only distantly related to their drosophila and mosquito relatives with no direct homolog in the tree, while GmmGR11 and GR12 could likely be duplicates of an ancestral gene. Overall, these may represent putatively the *G. m. morsitans* unique GRs.

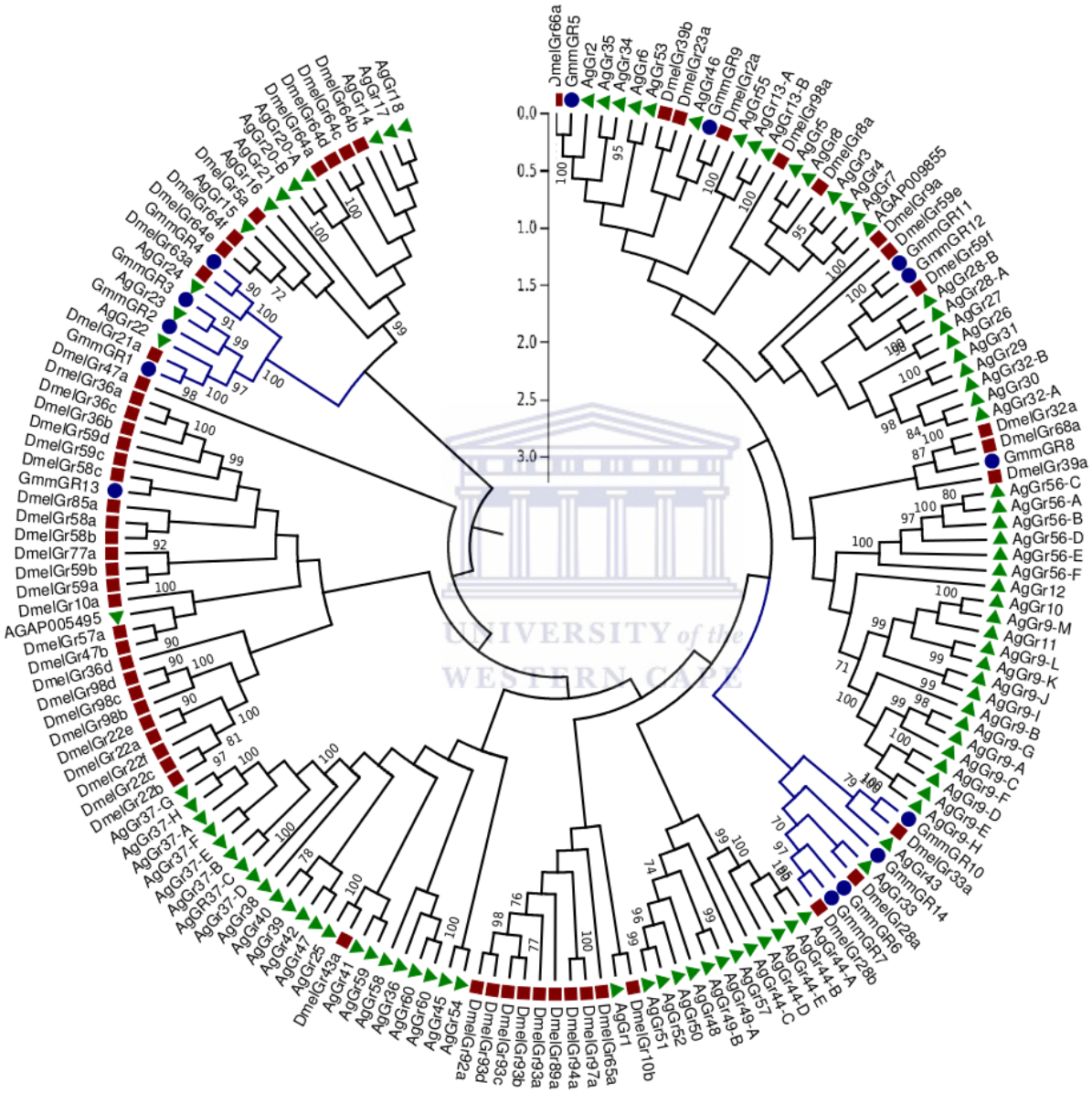


Figure 3.7 Maximum likelihood tree for *G. m. morsitans* GRs with homologs from *D. melanogaster* and *An. gambiae*.
The terminals in blue dots are Glossina GRs, red squares are Drosophila GRs, while green triangular shapes represent Anopheles residues. The bootstrap values are indicated on the tree nodes. The blue colored tree branches indicate clusters with Glossina orthologs for DmelGr21a, 63a and 28b. Multiple sequence alignment was done using MUSCLE tool (Edgar, 2004) and edited via Jalview tool (Waterhouse *et al.*, 2009). The distance tree was constructed using MEGA5 tool (Tamura *et al.*, 2011).

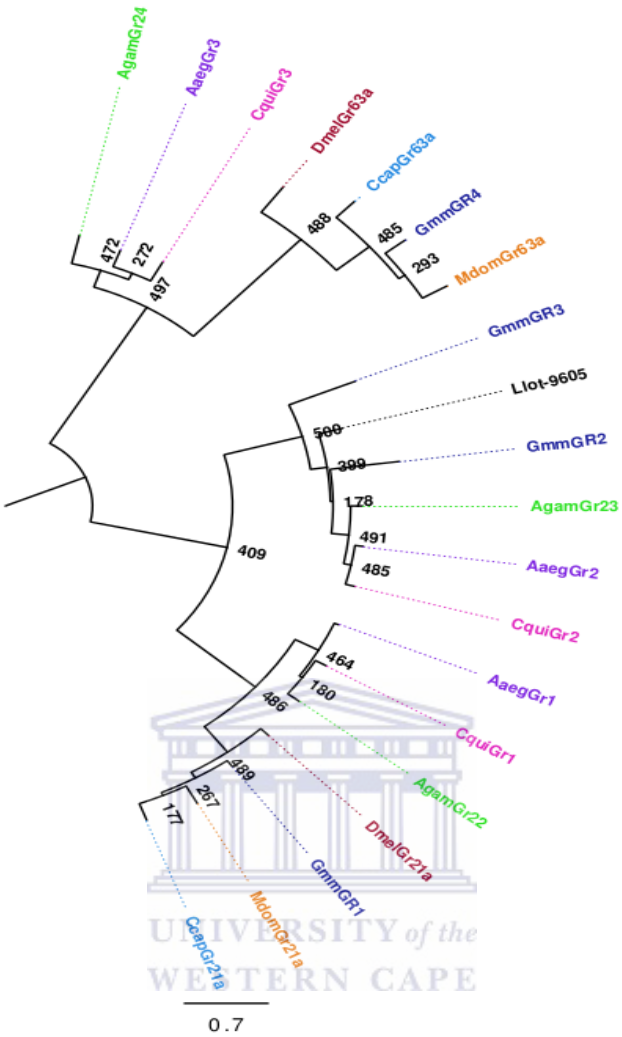


Figure 3.8 Phylogenetic tree of carbon (IV) oxide sensitive gustatory receptors amongst selected diptera.
The node branch bootstrap support is shown at nodes relative 500 iterations. Agam - *An. gambiae;* Cqui - *Cx. quinquefasciatus;* Aaeg - *Ae. aegypti;* Llot - *Lutzomya longipalpis;* Mdom - *Musca domestica;* Ccap - *Ceratitis capitata*. The tree was constructed using a standalone PhyML program (Quindon *et al*., 2009) and edited using Figtree tool (http://tree.bio.ed.ac.uk/software/figtree/).

3.3 The annotated glutamate-gated ion receptor GluR (IRs, iGluRs and mGluRs) genes in *G. m. morsitans*

3.3.1 Repertoire of *G. m. morsitans* iGluRs

Fifteen iGluRs genes were identified and annotated in *G. m. morsitans* genome, of which 13 corresponded to those computationally predicted by VectorBase team but improved via manual curation in this study, and two genes were manually identified and assigned temporary identities as TMP_iGluR* (Table 4). Majority of *G. m. morsitans* iGluRs were encoded by large genes of at least 10 exons, with exceptions above 1000 amino acids. Surprisingly, there was no homolog

of DmelGluRIID identified in the *G. m. morsitans* genome. The *G. m. morsitans* iGluRs were highly conserved in sequence (with at least 70% similarities, e-value 0.0) and functional domains between themselves and also with their homologs in *D. melanogaster*. There were three sub-types of iGluRs identified in the *Glossina*, namely; Ampa-like (one gene), kainate-like (12 genes), and NmdaR-like (two genes), which were assigned names following the nomenclature used for their homologs in *D. melanogaster.* For instance, the ampa-like gene was named *GmmGluRIB*; the kainate-like genes were named as *GmmGluR** and *KaiR**, and NmdaR-like as *GmmNMDAR1* and *NMDAR2*. Portions of *GmmGluRIA-b* and *GluRIIE-b* overlapped unsequenced genome regions; therefore they were annotated as incomplete. *GluRIIA* was annotated with two alternative splice variants, differing in length of their internal exons, and were subsequently named *GluRIIA.1* and *GluRIIA.2*. Additionally, three splice variants were detected in *GmmKaiRIA*, and two variants each in *GmmNMDAR2* and *GmmKaiR2-like d*. These splice variants differed in either their C-terminal exons (*GmmKaiRIA*), some had extra C-terminal internal exons (*GmmNMDAR2* and *GmmKaiR2-like d*) (data not shown). Probable duplicate genes annotated on different scaffolds were *GmmGluRIIE-a* and *GluRIIE-b; GluRIA-a* and *GluRIA-b;* and *KaiR2, KaiR2-like-c* and *KaiR2-like-d*. Further, the *G. m. morsitans iGluR* genes were encoded scattered in their respective genome scaffolds with no neighboring family members. However, *G. m. morsitans Clumsy*, *GluRIIA* and *GluRIIB* clustered in tandem on scaffold scf7180000649055*;* the clumsy gene encoded on the complementary strand (Appendix 8). The *G. m. morsitans GluRIIA* and *GluRIIB* were separated by a 258 amino acids, a sequence length that is probably shorter for an independent upstream regulatory promoter region.

Table 4: Annotated ionotropic glutamate-gated receptors (iGluRs) in *Glossina morsitans morsitans*

Glossina morsitans morsitans					*D. melanogaster*	BLAST2GO HITS		
Gene Name	**Identities**	**Exons**	**AA**	**Pfam Domain**	**Ortholog ID**	**B2G Identity**	**Similarity (%)**	**Description**
GluRIA-a	GMOY001262	14	862	ANF, LC	CG8442	XP_002054165	96.5	Ionotropic glutamate receptor ia
GluRIA-b#	GMOY006890	4	208	LC, SBP	CG8442	XP_002054165	83.6	Ionotropic glutamate receptor ia
GluRIB	GMOY012136	15	1108	ANF, LC, SBP	CG4481	XP_001983344	88.1	Ionotropic glutamate receptor ib
GluRIIA.1	GMOY012165-RA	13	934	ANF, LC, SBP	CG6992	XP_002051413	80.9	Glutamate receptor iia
GluRIIA.2	GMOY012165-RB	11	656	ANF, LC, SBP	...	...	...	
GluRIIB	**TMP_iGluR1**	10	805	ANF, LC, SBP	CG7234	XP_002051414	72.9	Glutamate receptor iib
GluRIIC	GMOY012186	13	1127	ANF, LC	CG4226	XP_002020926	72.3	Glutamate receptor iic
GluRIIE-a	**TMP_iGluR2**	14	989	ANF, LC, SBP, PBP6	CG31201	XP_001979403	84.1	Glutamate receptor iie
GluRIIE-b#	GMOY009209	3	151	...	...	...	...	
NmdaR1	GMOY007988	15	1034	ANF, LC, SBP, CaM	CG2902	XP_002038420	88.6	Nmda receptor 1
NmdaR2	GMOY012037	8	934	ANF, LC, SBP	CG33513	XP_002058334	97.9	Nmda receptor isoform c
Clumsy	GMOY006490	14	1199	ANF, LC, SBP	CG8681	XP_002065409	79.8	Isoform b
KaiRIA	GMOY001514	16	991	ANF, LC, SBP	CG18039	XP_002044295	79.1	Isoform a
KaiR2-lk-c	GMOY004959	11	853	NCL, NCM	CG9935	XP_002069012	70.2	Glutamate ionotropic kainate 2-like
KaiR2-lk-d	GMOY012113	11	1260	NCL, NCM	CG11155	XP_002011450	58.4	Glutamate ionotropic kainate 2-like
KaiR2	GMOY004222	15	937	NCL, NCM	CG3822	XP_002072149	97.5	Gluatamate ionotropic kainate 2

Columns from left are *G. m. morsitans* assigned gene names, gene identities (**TMP_iGluR*** are identities for manually curated, after they were not computationally captured), number of protein coding exons per gene, and AA- length of amino acids in mature protein. # designates incomplete genes. The gaps (…) indicate no data available. Pfam conserved domains: ANF – atrial natriuretic factor receptor (Pfam01094); LC – Ligand-gated channel (Pfam00060); LigC-G – Lig_Chan_Glu_bdg (Pfam10613); SBP – SBP_bac_3 (Pfam00497); PBP – Periplasmic binding protein domain typical of bacteria; CaM - calmodulin binding domain located at the C-terminal (Pfam10562); NCL – Neur_Chan_LBD (Pfam02931.18); NCM – Neur_Chan_Membrane (Pfam02932.11). The *D. melanogaster* query ortholog identities (CG-----). BLAST2GO searches were done using *G. m. morsitans* protein sequences at e-value less than 0.0001 via non-redundant Swiss-Prot database reporting only the best hit identity, hit species and the protein description.

Multiple sequence alignments for ligand binding domains LBD S1 and S2 and trans-membrane domains TM1-TM4 are presented in Appendix 9 and Appendix 10 respectively. The structural domain arrangements of *G. m. morsitans* iGluRs were typical of insect glutamatergic ionotropic receptors. The channel pore trans-membrane domains (TM1, P-loop/TM2, TM3, and TM4) and ligand-biding domains (S1, S2) sequences were the most conserved in all the *G. m. morsitans* iGluRs. However, analysis of the Venus fly trap structure of S1 and S2 domains, revealed a diversity of residues at known conserved sites dedicated for interaction with glutamate ligand. The R, T, and D/E sites were conserved in all iGluRs except NMDAR1 that had V residue at T site and both NMDA receptors had D instead of E residue at D/E site (Appendix 9). The LBD S2 extracellular loop was over 150 residues flanked by TM3 and TM4, but separated from both by linker sequences of variable lengths. All the *G. m. morsitans iGluRs* contained a conserved cysteine site after LBD S2 – its role yet unknown. At the ATD (amine terminal domain) site, the amino acid sequences of NMDAR1 and NMDAR2 resembled conserved LIVBP-like (leucine/isoleucine/valine-binding protein) domains, with similarity to heterodimer epsilon subunits of NMDA receptor family. Downstream at the LBD S1, PBP (periplasmic binding protein) type b domain overlapped the L-glutamate and glycine-binding site. Amongst the Kainate receptors, the LIVBP-like domain subunits of non-NMDA ionotropic receptors that are stimulated by neurotransmitter glutamate were detected.

Functional classification distributions in *G. m. morsitans iGluRs* revealed they are neuron membrane proteins with cellular location at neuromuscular junctions and post-synaptic membranes (Figure 3.9). The molecular functional annotations common in iGluRs were trans-membrane ion transporter, signal transduction, ligand binding and, interestingly, enzymatic hydrolase activity (GO: 0016787, class EC: 3 enzymes, which act on carbon-nitrogen, but not peptide bonds). The *G. m. morsitans GluRIB* and *GluRIA* had functional motifs that regulate peptidoglycan catabolic process, while *GluRIIA, GluRIIB, GluRIIC* and *GluRIIE* had functions for regulation of neurotransmitter secretion at the synapse and control of rhythmic excitations. The *G. m. morsitans NMDAR1* functional classes included long term memory, olfactory learning, Ca^{2+} ion homeostasis, photo-taxis, cell-cell signaling pathway and regulation of membrane potentials.

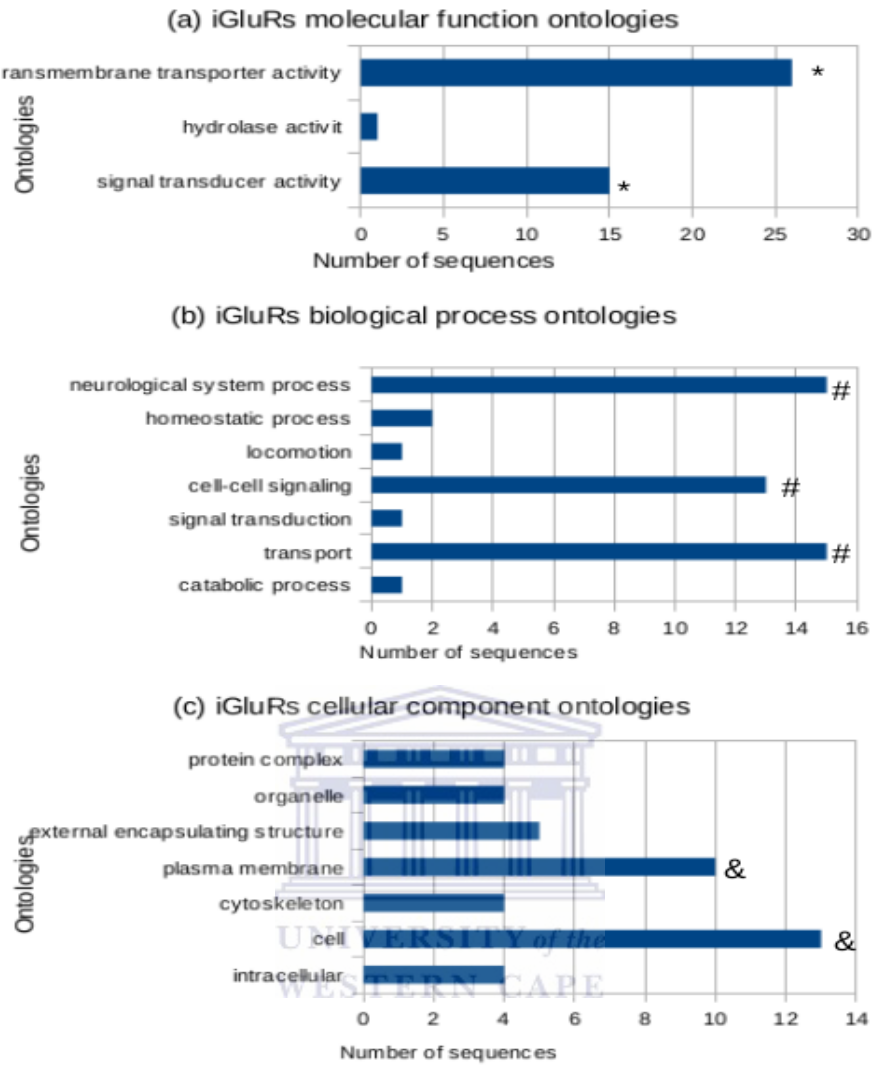


Figure 3.9 Gene ontology functional classifications of iGlu Rs in *G. m. morsitans*.
Functional classes with asterices (*) are the major molecular activities with likelihood that there are multiple functional motifs per sequence; # major biological process mediated; symbol & represent major expression sites of the proteins. The functional classifications were estimated via B2G web server (Conesa *et al*., 2005).

3.3.2 Repertoire of *G. m. morsitans* IRs

Nineteen IRs genes were identified and annotated in *G. m. morsitans* genome, of which 15 have since been assigned permanent identities and integrated into the VectorBase database, while four of them were manually curated and assigned temporary identities, TMP_IR* (Table 5). The *G. m. morsitans* IR genes were encoded on diverse scaffolds, each gene encoded by up-to ten exons except *GmmIR8a* and *IR76b* that had 16 exons each. This contrasts with fruit flies non-olfactory

IRs that are encoded as single exons (Ai *et al.*, 2010). The amino acids of the IRs ranged from 301 – 2832, with GmmIR8a and IR76b having over 2000 amino acids. Out of 19 IRs, 16 matched the ancestrally conserved antennal specific IRs, having one-on-one homologs in *D. melanogaster,* while three (*GmmIR10a, IR56b* and *IR56d*) had only low sequence similarity to species-specific divergent non-antennal homologs in the drosophila. Portions of three *G. m. morsitans* IRs (*GmmIR75b, IR75c* and *IR10a*) were located in undetermined genomic regions, and together with IR76a were annotated as incomplete since they lacked some conserved domains. All the *G. m. morsitans* IR genes exhibited high sequence diversity, sharing low sequence similarities except three genes - *GmmIR8a, IR25a* and *IR76b,* which were also highly conserved in *D. melanogaster.* One antennal IR gene lineage of *GmmIR84a* had three related copies (named here as *GmmIR84a-A, -B* and *-C*) located on different scaffolds. There was a putative splice variant detected at gene locus *GmmIR8a,* in which variant 2 had a split in its fourth exon (data not shown). There was no homolog of DmelIr93a (a glutamate-sensitive ion receptor reportedly conserved across all arthropods) identified in the *G. m. morsitans* genome.

All the 19 *G. m. morsitans* IR genes contained glutamate-gated receptor specific domains, PF00060 and PF10613 except IR56b-like, IR56d-like, IR75b, and IR76a that had no detectable domain (see Table 5). In addition, G.mmIR76b had six tandem copies of conserved neuralized protein domain family (IPR006573, PF07177) towards N-terminal and a multi-domain ligand-gated ion channel at the C-terminal. The longest IR gene GmmIR8a with extensive N-terminal end had no extra known functional domain. The IR75-family was represented by four genes - *GmmIR75a, IR75b, IR75c* and *IR75d.* Structurally, *GmmIR75a, IR75b* and *IR75c* were encoded in tandem but had poor RNA-seq read coverage; IR75c contained secretin and thiopeptide domains of bacterial origins at the N-terminal, and also increased presence of cysteine and serine residues. The IR75d had a highly conserved domain that belongs to the periplasmic-binding protein type 2, PBPII superfamily (cd13717) and ligand_channel domain (pfam00060).

All the *G. m. morsitans* IRs residues had shorter C-terminals and lacked the subunit assembly ATD domains, except in *G. m. morsitans* IR8a and IR25a that resembled their iGluR relatives. Like in iGluRs, a cysteine site (C site) after LBD S2 was conserved in all IRs except in GmmIR76b (G), GmmIR40a (G), and GmmIR21a (S). The C site was flanked by glycine-rich sites. Though the channel pore trans-membrane domains (TM1, P-loop/TM2, TM3, and TM4) and ligand-biding domains S1 and S2 were the most conserved in all the *G. m. morsitans* IRs, specific residues in S1 and S2 that interact with glutamate ligand were of diverse nature (see Appendix 10).

Table 5 Annotated ionotropic receptors (IRs) in *Glossina morsitans morsitans*

Glossina morsitans morsitans					*D. melanogaster*	BLAST2GO HITS		
Gene Name	**Identities**	**Exons**	**AA**	**Pfam Domain**	**Orthologs ID**	**Homologs**	**Similarity (%)**	**Description**
IR8a	GMOY012127	16	2832**	LC, LigC-G, SBP	CG32704	XP_002011930	76.9	Ionotropic receptor 8a
IR21a	GMOY006751	7	904	LC	CG2657	XP-001961590	77.5	Ionotropic receptor 21a
IR25a	GMOY001810	7	837	LC, LigC-G, SBP, IT	CG15627	AFP89966	91.6	Ionotropic receptor 25a
IR31a	GMOY012048	7	548	LC	CG31718	NP_723585	67.2	Ionotropic receptor isoform d
IR40a	GMOY004663	9	714	LC, LigC-G	CG42352	XP_002051306	70.7	Ionotropic receptor isoform f
IR64a	GMOY000804	10	743	LC, LigC-G, SBP, Sec	CG10633	XP_002007888	79.6	Ionotropic receptor 64a
IR68a	GMOY005753	4	572	LC, LigC-G	CG6185	XP_001956515	60.6	Ionotropic receptor 68a
IR75a	GMOY008540	6	529	LC	CG14585	XP_002008655	65.7	Ionotropic receptor 75a
IR75b&	**TMP_IR4**	4	301	...	...	XP_002068119	52.1	Ionotropic receptor 75a
IR75c&	**TMP_IR3**	6	541	LC	…	XP_002068119	68.5	Ionotropic receptor 75a
IR75d	GMOY007825	5	563	PBPII, LC	CG14076	XP_002042636	81.7	Ionotropic glutamate receptor
IR76a	GMOY008789	5	348	...	CG42584	XP_002050044	58.3	Iontropic receptor 41a#
IR76b	GMOY009750	16	2490**	LC, Neuralized	CG7385	XP_002008923	81.6	Neuralised-like protein 4-like
IR84a-A	GMOY002585	5	462	LC	CG10101	XP_001953250	50.9	Iontropic receptor 84a-A
IR84a-B	GMOY008188	5	424	LC	...	XP_002070349	60.4	Iontropic receptor 84a-B
IR84a-C	GMOY004518	5	581	LC	...	XP_001998709	60.7	Iontropic receptor 84a-C
IR10a-like&	GMOY004578	2	422	LigC-G	CG34143	XP_002055742	42.2	Ionotropic receptor partial
IR56b-like	**TMP_IR6**	2	413	...	CG15121	...	...	…
IR56d-like	**TMP_IR5**	1	640	...	CG15904	XP_002063825	40.1	Iontropic receptor 56d

Columns from left: *G. m. morsitans* assigned gene names, gene identities (**TMP_IR*** are manually curated and assigned identities that were not computationally predicted), number of gene exons, and AA- length of amino acids in mature protein. & denote incomplete genes. The gaps (…) indicate no data found. Pfam conserved domains: LC – Ligand-gated channel (Pfam00060); LigC-G – Lig_Chan_Glu_bdg (Pfam10613); SBP – SBP_bac_3 (Pfam00497); Sec – secretin (Pfam07655); Neuralized – ubiquitin-like ligase domain (Pfam07177). (CG*) - The *D. melanogaster* identities. BLAST2GO searches were done using *G. m. morsitans* protein sequences at e-value less than 0.0001 against non-redundant Swiss-Prot database reporting only the best hit identity, similarity percentage and the protein description. Double asterisks (**) – show disproportionate amino acid above average. # - evidence gleaned from PhylomeDB trees indicate Dmel/Ir41a and Dmel/Ir76a as duplicates after speciation from most recent common ancestor *Glossina* gene.

Distribution of gene ontology functions in *G. m. morsitans IRs* (Figure 3.10) revealed that *GmmIR8a* had kainate selective glutamate receptor molecular activity specific for sensory dendrites; *IR25a* had post-synaptic and cell junction molecular activities; while *IR64a* and *IR84a-B* had specific sensory dendrite odor detection activity. The *G. m. morsitans IR56d* had G-protein coupled receptor (GPCR) activity, corresponding to ontologies for the species-specific divergent antennal sub-group of IRs. Predictably, *G. m. morsitans* IR76b contained ontology functions for reproduction, developmental processes, and neuronal cell biogenesis by regulating pole plan oskar mRNA localization, oocyte microtubule cytoskeleton polarization, and cytoskeleton self-organization of the ion channel subunits into functional membrane structures – consistent with its drosophila homologs that participate in cell-fate determination in the notch pathway.

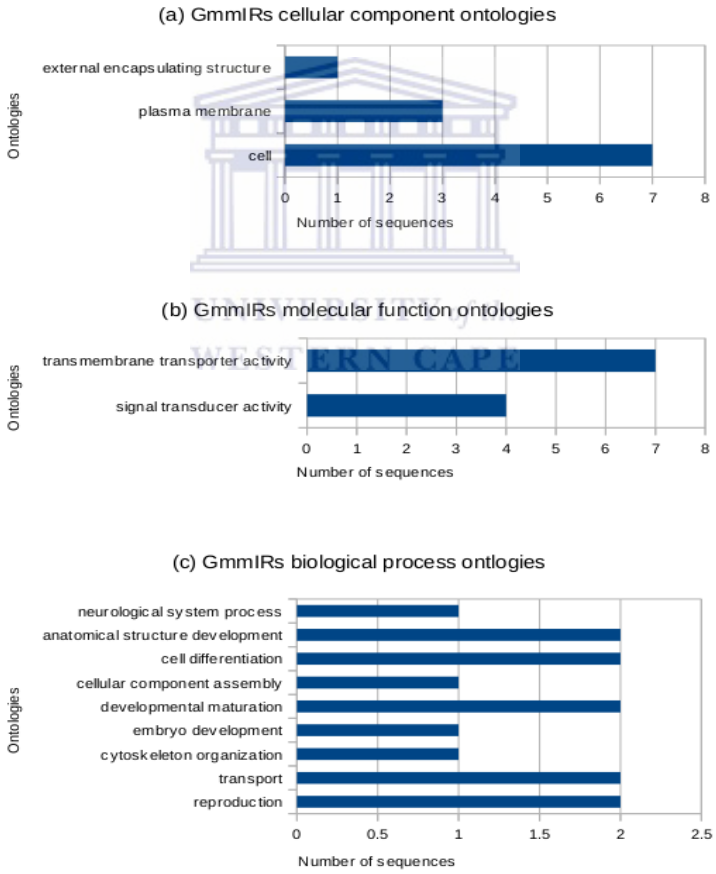


Figure 3.10 G**ene ontology functional classifications in GmmIRs.**
The IRs are membrane proteins with transmembrane transporter and signal transducer activities. They participate in neuronal system development processes. The functional classifications were estimated via B2G web server (Conesa *et al.*, 2005).

3.3.3 Repertoire of *G. m. morsitans* mGl uRs

Six *G. m. morsitans* mGluR-like genes named mGluRA, mGluR-a, mGluR-b, mGluR-c, mGluR-d, and mGluR-e were identified in the *G. m. morsitans* genome (Table 6). The *G. m. morsitans* mGluRs genes contained multi-exons ranging 7-16, and amino acid lengths ranging 522-1312. Conserved domains among *G. m. morsitans* mGluRs included G-protein coupled receptor GPCR family 3 (IPR000337), some members had nine cysteines (NCD3G, IPR011500), and periplasmic binding protein type1 (PBP 1) (IPR028082) - an extracellular ligand-binding receptor (IPR0001828) often coupled with atrial natriuretic factor, ANF (PF01094). Others had GPCR metabotropic glutamate receptor family (IPR00248 or IPR000162) and/or predicted metabotropic glutamate receptor (PTHR24060).

Table 6 Annotated metabotropic glutamate-gated ion receptors (mGluRs) in *Glossina morsitans morsitans*

Glossina morsitans morsitans					*D. melanogaster*	BLAST2GO HITS	
Gene Name	**Identities**	**Exons**	**AA**	**Pfam Domain**	**Orthologs ID**	**Homologs**	**Description**
mGluRA	GMOY000333	7	1312	ANF, PBP1, 7tm-3, NCD3G	CG11144	XP_001982701.1	Metabotropic glutamate receptor
mGluR-a	GMOY003230	12	1030	PBP, 7tm-3	...	XP_004518781.1	Uncharacterised proein
mGluR-b	GMOY005828	16	1279	ANF, Pk, HNO, PBP6	...	AAM94353.1	Guanyly cyclase receptor
mGluR-c	GMOY010637	3	1024	ANF, PBP1	...	XP_001958833.1	GPCR Glutamate receptor
mGluR-d	GMOY010638	5	522	PBP1, NCD3G	...	CAE46392.1	Metabotropic X receptor
mGluR-e	GMOY010639	16	835	PBP1, 7tm-3, NCD3G	...	XP_004534864.1	Met. Glutamate receptor 4

Columns from left: *G. m. morsitans* assigned gene names, gene identities, number of protein coding exons, and AA- length of amino acids in mature protein. Pfam conserved domains: ANF – ANF_receptor (atrial natriuretic factor) family binding region (Pfam01094); PBP – Periplasmic binding protein domain typical of bacteria; NCD3G – nine cysteines domain of family 3 GPCR (Pfam07562); 7tm-3 – Class C GPCR super-family taste receptor domain (Pfam00003); Pk – Pkinase (Pfam00069); HNO – Heme NO binding associated (Pfam07701). The *D. melanogaster* query ortholog (CG11144). BLAST2GO searches were done using *G. m. morsitans* protein sequences at e-value less than 0.0001 via non-redundant Swiss-Prot database reporting only the best hit identity, hit species and the protein description.

The *G. m. morsitans* mGluRs had gene ontology functions commonly associated with CNS neuronal genes, similar to those occurring in iGluRs. In addition to glutamate binding and receptor activities, and G-protein coupled receptor signaling pathway (for phospholipase C) that were also found in iGluRs, the *G. m. morsitans* mGluRs contained ontologies for regulating adult feeding behavior, response to insecticide, neuromuscular junction development, and synaptic transmission. In addition, they also had ontologies for positive regulation of cell proliferation and mitotic cell cycle with guanylate cyclase and protein tyrosine kinase molecular activities (specifically mGluRb) (Figure 3.11).

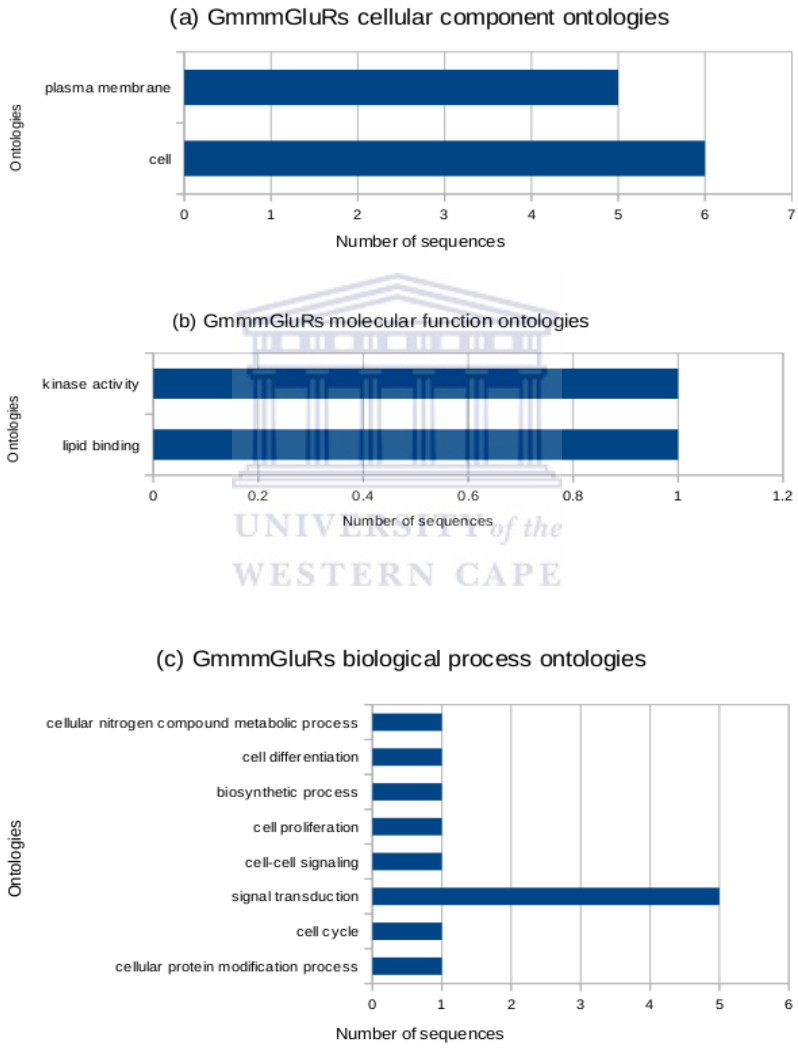


Figure 3.11 Gene ontology functional classifications of mGluRs in *G. m. morsitans*.
The mGluRs are transmembrane proteins annotated with kinase activity and lipid binding, and the main functional biological process as signal transduction. The functional classifications were estimated via B2G web server (Conesa *et al.*, 2005).

3.3.4 Expression levels of IRs, iGluRs, and mGluRs genes from *G. m. morsitans*

Comparatively, there were fewer RNA-seq reads supporting expression of peripheral glutamate receptor *GmmIR*s (n = 48,930.93 RPKM) than the amounts for *iGluRs* (n = 314,643.84 RPKM) and *mGluRs* (n = 358,470.69 RPKM) (Figure 3.12). This reveals differential gene expression between receptors located in peripheral sensory organs, which could be more adaptive to different environments, and those known to be ubiquitously expressed in the CNS. The antennal *G. m. morsitans IR8a* (17,377.45 RPKM), *IR64a* (9,739.9 RPKM), *IR76b* (12,598.34 RPKM), *IR84a-B* (2,637.09 RPKM) and *IR84a-C* (1,703.69 RPKM) had higher expression levels (Figure 3.12). The remaining *G. m. morsitans* IRs had expression profiles below 1000 RPKM. The *GmmIR8a* and *GmmIR25a* encode receptors thought to act as co-receptors, and thus one would expect them to have a higher expression profiles. However, the expression level of *GmmIR25a* (1,656.52 RPKM) was lower than the level for *GmmIR8a* (17,377.45 RPKM) relative to n = 48,930.93 RPKM - but this difference could be attributed to the long N-terminal of IR8a whose role is yet unknown.

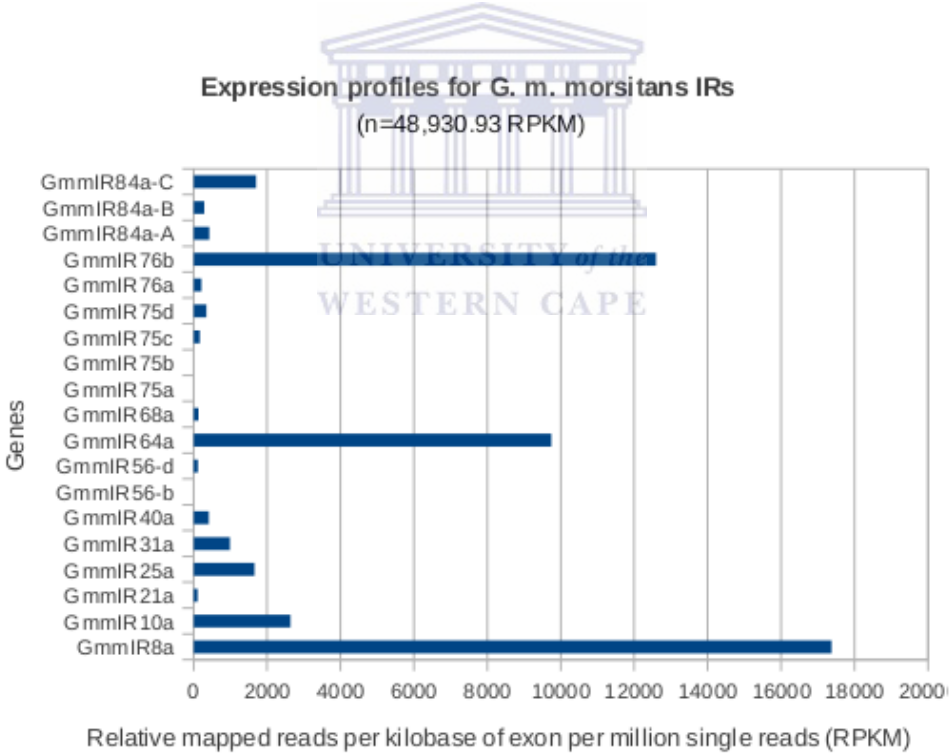


Figure 3.12 Expression levels of IRs genes in *G. m. morsitans*.
The overall transcripts for IRs n=48931 RPKM. The expression levels were estimated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

Most *G. m. morsitans iGluRs* had high expression levels; *GmmNMDAR1* had 56,088.49 RPKM profile, *GmmKaiR-d* 55,543.29 RPKM, *GmmKaiR1A* 52,877.45 RPKM, *GmmClumsy* 35,328.9 RPKM, and *GmmKaiR-c* 27,057.59 RPKM. The remaining *G. m. morsitans* iGluRs had expression levels ranging 0 - 11,272.86 RPKM (Figure 3.13). Similarly, GmmmGluRs exhibited high expression levels (Figure 3.14).

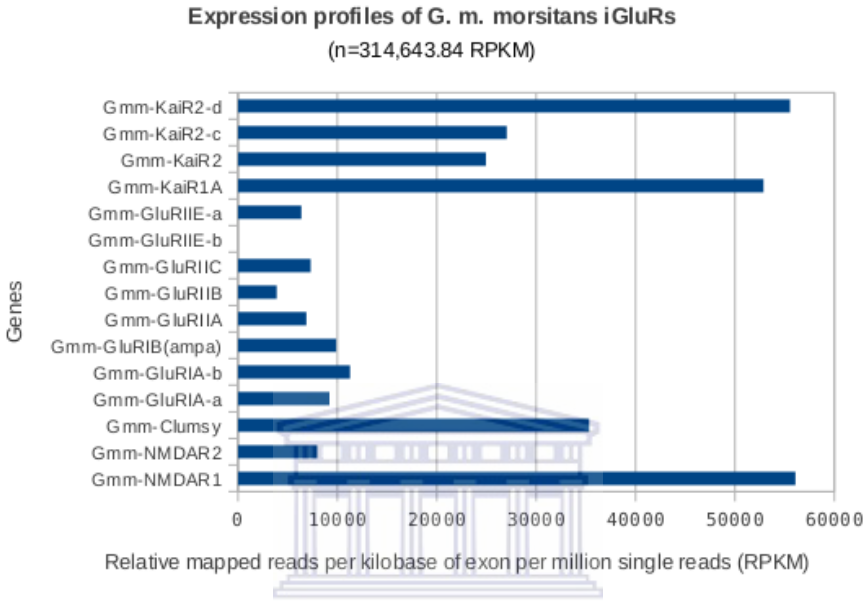


Figure 3.13 Expression levels of iGluR genes in *G. m. morsitans*.
The overall expression transcripts for iGluRs n=226438 RPKM. The expression levels were estimated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

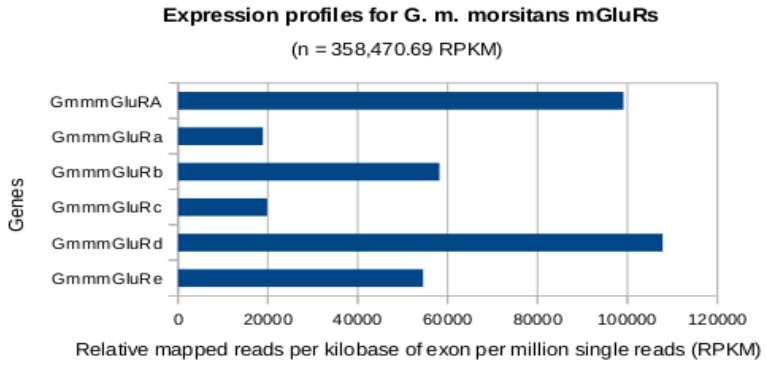


Figure 3.14 Expression levels of mGluRs genes in *Glossina m. morsitans*.
The overall expression transcripts for mGluRs n=358470 RPKM. The expression levels were estimated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

3.3.5 Phylogenetic relationships of IRs, iGluRs and mGluRs in *G. m. morsitans* with homolgs from *D. melanogaster* and *An. gambiae*

Maximum likelihood phylogenetic tree revealed majority of the *G. m. morsitans* iGluRs and IRs clustered correctly with their homologs from *D. melanogaster* and *An. gambiae*, node branches supported by over 70% bootstraps at 500 replications (Figure 3.15). There were four distinct clusters for antennal IRs: (i) the two co-receptors - IR8a and IR25a; (ii) nine receptors - IR31a , IR64a, IR75a, IR75b, IR75c, IR75 d, and IR84a-A, IR84a-B, IR84a-C; (iii) six receptors - IR21a, IR40a, IR68a, IR76b, and (iv) the species-specific divergent group with IR10, IR56b and IR56d. The *G. m. morsitans* iGluRs had two sub-clusters: (i) NMDAR1 and NMDAR2; and (ii) non-NMDAR including AMPA-like receptor – GluRIB; kainate receptors GluRIA-a, -b; GluRIIE-a, -b, KaiR2, KaiR2-like-c, -d, KaiRIA, Clumsy, GluRIIA, GluRIIB and GluRIIC. The *G. m. morsitans* mGluRs clustered into three sub-branching (blue branches). Unexpectedly, *G. m. morsitans* IR76a clustered deeply in the tree with the mGluRs and not with its ortholog DmelIr76a despite having the best non-redundant protein database homology *D. melanogaster* (ACN81897.1) at e-values 5e-44 and sequence identity 31%, both of which correspond to *D. melanogaster* Ir76a (Figure 3.16). The *G. m. morsitans* IR76b clustered with their drosophila *blue* gene product including out-group species from the phylome database (data not shown).

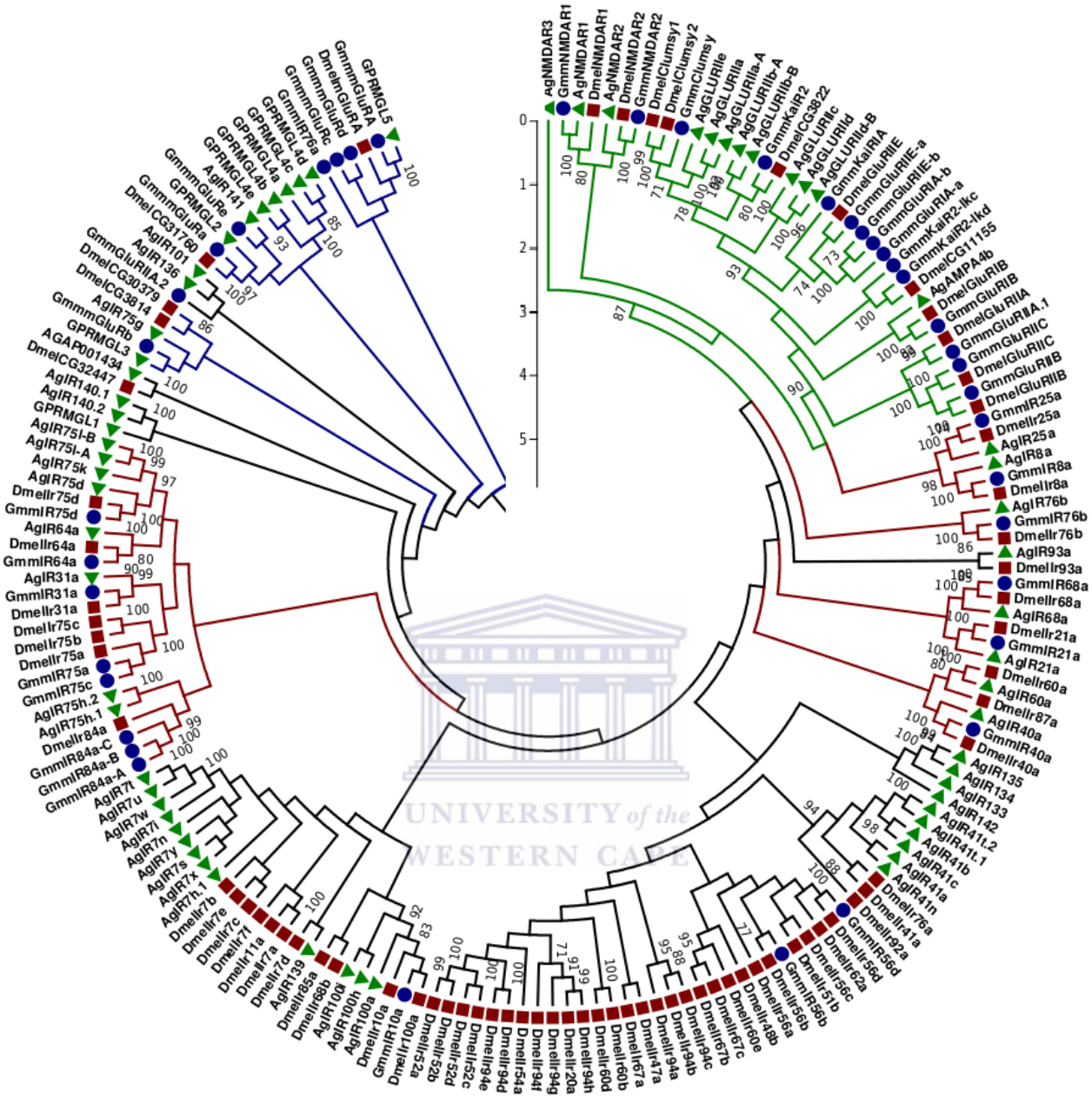


Figure 3.15 Maximum likelihood phylogenetic tree for glutamate-gated receptors of *G. m. morsitans* with homologs from *D. melanogaster* and *An. gambiae*.

The green branches show the iGluRs clusters; the red branches represent IRs orthologs, with the co-receptor IR8a and IR25a branching from the iGluRs and they form two distinct groups; while sky blue cluster contain the mGluRs, and surprisingly with GmmGluRIIA.2. Multiple sequence alignment was done using MUSCLE tool (Edgar, 2004) and edited via Jalview tool (Waterhouse *et al.*, 2009). The distance tree was constructed using MEGA5 tool (Tamura et al., 2011).

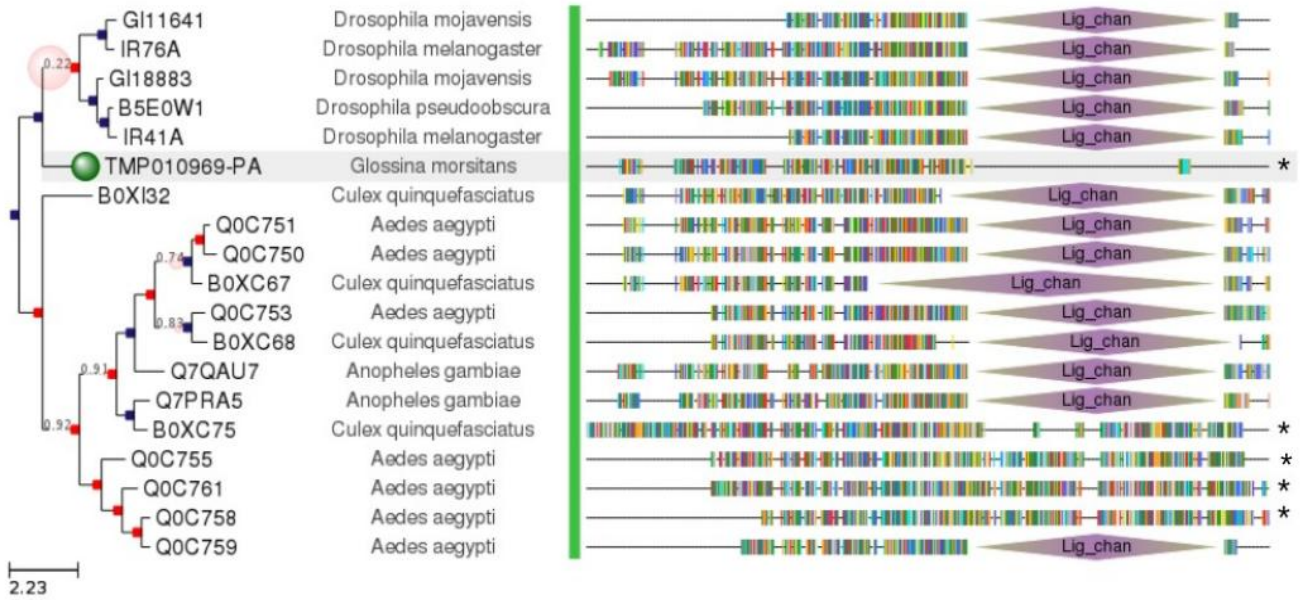


Figure 3.16 Phylome tree clustering of GmmIR76a (TMP010968-PA).
Green ball - the GmmIR76a (TMP010969-PA); * - lack conserved Lig_chan domain. Mosquito species have multiple copies of IR76a homolog, while among the fruit flies, the *D. melanogaster* have two closely related members (IR76a and IR41a). The nodes marked blue are speciation events while those marked red indicate duplication events. The assessment was done via phylome database (Huerta-Cepas *et al*., 2011).

3.4. Annotated chemosensory responsive genes – OBPs, CSPs and CD36-like SNMPs in *G. m. morsitans*

3.4.1 Repertoire of *Glossina m. morsitans* OBPs

There were 32 classical full length recoverable genes for OBPs in *G. m. morsitans* genome, named OBP1-29, with three likely duplicates - OBP2A/B, OBP5A/B, and OBP8A/B (Table 7). This study recovered 12 extra loci in the genome (bold typed), thus improving the 20 OBPs previously identified from transcriptome library by Liu *et al*. (2010). Though all the OBP genes were computationally predicted, some were edited by splitting those that had double functional domains to form independent genes– for instance: *G. m. morsitans*, OBP2A/B, OBP5A/B, OBP8A/B, and OBP20/21, which were encoded in tandem on their respective scaffolds. This study provides the complete coding sequences of all the OBPs, further extending the completeness of those reported in Liu et al. The OBP genes in other insects usually have molecular weights between 15 and 20 kDa, but in *G. m. morsitans*, some OBPs had below 15kDa (OBP15, OBP16, OBP23 and OBP26) and others had above 20kDa (OBP3, OBP7, OBP8B, OBP20, OBP27, OBP28 and OBP29) (Table 7). This study also identified a gene that encodes a LUSH protein homolog, *GmmOBP26* (Appendix 11).

The *G. m. morsitans* OBP genes were small proteins encoded by up-to five exons. Most *G. m. morsitans* OBP

residues had below 200 amino acids, with exceptions above 200 residues. Out of the 32 *G. m. morsitans* OBPs, 23 (representing 70%) had predicted secretory signal peptides within their first 24 N-terminal amino acids. All the *G. m. morsitans* OBPs had the conserved domain signature pheromone binding protein/general odorant binding protein (PBP/GOBP, pfam01395), with an insect-specific pheromone binding motif signature (smart00708). However, *G. m. morsitans* OBP7 had, in addition, anillin superfamily signature that is implicated in stabilization of septins during cell division.. The *G. m. morsitans* OBPs functional classifications included: GOBP/PBP (general odorant binding protein/pheromone binding protein) OBPs - *GmmOBP2B, OBP4, OBP5B, OBP6, OBP8A, OBP8B, OBP9, OBP10, OBP11-16, OBP18* and *OBP24-29* – (*OBP8A, OBP8B, OBP9, OBP10*are identifiable by the presence of a conserved proline site after the sixth cysteine site, Figure 3.17); Plus C OBPs - *GmmOBP5A, OBP7* and *OBP23*; and Minus C OBPs - *GmmOBP1, OBP2A, OBP3, OBP17, OBP19, OBP20, OBP21 and OBP22* . *GmmOBP13, OBP26* and *OBP28* had functional classifications for response to pheromone in olfactory behavior. In fact, *OBP26* had specific classification with potential to participate in courtship behaviors with binding motifs for dibutyl phthalate, diphenyl phthalate and ethanol (data not shown).

Table 7 Annotated odorant binding protein (OBP) genes in *G. m. morsitans*

Name	Scaffold[&]	Identity	Length (aa)	Exons	Mol. Wt. (kDa)	BlastP homolog (e-value)	Signal P (N-terminal)	Classification
GmmOBP1	640662	GMOY000890	141	2	16.31	GmmOBP1 (2e-37)	1-17	Minus C
GmmOBP2A	644614	GMOY002825	144	2	16.83	Dsim-GD21452 (2e-14)	1-18	Minus C
GmmOBP2B	648638	GMOY002826	153	2	17.9	GmmOBP2 (3e-44)	1-18	GOBP
GmmOBP3	650660	GMOY005549	284	3	33.93	GmmOBP3 (4e-61)	Null	Minus C
GmmOBP4	648638	GMOY007757	163	1	20	GmmOBP4 (2e-59)	1-17	GOBP
GmmOBP5A	649084	GMOY006521	184	3	20.87	GmmOBP5 (7e-61)	1-21	Plus C
GmmOBP5B	649084	GMOY006522	157	4	18.18	DpseuOBP19b (2e-23)	1-23	GOBP
GmmOBP6	651846	GMOY009708	145	3	16.02	GmmOBP6 (3e-38)	1-19	GOBP
GmmOBP7	648638	GMOY005548	240	3	28.32	GmmOBP7 (1e-75)	1-19	Plus C
GmmOBP8A	648041	GMOY004317	150	4	17.48	GmmOBP8 (3e-52)	1-24	GOBP/PBP
GmmOBP8B	648041	GMOY004316	260	5	30.06	AgamOBP43 (6e-30)	1-23	GOBP/PBP
GmmOBP9	648462	GMOY005184	150	4	17.41	GmmOBP9 (6e-54)	1-26	GOBP/PBP
GmmOBP10	639213	GMOY012275	152	1	17.54	GmmOBP10 (3e-10)	Null	GOBP/PBP
GmmOBP11	648638	GMOY005550	140	2	16.87	DwilGK14209 (3e-31)	1-17	GOBP
GmmOBP12	648462	GMOY005184	160	4	17.52	GmmOBP9 (6e-73)	1-19	GOBP/Os-E
GmmOBP13	644671	GMOY002859	134	2	15.11	MdomOBP99a (7e-23)	1-20	GOBP
GmmOBP14	649084	GMOY006523	149	1	16.59	GmmOBP14 (5e-23)	1-22	GOBP
GmmOBP15	650660	GMOY012229	119	1	13.87	GmmOBP15 (6e-20)	Null	GOBP
GmmOBP16	648453	GMOY005163	115	2	13.59	DmelOBP99a (2e-17)	Null	GOBP
GmmOBP17	648833	GMOY012281	126	1	15.79	GmmOBP17 (3e-28)	Null	Minus C
GmmOBP18	647856	GMOY003978	144	5	16.72	DmelOBP83a (2e-24)	1-20	GOBP
GmmOBP19	650289	GMOY007314	138	1	14.75	AgamOBP24 (5e-09)	1-19	Minus C
GmmOBP20	649017	GMOY006417	261	3	30.51	GmmOBP20 (3e-39)	1-19	Minus C
GmmOBP21	649017	GMOY006418	150	2	17.52	GmmOBP21 (1e-30)	Null	Minus C
GmmOBP22	641423	GMOY001476	143	2	17.32	GmmOBP22 (2e-30)	Null	Minus C
GmmOBP23	640257	GMOY000657	125	5	14.25	DmelOBP84a (9e-19)	Null	Plus C

GmmOBP24	648567	GMOY005400	144	5	15.81	ScalOBP- (4e-26)	1-21	GOBP
GmmOBP25	648778	GMOY005876	134	2	15.2	Dmel OBP56d (1e-17)	1-18	GOBP
GmmOBP26	648792	GMOY005931	114	4	12.95	GmmOBP2 (2e-18)	Null	GOBP
GmmOBP27	648825	GMOY006081	211	5	25.1	DmelOBP59a (7e-40)	1-15	GOBP
GmmOBP28	650245	GMOY012237	149	1	8.98	DantOBP5 (8e-27)	1-22	GOBP
GmmOBP29	650279	GMOY007293	138	3	25.9	DpseuOBP99a (2e-08)	1-18	GOBP

Bold face are novel genes identified in this study in addition to those previously identified by Liu et al. & the scaffold identities preceeded by Scf7180000*. Classification of *G. m. morsitans* OBPs: GOBP (include Atypical OBPs and putative PBPs with conserved proline sites after C6), OBPs with 6 Cs = 20; Plus C OBPs =3; Minus C OBPs (less than 6 Cs) = 8. Reciprocal blastat e-value cut off 0.0001: Gmm – *Glossina morsitans morsitans;* Agam – *Anopheles gambiae*; Dmel – *Drosophila melanogaster*; Mdom – *Musca domestica*; Dwil – *Drosophila willistoni*; ; Dpseu – *Drosophila pseudoobcura*; Dsim – *Drosophila simulans*; Dant – *Delia antiqua.* Protein molecular weight predictions were done using sequence manipulation suite (SMS) (http://www.bioinformatics.org/sms/prot_mw.html).

Glossina m. morsitans OBPs conserved cysteine sites were variable, majority of them having the standard six C-sites while others either had fewer or additional C-sites at other locations not conserved - potentially compensating for the missing sites. Nonetheless, *G. m. morsitans* OBPs had the typical cysteine motif signature C1-X_{26-80}-C2-X_3-C3-X_{28-45}-C4-X_{8-21}-C5-X_8-C6 (Figure 3.17). This means the tertiary structures of these proteins could be able to form three disulfide bridges as their counterparts in *D. melanogaster* and *An. gambiae*.

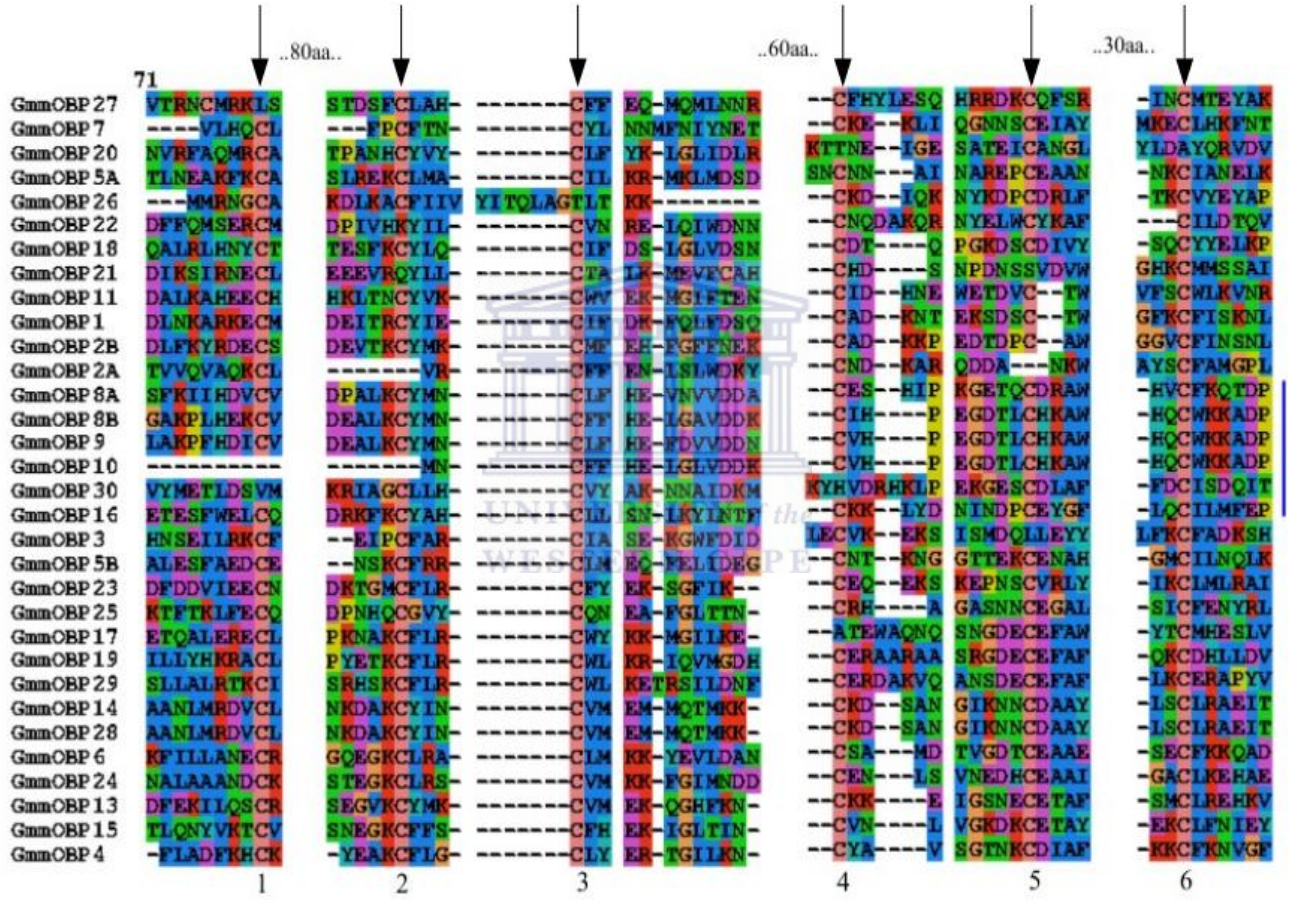


Figure 3.17 Truncated multiple sequence alignment of OBPs in *G. m. morsitans*.
The conserved cysteine sites are numbered respectively 1-6 (downward arrows). The amino acids truncated in between the C-site alignments are written on top the alignments. *G. m. morsitans* OBPs with extra C sites include GmmOBP3, OBP7, OBP8B, OBP10, OBP13, OBP17, OBP20, OBP21, OBP24, and OBP27. Blue vertical line (on the extreme right) on OBP8A, OBP8B, OBP9, OBP10 and OBP16, represent putative pheromone-specific PBP classification. The sequences were aligned using MUSCLE tool (Edgar, 2004) and viewed using Jalview tool (Waterhouse *et al.*, 2009).

Out of 32 *G. m. morsitans* OBPs, three genes *OBP20*, *OBP28*, and *OBP29* had no presence of gene ontology functional classes detected (Appendix 12). All the *G. m. morsitans* OBPs had ontologies for odorant binding (GO:0005549), with specificity for sensory perception of smell, suggesting that all the *G. m. morsitans* OBPs may participate in the antennal sensory system. The blast search revealed *G. m. morsitans OBP8A* as a homolog of pheromone binding protein 6-like, and *OBP26* as a homolog of general odorant binding LUSH-like protein.

3.4.2 Repertoire of *Glossina m. morsitans* CSP genes

Five copies of *G. m. morsitans* CSPs (*CSP1-5*) encoding full length proteins were identified in the genome (Table 8). There was one putative alternative splice variant detected in CSP2, in which the second exon could be encoded on different reading frames, thus generating slight changes in amino acid set. Nevertheless, the numbers of *G. m. morsitans* CSPs genes were similar to those earlier identified via cDNA and EST libraries in the same *G. m. morsitans* species (Liu *et al.*, 2012), and also comparable to four and eight CSP genes in *D. melanogaster* and *An. gambiae* respectively. The mature residues ranged 108 – 178 amino acids, encoded by up to five exons. The *G. m. morsitans* CSP1 and CSP3 were encoded in tandem on one scaffold (on which CSP4 was also located), while CSP2 and CSP5 were on their individual scaffolds. The *G. m. morsitans* CSP1 and CSP2 lacked signal peptides, the rest of CSPs had the peptide signals ranging 19 - 26 amino acids from their N-terminal.

Table 8 Annotated chemosensory-specific protein (CSP) genes in *G. m. morsitans*

Name	Scaffold	Identity	Length (aa)	Exons	Mol. Wt. (kDa)	Blast homolog (e-value)	Signal P (N-terminal)
GmmCSP1	652157	GMOY012164	192	3	22.41	GmmCSP1 (2e-67)	null
GmmCSP2a	652014	GMOY010026	178	2	20.16	GmmCSP2 (2e-95)	null
GmmCSP2b	652014	GMOY010026	173	2	19.85	GmmCSP2 (2e-49)	null
GmmCSP3	652157	GMOY012165	128	1	14.63	GmmCSP3 (4e-69)	1-19
GmmCSP4	652157	GMOY010874	123	2	14.00	GmmCSP4 (1e-61)	1-21
GmmCSP5	651861	GMOY009731	108	2	12.42	Cqui-gb|EDS38047.1 (5e-38)	1-26

Columns from left: gene names, scaffold identities preceded by Scf7180000*, assigned identities, length of mature protein, number of exons per gene, best Delta Blast homologs, and signal peptide. GmmCSP2 had potential splice variants supported by RNA-seq reads, only differing in terms of reading frames thus giving different sets of amino acids for exon 2. Reciprocal Delta Blast done using annotated amino acids at e-value cut off 0.0001 against non-redundant Swiss-Prot database: Gmm – *Glossina morsitans morsitans;* Cqui – *Culex quinquefasciatus* Signal peptides were determined via Signal P v4.1 server (www.cbs.dtu.dk/services/SignalP/). Protein molecular weight predictions were done using sequence manipulation suite http://www.bioinformatics.org/sms/prot_mw.html).

All the GmmCSPs had four absolutely conserved C sites, typical of regular expression of C1-X_6-C2-$X_{6\text{-}18}$-C3-X_2-C4; they also contained the conserved odorant sensitive domain (OS-D) signature (pfam03392), an insect olfactory-specific pheromone binding protein domain (Figure 3.18). Gene ontology functions were detected in *G. m. morsitans* CSP1, CSP2, CSP3 and CSP4, but not in CSP5. The *G. m. morsitans* CSP2 had functional classsification for pheromone binding (GO:0005550) and sensory perception of chemical stimulus (GO:0007606). The CSP3 and CSP4 were both homologous to an ejaculatory bulb-specific protein 3-like, PebIII (IPR005055) and had motifs with potential to participate in anatomical structural development via metamorphosis process (GO:0007552), and molecular activity of stress response against viral infection (GO:0009615). These functional motifs confirm that *G. m. morsitans* CSP expression may not be restricted to the chemosensory system only (see also Appendix 12).

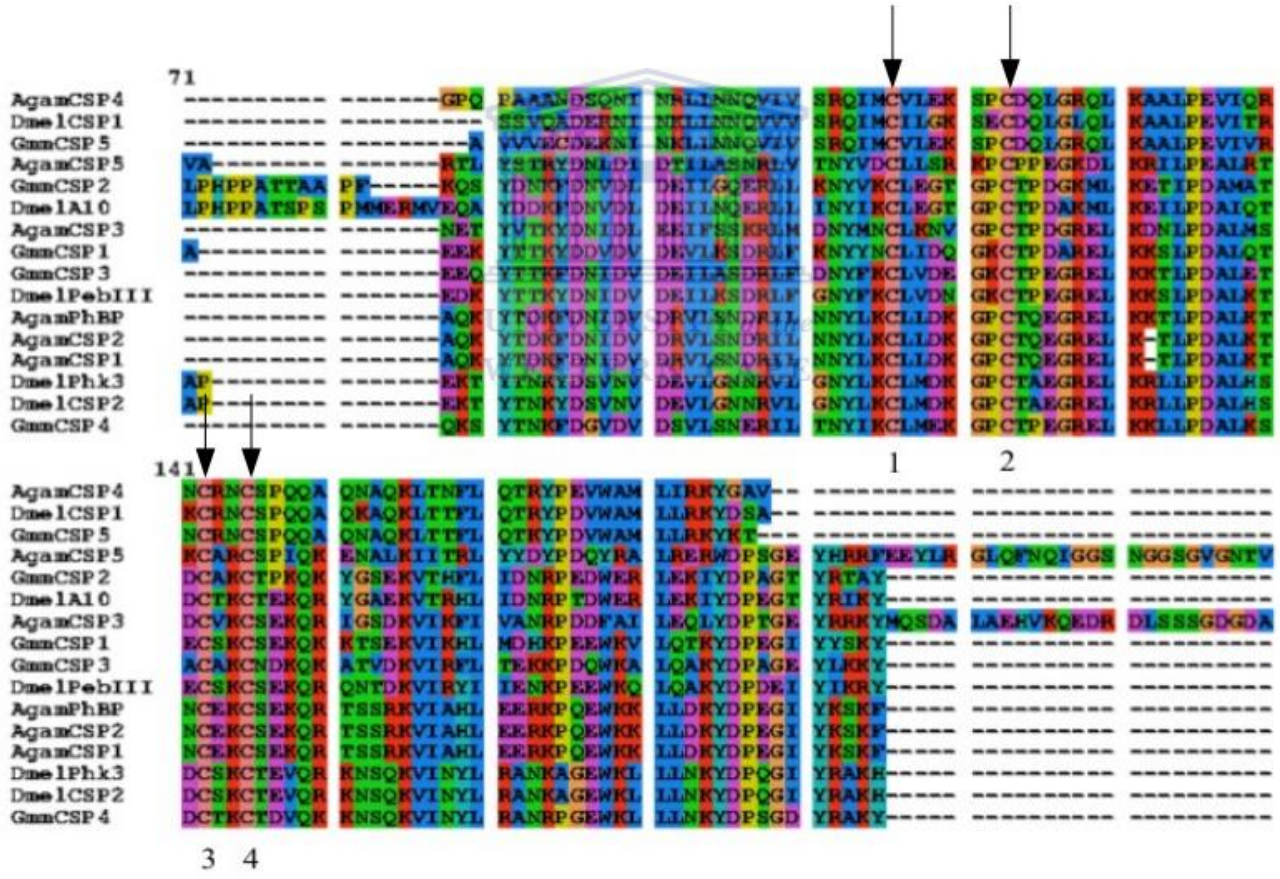


Figure 3.18 Truncated multiple sequence alignment of CSP proteins for *G. m. morsitans, D. melanogaster and An. gambiae*.
The cysteine sites are numbered respectively 1-4 below the columns (shown by arrows). The sequences were aligned using MUSCLE tool (Edgar, 2004) and edited using Jalview (Waterhouse et al. 2009).

3.4.3 Repertoire of *Glossina m. morsitans* CD36-like genes (including SNMPs)

There were 15 CD36-like genes recovered in *G. m. morsitans* genome, of which two were annotated as SNMPs (*SNMP1* and *SNMP2*), a similar number as in other insects. *Glossina m. morsitans CD36-6* was annotated with an alternative splice variant (data not shown). Two gene sets were encoded in close proximity to each other on their respective scaffolds: *GmmCD36-3, CD36-4, CD36-5, CD36-6* (scf7180000648975) and *CD36-9, CD36-10, CD36-11* (scf7180000652157). Additionally, structural examination revealed that *GmmCD36-9, CD36-10* and *CD36-11* gene sets were flanked on both sides by chemosensory-specific protein genes CSP4 and *CSP1*. All the *G. m. morsitans* CD36-like genes encoded large proteins of between 414 and 633 amino acids with 4-10 exons. The molecular weights of *G. m. morsitans* CD36-like receptors averaged 60.9kDa. The *Glossina* CD36 genes were almost identical (e-value 0.00) to homologs in *C. capitata, M. domestica,* and drosophila species (Table 9).

Table 9 Annotated sensory neuron membrane protein (SNMP) genes in *G. m. morsitans*

Name	Scaffold[&]	Identity	Length (aa)	Exons	Mol. Wt. (kDa)	Blast homolog (e-value)
GmmSNMP1	644980	GMOY002994	540	7	61.01	DgrimSNMP1 (5e-177)
GmmSNMP2	648879	GMOY006180	523	8	61.10	DmelSNMP2 (0.00)
GmmCD36-3	648975	GMOY006342	535	7	61.89	Dmelpestle (0.00)
GmmCD36-4.1	648975	GMOY006344	489	6	56.08	Mdomcroq (0.00)
GmmCD36-4.2	648975	GMOY006344	472	5	55.54	MdomCroq (8e-78)
GmmCD36-6.1	648975	GMOY006345	448	6	62.72	CcapCroq (0.00)
GmmCD36-6.2	648975	GMOY006345	414	4	47.17	CcapCroq (0.00)
GmmCD36-8	648454	GMOY005165	502	5	58.01	Ccap CroqA (0.00)
GmmCD36-9	652157	GMOY012013	532	6	57.78	GmmSCR (0.00)
GmmCD36-10	652157	GMOY010875	540	7	63.45	MdomSCRB1 (0.00)
GmmCD36-11	652157	GMOY010881	625	6	72.50	GmmSCR (0.00)
GmmCD36-12	642607	GMOY002035	557	7	63.54	MdomSCRB1 (0.00)
GmmCD36-13	647496	GMOY003843	579	8	66.12	DwilSCRB1 (0.00)
GmmCD36-14	649782	GMOY006978	477	5	53.79	DmelninaD (0.00)
GmmCD36-15	648956	GMOY006317	633	10	73.30	MdomSCRB1 (0.00)

Bold types are the *G. m. morsitans* SNMPs. & - scaffold identities preceded by Scf7180000Reciprocal blast e-value cut off 0.0001 against non-redundant Swiss-Prot database. Gmm – *Glossina morsitans morsitans;* Dmel – *Drosophila melanogaster*; Ccap – *Ceratitis capitata*; Mdom – *Musca domestica*; Dwil – *Drosophila willistoni* ; Dgrim – *Drosophila grimshawi.* Protein molecular weight predictions were done using sequence manipulation suite

(http://www.bioinformatics.org/sms/prot_mw.html).

All the *G. m. morsitans* CD36-like receptors had the functional domain signature CD36 (pfam01130) typical of scavenger receptor B1 sub-class and insect-specific Croquermort glycolipid-like receptors. They also had the characteristic two trans-membrane alpha helices located towards each of the sequence termini, coincident with regions having high sequence identity. However, in *G. m. morsitans* SNMP2, all of the predicted helices were located at the C-terminal end (see Appendix 13). Instead of 10 known CD36 cysteine residues in vertebrates, the *G. m. morsitans* CD36-like proteins had six extracellular cysteine sites in proline-rich segment, with regular expression C1-X_{28-32}-C2-X_{38-40}-C3-X_{1}-C4-X_{3-13}-C5-X_{10}-C6 (Figure 3.19). Further, *G. m. morsitans* SNMP1 lacked any of the C terminal palmitoylation sites necessary for proper anchorage on the membrane, while SNMP2 had one at C497 at C-terminal. In addition, *G. m. morsitans* SNMP1 receptors had five potential asparagine N-glycosylated motif sites, while SNMP2, surprisingly, had 31 sites concentrated in the first half of the sequence (Figure 3.20). Ontology analyses revealed CD36-3, CD36-4, CD36-5, CD36-6, CD36-7, and CD36-14 as likely homologs to fruit fly Croquermort gene; CD36-9, CD36-12, CD36-13, and CD36-15 are homologous to scavenger receptor b-like. Functional annotations in the *G. m. morsitans* CD36-like genes included motifs for molecular response to stress, participation in apoptotic signaling pathway, phagocytosis and photo-transduction among others. Specifically, *G. m. morsitans* SNMP1 and SNMP2 had ontology functions for components of neuronal cell membrane (GO:0005886; GO:0005887) and participation in trans-membrane signaling receptor activity (GO:0004888) (see Figure S9).

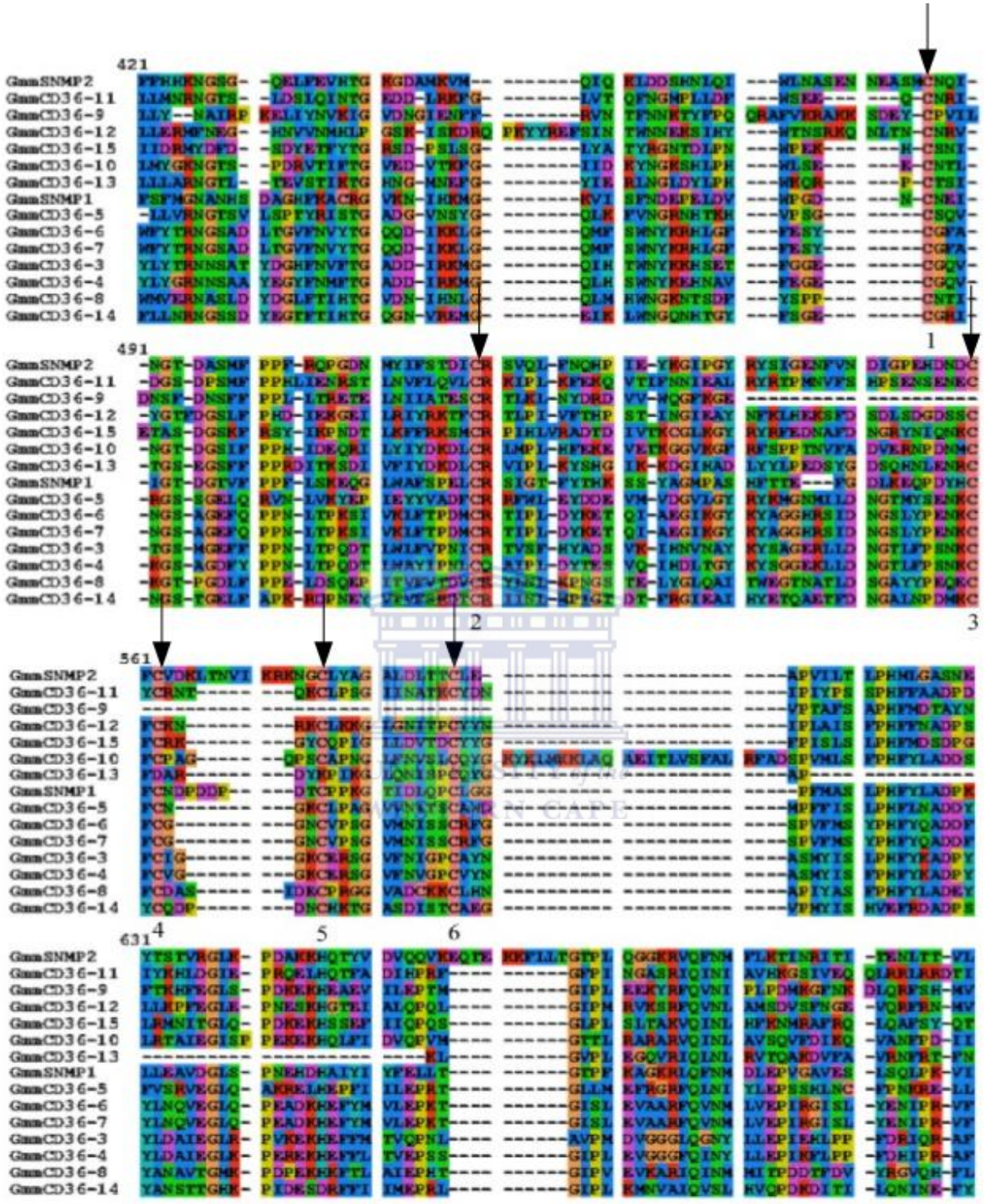


Figure 3.19 Truncated multiple sequence alignment of CD36-like proteins in *G. m. morsitans*.
The cysteine sites are numerated below the respective columns 1-6 (shown with arrows). The cysteine conserved sites signature is C1-X_{28-32}-C2-X_{38-40}-C3-X_{1}-C4-X_{3-13}-C5-X_{10}-C6. The sequences were aligned using MUSCLE tool (Edgar, 2004) and edited using Jalview (Waterhouse et al. 2009).

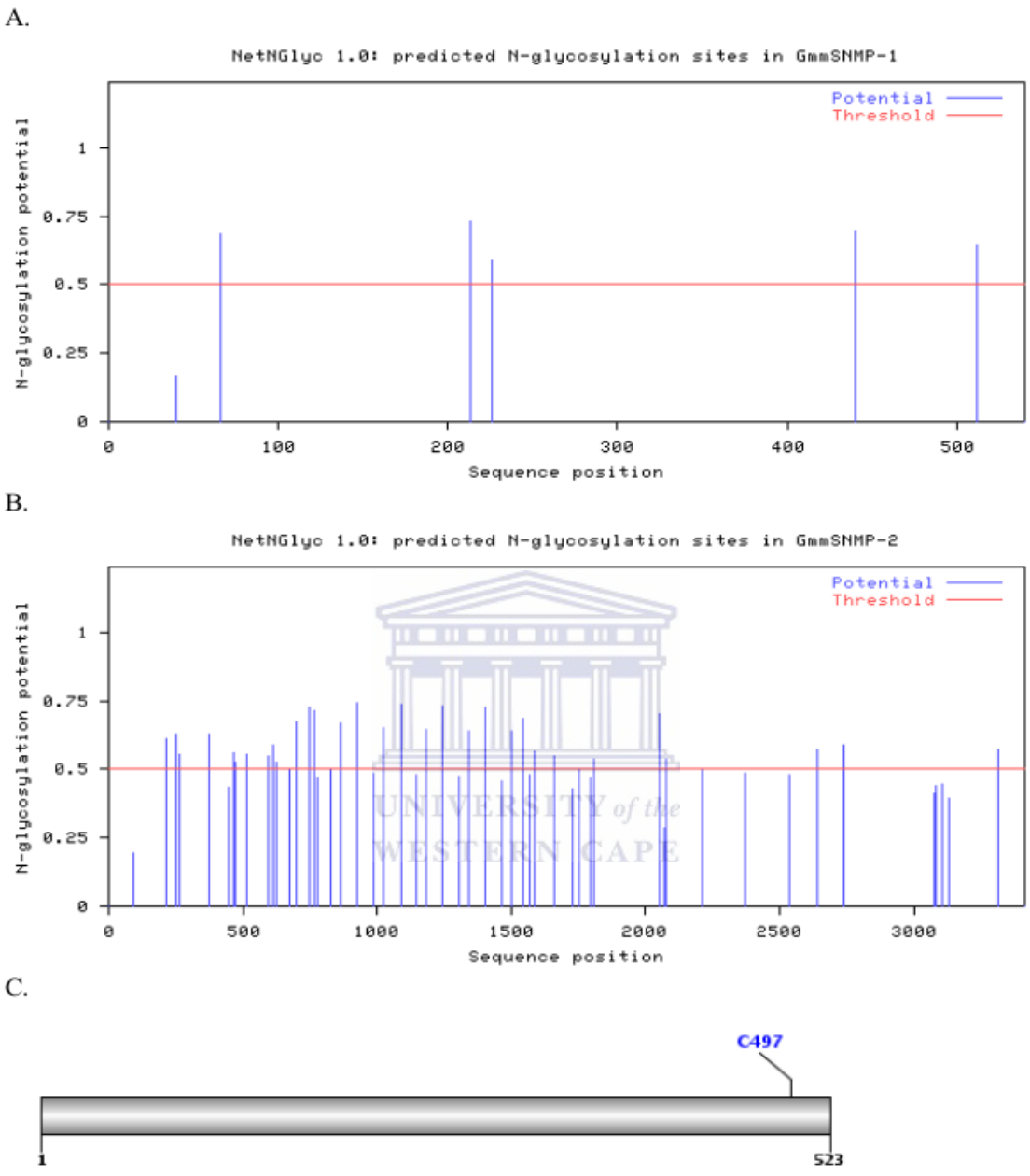


Figure 3.20 Predicted N-glycosylation and palmitoylation sites of SNMP1 and SNMP2 in *G. m. morsitans*.
A. – Asparagine N-glycosylation sites in SNMP1; B. – Asparagine N-glycosylation sites in SNMP2; C. – palmitoylation site in SNMP2. N-glycosylation sites were predicted via NetNGlyc 1.0 sever (www.cbs.dtu.dk/services/NetGlyc/), while palmitoylation sites were predicted using CSS-Palm 4.0 tool (http://csspalm.biocukoo.org).

3.4.4 Expression levels of OBPs, CSPs and CD36-like genes in *G. m. morsitans*

Glossina m. morsitans OBP21 (28%), OBP1 (19%), OBP2B (15%), OBP22 (14%), and OBP8A had higher expression levels out of a total of n= 865,561 RPKM supporting expression of all OBP genes. The remaining *G. m. morsitans* OBP genes had lower expression levels (Figure 3.21).

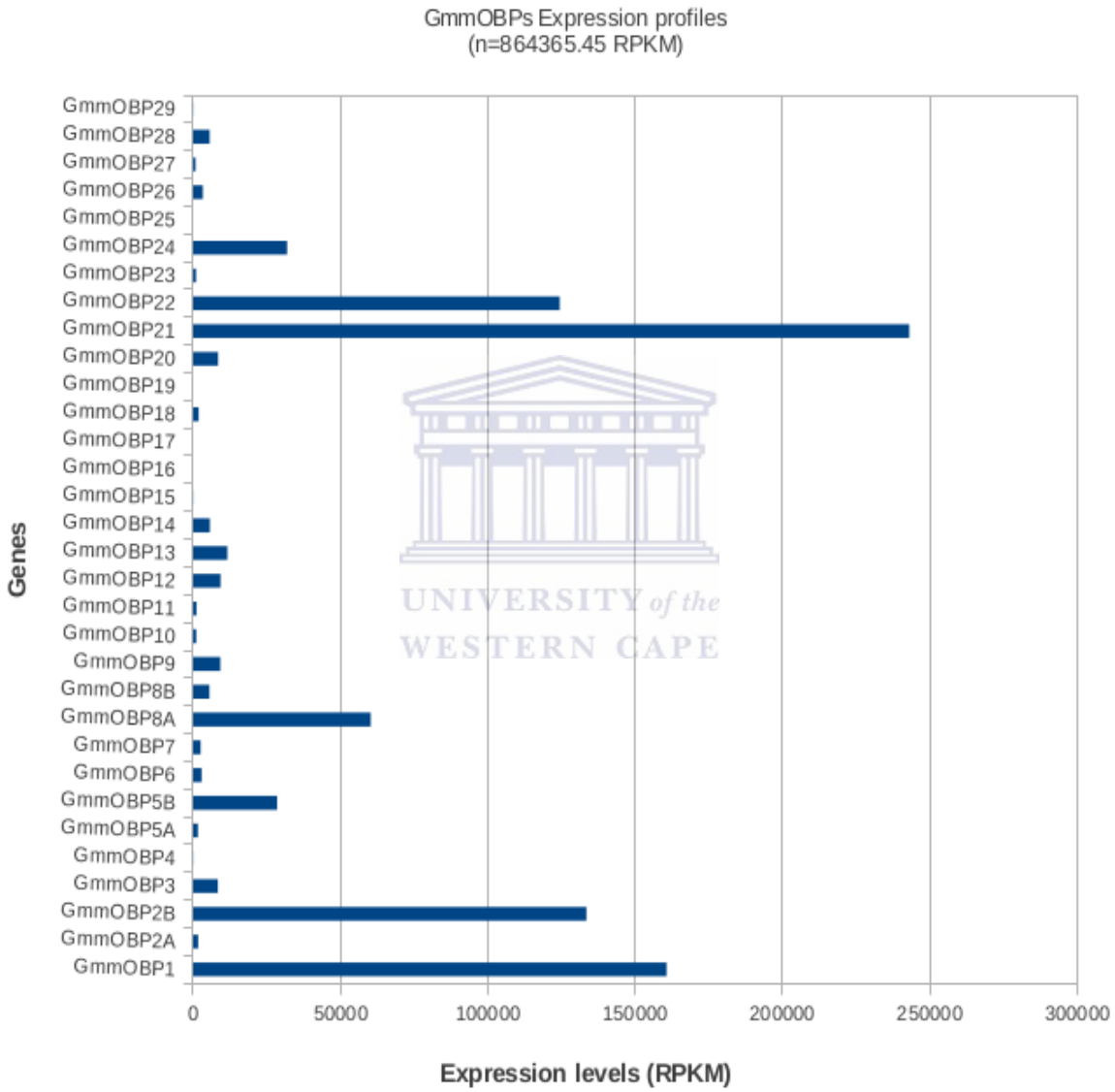


Figure 3.21 Expression levels of OBP genes in *G. m. morsitans*
Total amount of supporting RNA-seq data is given by 'n'. OBP21, OBP1, OBP2B, OBP22 and OBP8A are highly expressed genes. The expression levels were generated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

Of the five *G. m. morsitans* CSP genes, CSP3 transcripts representated 53% and CSP1 43% of the total 1,360,124 RPKM reads supporting expression of CSPs in the genome. This probably suggests that they might play key role in either binding of hydrophobic odors or in other critical physiologies. The others, CSP2, CSP4 and CSP5 had lower expression levels. This contrasts earlier reports for CSP2 expression (Liu *et al.*, 2012) and presence of functional ontologies that point to its crucial role in antennal chemosensory specific binding activities (Figure 3.22).

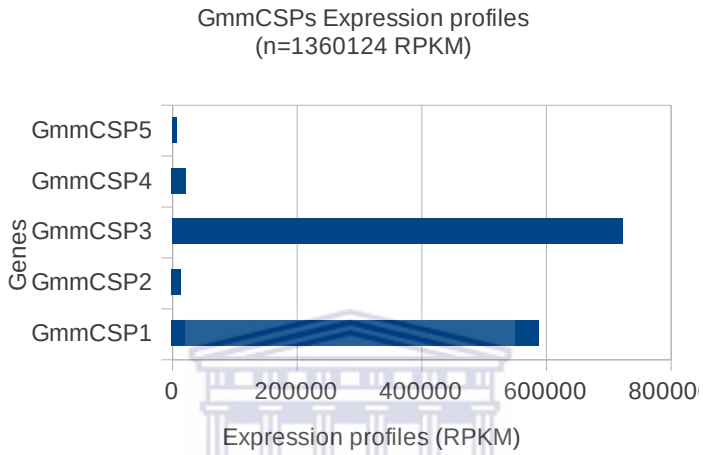


Figure 3.22 Expression levels of CSP genes in *G. m. morsitans*
Total amount of supporting RNA-seq data is given by 'n'. CSP3 and CSP1are highly expressed genes. The expression levels were estimated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

All the *G. m. morsitans* genes coding for CD36-like genes analyzed were represented by over 600,000 RPKM RNA-seq expression data (Figure 3.23). The Croquermort-like *G. m. morsitans* homologs (*CD36-14, CD36-8, and CD36-3*) had the most amounts of expression data. The *G. m. morsitans SNMP* genes had low expression level, representing just over 1% (5,238 out of 611,505 RPKM) of all *G. m. morsitans* CD36-like reads. Further, the *G. m. morsitans* SNMP2 gene was highly expressed at 80% of 5,238 RPKM, suggestive of its differential cellular location/function relative to SNMP1 (Liu *et al.*, 2014).

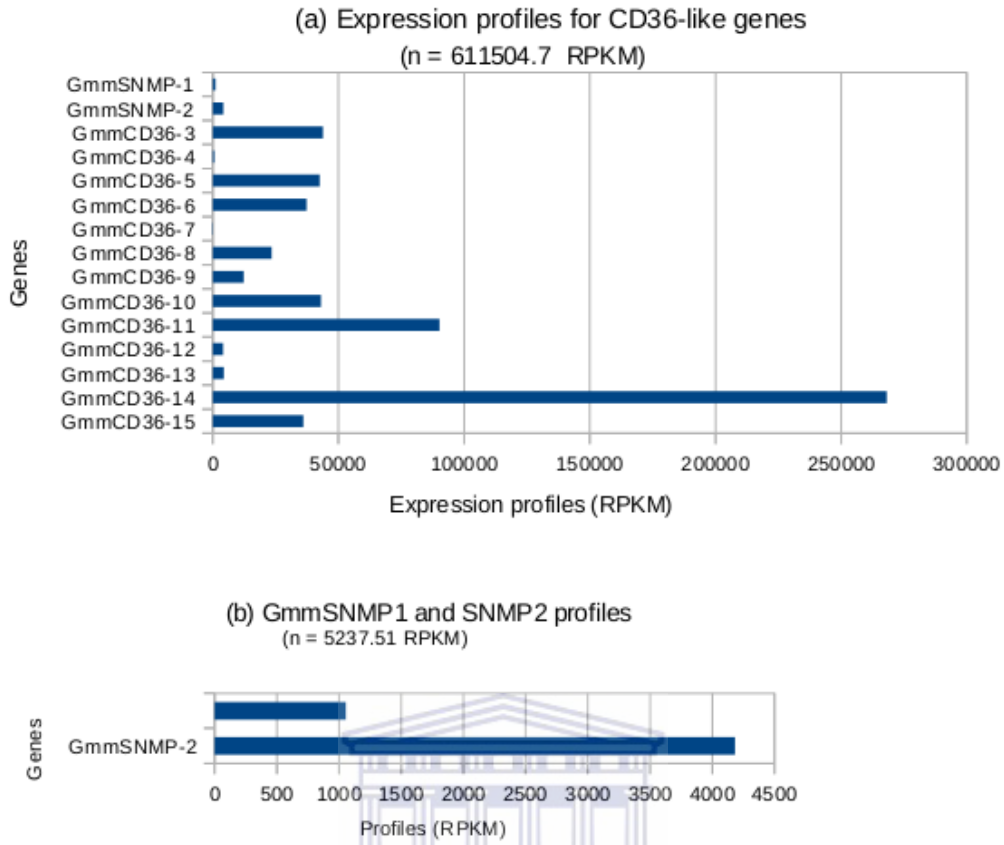


Figure 3.23 Expression levels of CD36-like genes in *G. m. morsitans*.
Total RNA-seq data is given by 'n'. (a). Expression levels of all the 15 CD36-like genes; (b). Expression levels of SNMPs. The expression levels were estimated using RNA-seq analysis pipeline in CLCGenomics Workbench (CLC bio, 2012).

3.4.5 Phylogenetic relationships of chemosensory responsive OBPs, CSPs, and CD36-like proteins of *G. m. morsitans* with those from *D. melanogaster* and *An. gambiae*

Maximum likelihood phylogenetic tree of the *G. m. morsitans* chemosensory-related proteins including OBPs, CSPs and CD36 with those in *D. melanogaster* and *An. gambiae* are summarized in Figures 3.24, 3.25 and 3.26 respectively. The tree containing *G. m. morsitans* OBPs (Figure 3.24) had three clusters, labeled 1, 2 and 3; some individual proteins had low terminal bootstrap support. Cluster 1 (green arcs) consisted of two sub-clusters, 1A and 1B. Sub-cluster 1A contained only *An. gambiae* OBPs – the atypical cluster of OBPs exclusively found only in this mosquito species. Sub-cluster 1B contained known homologs of Minus C (the fruit fly DmelOBP8a, DmelOBP44a, DmelOBP83g, DmelOBP99a-d, and anopheles member AgamOBP9); see section 1.2.5.4 for review of OBPs classifications. The *G. m. morsitans* OBPs that can be putative Minus C due to absence of C2 and/or C5 were GmmOBP1, OBP2A, OBP3, OBP21 and OBP22, and also

distantly related to GmmOBP17, OBP19, OBP20 and OBP28 which have C2 and C5 sites. However, GmmOBP20 lacked C4 and C6 while GmmOBP17 lacked C4, but had two additional cysteine sites that may functionally compensate in making a needed disulfide bridge. Cluster 2 (blue arcs) comprise putative Plus C homologs dominated by members from the fruit fly and mosquito homologs. The homologs from *G. m. morsitans* include GmmOBP5A and GmmOBP23, distantly rooted by GmmOBP7, DmelOBP5b and DmelOBP83c. Cluster 3 (red arcs) has sub-clusters 3A, 3B and 3C, generally corresponds to the GOBP/PBP class (also referred to as ABPX or OS-E/OS-F OBPs, and are rooted by GmmOBP29, DmelOBP73a, OBP57d and OBP57e. In sub-cluster 3A, the *G. m. morsitans* OBPs are GmmOBP8A, OBP8B, OBP9, OBP10, OBP15, OBP16, OBP18, OBPOBP25, and OBP27. The sequence alignment similarities of GmmOBP8A, OBP8B, OBP9, OBP10, and OBP16 indicate they have at least two proline sites after C6. Sub-cluster 3B contains GmmOBP5B, OBP6, OBP14, OBP24 and OBP26; and sub-cluster 3C has two tsetse fly members, GmmOBP4 and OBP13. These could be *G. m. morsitans* putative general odorant and/or pheromone binding proteins. However, the proof of functionality of all the *G. m. morsitans* OBPs require in depth experimental validation beyond the present data. In general, the *G. m. morsitans* OBPs have diverse relationship with their counter-parts in the *D. melanogaster* and *An. gambiae.*

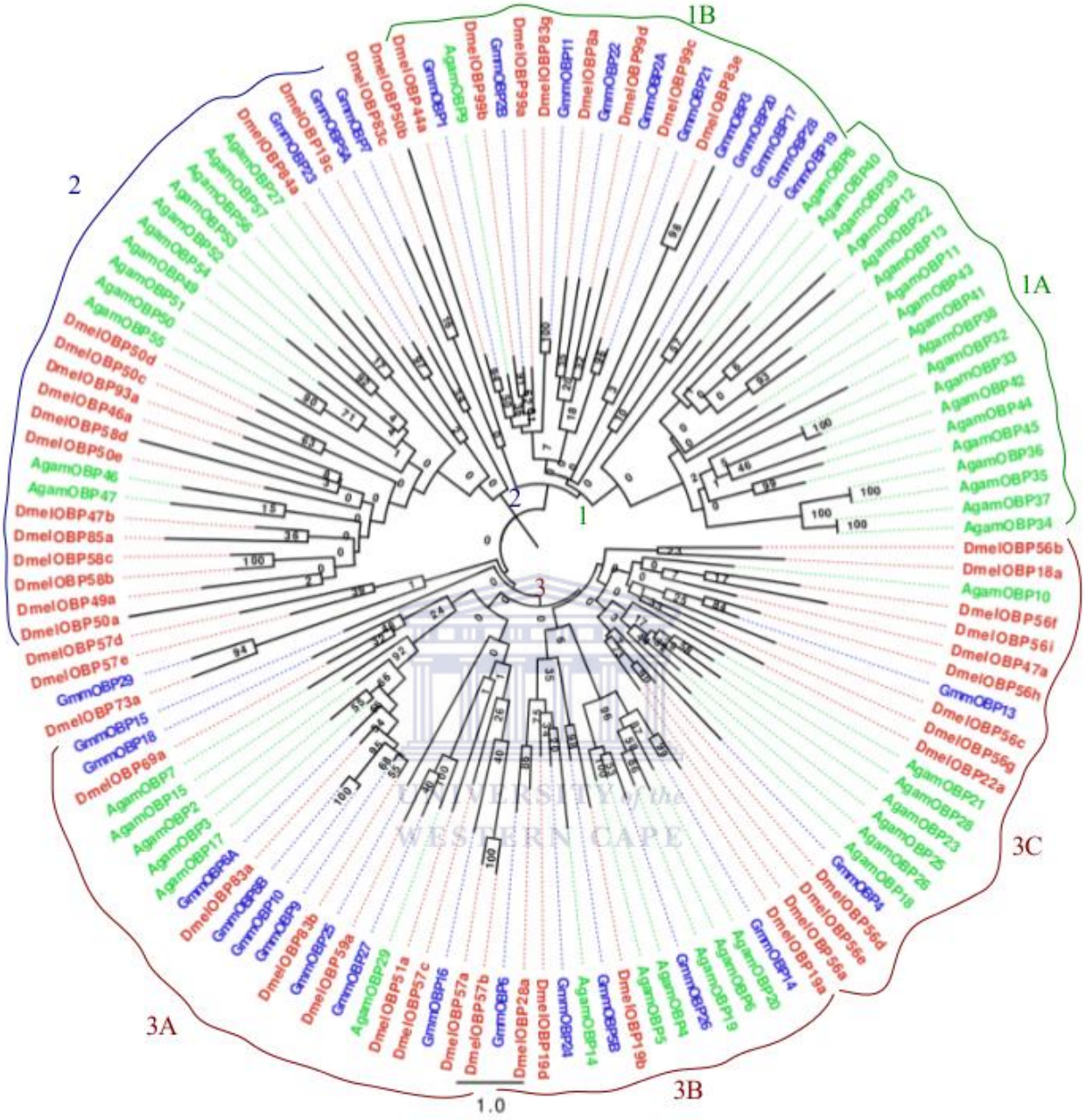


Figure 3.24 Phylogenetic tree of *G. m. morsitans* OBP proteins with orthologs from *D. melaogaster* and *An. gambiae* .
The blue leaves are tsetse fly *G. m. morsitans* OBPs; red leaves are fruit fly *D. melanogaster* OBPs; while green leaf nodes are mosquito *An. gambiae* OBPs. The bootstrap supports are shown on the internal branch nodes. The annotations 1A and 1B (Minus C), 2 (Plus C), and 3A, 3B and 3C (GOBP/PBP) indicate cluster names. Multiple sequence alignment was done using MUSCLE tool (Edgar, 2004) and alignment edited via Jalview tool (Waterhouse *et al*., 2009). The tree was constructed using PhyML program (Quindon *et al*., 2009) and edited using Figtree tool (http://tree.bio.ed.ac.uk/software/figtree/).

The *G. m. morsitans* CSPs maximum likelihood tree revealed GmmCSP4 more divergent from the other members whereas GmmCSP1 and CSP3 are closely related to DmelPebIII, GmmCSP2 to DmelA10 and AgamCSP3, and GmmCSP5 clusters with DmelCSP1 and AgamCSP4. DmelPhk3 and DmelCSP2 and AgamPhBb, AgamCSP1 and AgamCSP2 did not cluster with any *G. m. morsitans* CSPs (Figure 3.25).

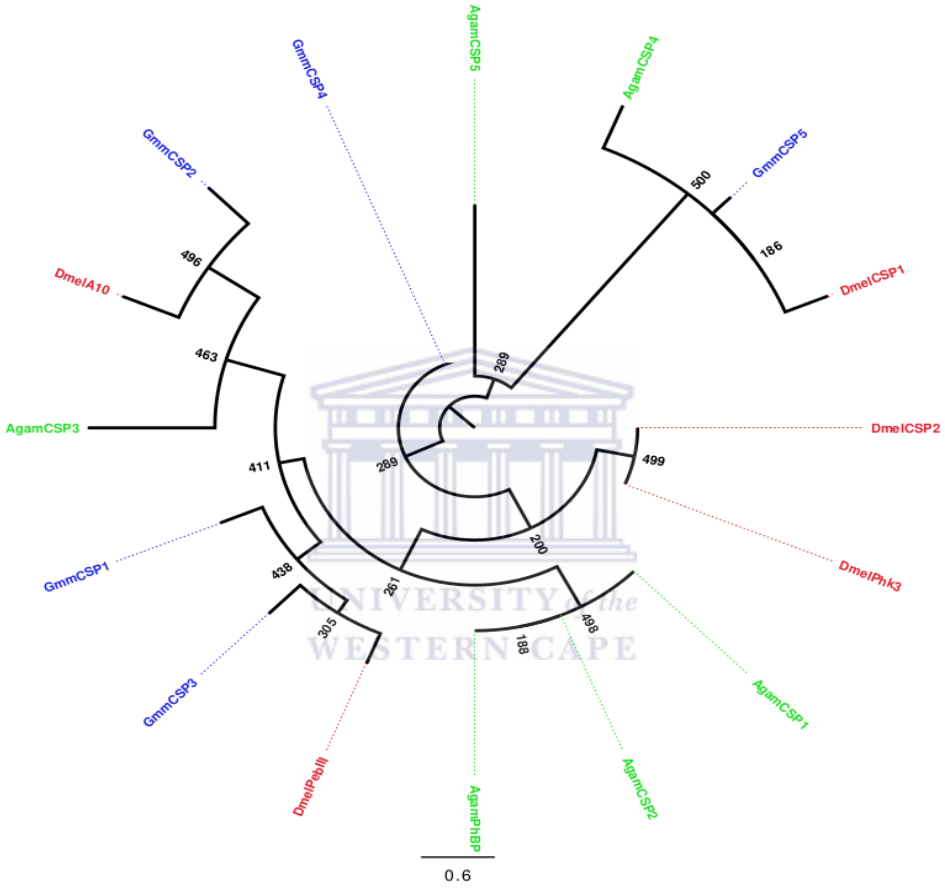


Figure 3.25 Phylogenetic tree of *G. m. morsitans* CSPs proteins with orthologs from *An. gambiae* and *D. melanogaster*.
The blue leaves are tsetse fly *G. m. morsitans* CSPs; red leaves are fruit fly *D. melanogaster* CSPs; while green leaf nodes are mosquito *An. gambiae* CSPs. The bootstrap supports are shown on the internal branch nodes. Multiple sequence alignment was done using MUSCLE tool (Edgar, 2004) and alignment edited via Jalview tool (Waterhouse *et al.*, 2009). The tree was constructed using PhyML program (Quindon *et al.*, 2009) and edited using Figtree tool (http://tree.bio.ed.ac.uk/software/figtree/).

Apis mellifera CD36 ortholog outgroups indeed formed the base of all the sub-trees (Figure 3.26). The proteins from the four species (*G. m. morsitans*, *D. melanogaster*, *An. gambiae* and *A. mellifera*) clustered into three main groups, each having two sub-clusters putting homologs together, with at least 70% bootstrap values. All the *G. m. morsitans* CD36-

like appeared singly in their respective clusters except for GmmCD36-3/4 and GmmCD36-6/7 that were in one sub-cluster. Cluster 1 comprised the SNMPs homologs, indicating each species has one member protein each except *An. gambiae* that have two representatives, AgSNMP1 and AgSCRB1, and AgSNMP2 and AgSCRB16 respectively. *Apis mellifera* outgroup AmSRBVII and AmSRBIV seemed homologous to the SNMP1 and distant from SNMP2 members of other species. Cluster 2 has two sub-clusters, with GmmCD36-10, CD36-11 and CD36-13 in one, and GmmCD36-9, CD36-12 and CD36-15 in the other cluster. Each of these *G. m. morsitans* pair up with a member from each of the other diptera, rooted by the *A. mellifera* except GmmCD36-9 that lacks a honey bee outgroup. Cluster 3 can also be sub-divided into two sub-clusters with seven *G. m. morsitans* CD36-like proteins each rooted by an outgroup member. In one sub-cluster, GmmCD36-3 and CD36-4 are homologous to the *D. melanogaster* pestle protein; CD36-6 and CD36-7 are homologous to *D. melanogaster* Santa Maria protein, and GmmCD36-5 homologous to *D. melanogaster* CG7227. The *G. m. morsitans* CD36-6 and CD36-14 are homologous to *D. melanogaster* NinaD and Croquermort proteins respectively.

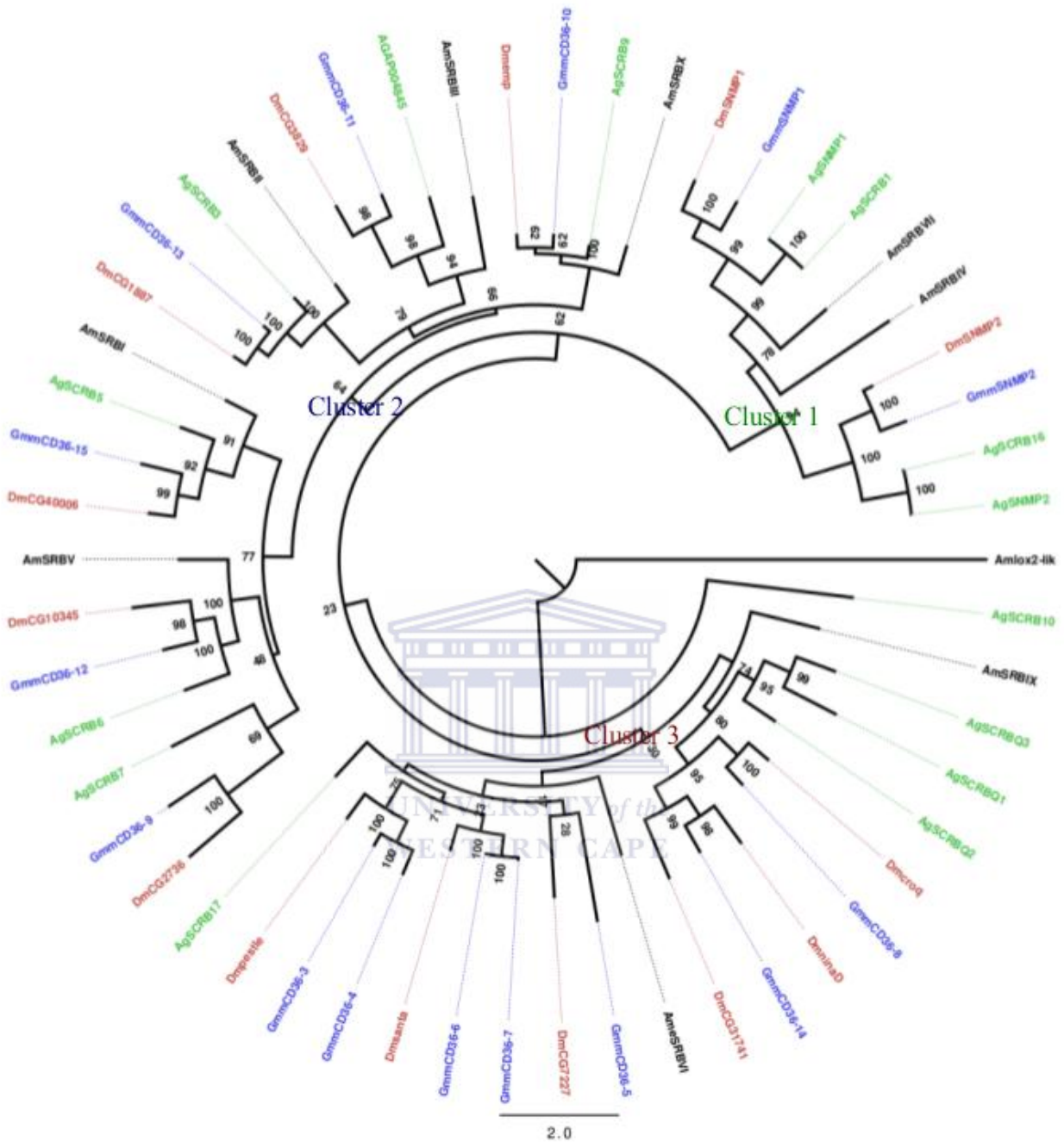


Figure 3.26 Phylogenetic tree of *G. m. morsitans* CD36-like proteins with orthologs from *An. gambiae* and *D. melanogaster*
The blue terminal brances are tsetse fly *G. m. morsitans* CD36-like; red branches are fruit fly *D. melanogaster* CD36; black branches are for *A. mellifera* CD36; while green branches are mosquito *An. gambiae* CD36. The tree clusters are labeled at the internal nodes. The values for 100 bootstraps are shown on the internal branch nodes. The tree was constructed using PhyML program (Quindon *et al.*, 2009) and edited using Figtree tool (http://tree.bio.ed.ac.uk/software/figtree/).

CHAPTER FOUR

DISCUSSION

Insect ecology contains heterogeneous mix of chemicals. The chemicals include volatiles (smell molecules) detected by different types of antennal hair cells (sensilla), and soluble or insoluble chemicals (taste molecules) detectable by contact using such organs as legs, proboscis, wings and ovipositor. The power or strength of any insect survival lies in its ability to detect, discriminate and respond to these multiple chemical cues found in their immediate environments. This process is referred to as chemoreception, and includes gustation for taste detection, and olfaction for smell detection. It is now established that many protein families are involved in insect chemoreception and signaling. These include multi-transmembrane smell receptors such as odorant (ORs), gustatory (GRs), and ionotropic (IRs); and chemosensory related proteins such as odorant binding proteins (OBPs), chemosensory-specific proteins (CSPs) and sensory neuron membrane proteins (SNMPs) which prime the functioning of the periceptors. Downstream in the brain, the signals from chemical sensory organs may be mediated by synaptic glutamate neurotransmitter that binds and activates ligand-gated ion channels, ionotropic glutamate-gated receptors (iGluRs), and second-messenger dependent metabotropic glutamate-gated receptors (mGluRs). These proteins help insects to quickly perceive and respond to the chemical stimuli, helping them to select important resources in their habitats including mates, resting sites, food source and sites to lay eggs. Many previous studies have reported different numbers of these protein families in different insect genomes (Robertson *et al.*, 2003; 2009; Robertson & Wanner, 2006; Gardiner *et al.*, 2008; Croset *et al.*, 2010; Hansson & Stensmyr, 2011; Zhou *et al.*, 2012; Missbach *et al.*, 2014). However, none of these earlier works reported on any tsetse fly. This chapter discusses the results of repertoires of chemosensory and related gene families identified in the tsetse fly *G. m. morsitans* genome, under three sub-headings: the seven transmembrane non-glutamate chemoreceptors (ORs and GRs), the glutamate-gated receptors (IRs, iGluRs, mGluRs), and chemosensory-related genes (OBPs, CSPs, SNMPs). An overall comparison of the *G. m. morsitans* annotated gene repertoires with those published in other selected insect species (see Table 1).

4.1 The *Glossina m. morsitans* chemoreceptors, ORs and GRs

Odorant receptors

Out of 46 recoverable OR genes in the *G. m. morsitans* genome reported here, 25 ORs were confirmed to correspond to the ones computationally predicted by Vectorbase ((Lawson *et al., 2009*), and have since been integrated into the VectorBase

database and assigned unigue identifiers; the remaining 21 novel ORs were manually identified and assigned temporary identities in this study. All the genes were rigorously edited and annotated manually to yield the gold standard repertoire of complete sequences presented here. All the *G. m. morsitans* OR genes were encoded with multiple exons and were scattered across different scaffolds, with no gene clustering in the genome as found in the *D. melanogaster* (Robertson *et al.*, 2003). The *G. m. morsitans* OR genes do not have alternative splice variants, except for only one gene, *GmmOR5;* other insects use alternative splicing as a mechanism to generate many receptors that respond to closely related odors in a species-specific manner – as found in fruit fly *D. melanogaster* (Robertson *et al.*, 2003), honey bee *A. mellifera* (Robertson & Wagner, 2006) and *Cx. quinquefasciatus* (Arensburger *et al.*, 2010). Overall, these *G. m. morsitans* chemoreceptor OR genes are fewer compared to those in other insects (Sánchez-Gracia *et al*., 2009; Benton, 2006). For instance, there were 16 less *GmmORs* than those in *D. melanogaster* and 33 less in *An. gambiae* (see Table 1). However, more genes could be recovered with improved genome assembly.

Despite having lower repertoire of OR genes *G. m. morsitans* encode specific OR genes in multiple copies; and also express some specific genes at higher levels. Such genes represent more than half of all the OR chemoreceptors in the genome. For instance, *G. m. morsitans OR6-OR9* revealed an expansion of related genes, and were all homologous to *D. melanogaster* Or42b. In drosophila larvae, Or42b have been implicated in mediating refractory responses against low concentrations of ethyl acetate (Kreher *et al*., 2005; 2008; Mathew *et al*., 2013). Ethyl acetate is a commonly used compound by entomologists as an asphyxiant to slowly kill the insects with little damage and also keep the tissues soft for microscopy; this means it is toxic to insects. It is also an abundant component of ripening fruits, and for fruit juice sucking insects it could signal source of unpalatable food. In overall, most insects would detect it and respond by trying to avoid it. It is yet unestablished what role these Or42b homologs could be playing in the tsetse fly chemo-biology. A cluster that had highest expression level was *OR14-OR16, which* seemed paralogous in the genome and were all homologous to DmelOr45a. The *GmmOR15* alone constituted over 90% of transcripts that supported expression of ORs. In *D. melanogaster* larvae, the products of DmelOr45a together with DmelOr33b are expressed on the same neuron that couple light and odor perception to induce an escape response (Kreher *et al.*, 2005; Bellmann *et al.*, 2010). The function of this GmmOR15 in any tsetse fly is yet to be determined. However, given that the RNA sequence data used to quantify the expression levels in this study were generated from reproductively active adult female flies (Benoit *et al.*, 2014), it is possible that *G. m. morsitans* OR15 products could in some way be associated with immature larval activity or the pregnant

female fly behaviors. Probably, this could be useful in coupling response to light and larviposition odors when searching for suitable places to deposit mature larvae. Another cluster in the phylogenetic tree is *GmmOR23-30*, and it suggests that they may be evolutionary related and also function in detecting and responding to similar stimulants. The closest homologs of these genes in the phylogenetic tree are DmelOr49a and DmelOr84f, which are co-expressed on same neurons (Couto *et al.*, 2005). Another cluster, *G. m.* morsitans *OR41-46*, seemed multiple copies of same gene, were all homologous to the only known volatile insect pheromone, (Z)-*11-cis*-vaccenyl acetate (cVA) receptor, DmelOr67d (Kurtovic *et al.*, 2007; Laughlin *et al.*, 2008; Dahanukar & Ray, 2011; Farine *et al.*, 2012). Probably this expansion indicates emphasis in regulating social and sexual mating behaviors in the *G. m. morsitans*, specifically deterrence of males from mated females, and making mated females to show refractoriness to further mating (Attardo *et al.*, 2006; Dahanukar & Ray, 2011). However, no volatilous pheromone has been identified in any tsetse species, thus making this finding important for future search for tsetse odorant pheromones. Nonetheless, the products of these sets of *G. m. morsitans* OR genes must be playing crucial roles, and may give clues to preferential behavior of the *G. m. morsitans* in their ecology in response to a mix of odorous plumes. For the other genes that had lower transcription levels, it can be hypothesized that they may be important for decoding specific but uncommon single file odors; alternatively they may be expressed at other physiological states and developmental stages other than the gestating female stage. The present data are insufficient to prove these thoughts.

Gustatory receptors

Similarly, the *G. m. morsitans* GRs repertoires reported here provide a complete set of recoverable gustatory genes in the current genome assembly. They also improve the VectorBase predictions by manually editing the eight computationally identified genes and recovering six extra genes by manual curation. Similar to the ORs, the *G. m. morsitans* genome encodes fewer chemoreceptor genes for GRs compared to those in other insects (Sánchez-Gracia *et al.*, 2009; Benton, 2006, see also Table 1). Some *GmmGRs* appeared in multiple copies and others had higher expression levels in *G. m. morsitans*. For instance, *G. m. morsitans* GR1-4 genes , whose proteins may function in binding similar stimulants; in this case CO_2 emanating from vertebrate hosts (Lehane, 2005) and resting sites plant hosts (Bouyer *et al.*, 2007). The four, GmmGR1-4, are homologous to CO_2 receptors (DmelOr21a and Or63a in *D. melanogaster*, and AgOr22, Or23 and Or24 in *An. gambiae*), and may be associated with host seeking behavior, to trigger attractive responses as elicited by the savanna tsetse fly species (Bogner, 1992). Evolutionarily, the *GmmGR1-4* have high sequence similarity to house flies (*M. domestica)* and Mediterranean fruit flies (*C. capitata)* than to either the fruit flies or to the facultative blood-

feeding mosquitoes, and these are flies that may be using CO2 as an attractive stimuli (see Figure 3.4 and Figure S1).

In stark contrast to other insects including sap sucking and blood sacking diptera, the *G. m. morsitans* genome had no recoverable genes for sweet taste receptors. In fact, in other insects the sweet/sugar receptors are encoded in multiple copies and are expanded in a species-specific manner (Robertson *et al*., 2003; Robertson & Wanner, 2006; Jiao *et al*., 2007; McBride *et al*., 2007; Kent *et al*., 2008). This suggests that the tsetse flies do not detect sweet taste molecules in their ecologies; consistent with the fact that they generate their flight energy from the alanine-proline metabolic pathway (Hargrove, 1976; Beenakers *et al*., 1984), as opposed to other insects that feed on sugary saps and nectaries (Robertson & Wanner, 2006; Gardiner *et al*., 2008).

Implications in biology of tsetse

Behaviorally, the factors underlying the apparent expansion and reductions of specific chemoreceptor genes or absence of other genes in the *G. m. morsitans* are yet unknown. Nevertheless, the lower chemoreceptor (ORs and GRs) repertoires suggest a smaller smell and tastant spectrum relevant to the survival of the flies in their ecology. Perhaps, the lower repertoires support the report that the tsetse flies compliment their chemical communication with other senses like vision better than other insects (Lehane, 2005). The tsetse flies are also less mobile, only covering a limited distance per year, thus limiting the possibility of encountering a diverse range of host odors and tastes. It can also be that the obligate blood-feeding of the tsetse fly (restricted to vertebrate hosts), might have necessitated evolutionary selection of specific chemoreceptor genes relevant to discriminate amongst specific vertebrate hosts. These are in contrast to the *D. melanogaster* that have expansive fruit species hosts (Gardiner *et al.*, 2008), the mosquitoes that feed on diverse plant saps and a range of mammalian blood to mature their female eggs (Fox *et al.*, 2001; Bohbot *et al*., 2007; Arensburger *et al.*, 2010), and to the honey bees that explore a whole range of floral nectaries (Robertson & Wanner, 2006). Environmental factors are also known to influence tsetse host choice, as they have been shown to have an acquired preference to specific hosts encountered early in life (Bouyer *et al.*, 2007), suggesting that their chemoreceptor system trains and retains the odor memory learnt early in life.

The *G. m. morsitans* lack the broadly expanded chemoreceptor lineages seen in other insects. For instance, the honey exhibit arrays of 60 genes of advent expansion defining bee-specific odorant receptors (Robertson & Wanner, 2006). In addition, while the mosquitoes, Anopheles, Aedes and Culex, share specific gene expansions, each have their individual gene lineage expansions, probably befitting their unique broad dietary ecologies (Bohbot *et al.*, 2007). This can be linked to

broad diversity of food sources ranging from different floral nectaries, plant saps and fruit juices that each of these insect species feed on. In contrast, the tsetse flies obligately feed on vertebrate blood, and their hosts are relatively abundant and stable within their habitats - thus eliminating the need for a broad chemosensory repertoire. Further, other blood-feeders such as mosquitoes possess expanded species-specific divergent taste receptors for seeking a variety of plant nectaries for sugar as energy source. The *G. m. morsitans* lack all known sugar receptor genes, which are highly conserved in other insects (Robertson & Wanner, 2006; McBride *et al.*, 2007). This may be related to the observation that tsetse flies derive their energy from the amino acids proline and alanine metabolic pathways (Hargrove, 1976), and may have no use for such expanded taste chemoreceptome.

Structurally, few specific groups of the *G. m. morsitans* OR and GR genes were clustered within their scaffolds, probably suggesting recent duplication events and may converge into performing related similar functions (like sensing a class of related semiochemicals). Similar clusters of genes performing common and related functions have been observed among chemoreceptor genes in *D. melanogaster* (Robertson & Wanner, 2006; McBride *et al.*, 2007; Sanchez-Gracia *et al.*, 2009), and more recently amongst twelve *G. m. morsitans* major milk proteins associated with lactation (Benoit *et al.*, 2014). It is common to find such clustered genes with more or less similar expression profiles. Such genes generally have common regulatory elements that drive their joint expression (Robertson *et al.*, 2003; Guo & Kim, 2007; Nozawa & Nei, 2007). Therefore, the few clusters of *G. m. morsitans* OR and GR genes in the genome might be under common regulatory mechanisms and in response to common or related stimuli.

In terms of sequence conservation, a part from the genes named above, the *G. m. morsitans* OR1 (homologous to Orco) was the most conserved amongst the *G. m. morsitans* ORs, retaining a single copy as is in other insects examined, sharing over 70% sequence similarity. This was not surprising since such conservation has been observed in other insects (Dahanukar *et al.*, 2005; Bohbot *et al.*, 2007), probably due to its critical role in modulating responses of the other receptors. The remaining *G. m. morsitans* ORs and GRs were less conserved, meaning they had higher sequence diversity, suggesting that the *G. m. morsitans* may detect slightly different sub-sets and mixtures of odor plumes from those detected by other insects. However, these diverse ORs and GRs had clear orthology with other related species. This was previously revealed amongst the fruit fly species in which the gene orthology and their neuronal response dynamics of chemosensory receptors were conserved, with a surprisingly slower evolution rate compared to their vertebrate counterparts (Guo & Kim 2007, Nozawa & Nei, 2007; Hansson & Stensmyr, 2011).

Conclusively, the *G. m. morsitans* reveals a contracted chemoreceptome of ORs and GRs, probably retaining only crucial orthologous genes in the genome. The fly seems to prioritize and invest in encoding a select few odorant and gustatory chemoreceptor genes necessary for feeding and reproduction behaviors. The few selected gene are then prioritized by being encoded in multiple copies or by increasing their transcription levels. Indeed, selectively prioritizing and investing in specific genes is not unusual phenomena in insects (McGraw *et al.*, 2004; Bionaz & Loor, 2011).

4.2 The *Glossina m. morsitans* glutamate-gated ion receptors - IRs, iGluRs and mGluRs

Ionotropic receptors, IRs

Recoverable numbers of glutamate-gated receptor IR genes in the *G. m. morsitans* genome indicate a drastic reduction in the number of IRs, 19 versus 62 in *D. melanogaster* (Benton *et al*., 2009; Croset *et al*., 2010; Zhou *et al*., 2010; 2012), see also Table 1. However, important findings are that: (i) the *Glossina* encodes one specific gene lineage in multiple copies - three copies of IR84a (*IR84a-A, IR84a-B, IR84a-C)*, in contrast to single copy in fruit fly; reportedly, it is also only found in drosophilids – where its products are implicated social and sexual mating responses (Grosjean *et al.*, 2011; Rytz *et al.*, 2013); (ii) key IR gene lineages not recoverable in *G. m. morsitans* may had been lost (or may had never evolved at all) e.g. the homologs of *D. melanogaster* IR93a and IR41a. In fact, it is surprising that the *Glossina* lacks IR93a that is an ancestrally conserved gene in all protostome lineages (Croset *et al*., 2010); (iii) the most remarkable reduction was amongst the divergent species-specific sub-type of IRs, in which the *Glossina* contain only three likely members – *GmmIR10a, IR56b,* and *IR56d*. This contrasts the scenario in other insects like fruit fly, mosquitoes, and honey bee that have wide expansions of such genes (Rytz *et al.*, 2013); (iv) the *Glossina* IRs are all encoded with multiple exons, in contrast with fruit flies where non-olfactory species specific IRs are encoded as single exons (Ai *et al.*, 2010). This latter point would agree with the observation that the *Glossina* genes not only lack alternative splice variants, but also have no signs of recent duplication events. This means the genome experiences no pressure to expand and diversify its chemosensory repertoire.

The *G. m. morsitans* encodes a highly conserved class of IRs, the antennal or olfactory IRs (Benton *et al.*, 2009; Ai *et al.*, 2010; Croset *et al.*, 2010). Amongst these are known co-receptor antennal IR genes (*GmmIR8a, IR25a,* and *IR64a*) that shared high homology above 70% sequence similarity, yet others have conserved multiple copies. (e.g. three copies of GmmIR84a – *IR84a-A, IR84a-B, IR84a-C* and four copies of IR75 family – *IR75a, IR75b, IR75c, IR75d*) compared to their

orthologs in *D. melanogaster* (Ai *et al.*, 2010) and *An. gambiae* (Croset *et al.*, 2010). Coincidentally, these conserved IR genes also had higher expression levels (*GmmIR8a, IR76b, IR64a, IR84a-B, and IR84a-C*), suggesting that *G. m. morsitans* heavily invests in them. Indeed, the homologs of *IR8a, IR25a, IR76b,* and *IR64a* are amongst the most conserved and expressed even in the ancestrally basal insects (Croset *et al.*, 2010; Ai *et al.*, 2013; Missbach *et al.*, 2014). In *G. m. morsitans,* the products of these genes may function in a combinatorial manner and to be crucial, not only for sensitivity to specific general odors, but also in chaperoning localization of other receptors to neuronal dendrites, as has been reported in *D. melanogaster* (Abuin *et al.*, 2011). For instance, the *G. m. morsitans* IR76b with six copies of ubiquitin ligase related neuralized domains (Pavlopoulos *et al*., 2011), could potentially be a multi-functional chemoreceptor participating both in developmental notch pathway-mediated neuronal cell-fate determinations, as well as chaperoning other ionotropic receptors to their correct localization and odor detection on the OSN dendrites. These particular sets of *G. m. morsitans* IR genes, may also be expressed and used ubiquitously as general receptors than others, a phenomenon established in other insect genomes (Bionaz & Loor, 2011). The *G. m. morsitans* may use this mechanism for exploiting specific odor resources in their ecology, in this case isolation of general odors, thus supporting earlier findings that the fly expresses more generalist sensilla than specialist sensilla (den Otter, 1991; den Otter & Natters, 1992; 1993).

The apparent reduction of species-specific 'divergent' IR subfamily in *G. m. morsitans* genome is interesting and not conclusive. It may have resulted from the obligate blood-feeding style of the *G. m. morsitans* in an ecology that is relatively stable with abundant vertebrate hosts, yielding a narrow range of odor plume. Therefore, this will save the tsetse flies the need to sample a wide range of odor plumes, compared to other insects like mosquitoes, honey bees, ants and fruit flies (Robertson & Wanner, 2006; Benton *et al*., 2009; Croset *et al.*, 2010; Zhou *et al*., 2012). Importantly, this may also mean that the *G. m. morsitans* IRs are dedicated to detecting important general vertebrate host volatiles. It is also that the ability of tsetse flies to search for hosts, successfully land, probe and feed to engorgement, may involve use of other senses such as tactile, thermal and contact stimuli (Gikonyo *et al.*, 2002; 2003). Therefore, this would enable the *G. m. morsitans* IRs to trade-off specific odor response for general odor sensitivity (den Otter & Natters, 1993).

Ionotropic glutamate-gated receptors and metabotropic glutamate-gated receptors

Insect glutamate-gated receptors are expressed on both PNS and CNS neurons, mediating fast impulse transmissions after binding glutamate neurotransmitter (Ramaekers *et al*., 2001; Bogdanik *et al*., 2004; Gladding *et al*., 2009). Though the molecular functions of these genes and their proteins have not been studied in tsetse, they may,

nonetheless, provide useful information on the physiology and behavior dynamics of how the tsetse fly navigates the odor space to select their preferred host. The *G. m. morsitans* iGluRs and mGluRs have high sequence conservation and also have high expression profiles than IRs. These may relate to the chemosensory role of IRs that requires a diversity to match the different ecological odors, while the iGluRs and mGluRs participate in the CNS neural transmission functions that are themselves highly conserved (Bogdanik *et al*., 2004; Braga *et al*., 2004; Gladding *et al*., 2009). The presence of these receptors may be useful in the tsetse fly, whereby the iGluRs may mediate quick neuronal responses, while the mGluRs modulate behavioral dynamics as the fly navigates its ecology. The conservation of these receptors may also be linked to the fact that tsetse flies generally inhabit ecologies with plenty of vertebrate hosts as potential sources of blood meal, a fact that may require a stringent modulation of all odor perceptions via the iGluRs and mGluRs so as to isolate just specific suitable hosts from their smell emanations.

Evolutionarily, tsetse flies have been thought to be closely related to the fruit flies. The current results reveal that majority of the *G. m. morsitans* glutamate-gated receptors have closer ancestral relations to other drosophilid species than to the model fruit fly, *D. melanogaster*, sharing over 80% sequence similarity (see Table 4 and 5). Each of the *G. m. morsitans* glutamate-gated receptors cluster with their *D. melanogaster* homologs and not to the facultative blood-sucker *An. gambiae*, except in mGluRs where some members cluster with homologs from *An. gambiae* (see Figure 3.15). Some *G. m. morsitans* GluRs have high sequence similarities to blood-feeding *Calcitrans irritans* partial genes compared to the mosquito species. This suggests the *G. m. morsitans* chemoreception and chemo-modulation proteins may have diverse relationships to different insect species, pointing to the need to carry-out an in-depth comparative evolutionary study with other Diptera. The diversity is also extended by the mixture of ligand-interacting residues; meaning some of the *G. m. morsitans* glutamate-gated receptors might bind some yet unknown ligands other than the glutamate and/or glycine.

4.3 The *Glossina m. morsitans* chemosensory responsive proteins – OBPs, CSPs, and SNMPs

Odorant binding proteins, OBPs

The repertoires of *G. m. morsitans* OBPs (32) are fewer than those reported in other diptera (Foret & Maleszka, 2006; Vieira & Rozas, 2011; Manoharan *et al*., 2013). They were less by 19, 35, 79 and 17 relative to *D. melanogaster, An. gambiae, Ae. aegypti* and *T. casteneum*, respectively, but were 11 more relative to the honey bee *A. mellifera* (see Table 1), and *Ae.*

albopictus (Deng *et al.*, 2013). The *G. m. morsitans* OBPs contained general odorant and/or pheromone binding ontologies located in sensillar fluid, a hydrophilic media that cannot be traversed by hydrophobic odorants. In addition, they had conserved canonical six cysteine sites and conserved signature patterns typical of their homologs in fruit fly (Xu *et al.*, 2009). It means the *G. m. morsitans* OBPs can potentially form three disulfide bridges to conform to 3D structure for odorant ligand binding. The *G. m. morsitans* OBPs clustered heterogeneously with their counterparts from the *D. melanogaster* and *An. gambiae*, into three main clusters 1-3, corresponding to Minus C, Plus C, and GOBP/PBP classifications (Hekmat-Scafe *et al.*, 2002; Pelosi *et al.*, 2005; 2006; Manharan *et al.*, 2013). The phylogenetic analysis revealed *G. m. morsitans* lacks the atypical class of OBPs that is only present in *An. gambiae* (Manharan *et al.*, 2013; see also cluster 1A in Figure 3.24).

Structurally, insect OBPs are known to be encoded with sets of members in close proximity in their respective genomes (Hekmat-Scafe, *et al.*, 2002; Gong *et al.*, 2009; Deng *et al.*, 2013). However, it is inconclusive how close these non-receptor genes are in the *G. m. morsitans* genome, because the genome scaffolds are yet to be organized into chromosomes. Nevertheless, few members of *G. m. morsitans* OBPs were encoded on same scaffolds. For instance, *GmmOBP5A*, *OBP5B* and *OBP14* were on scaffold 7180000648041; *GmmOBP4*, *OBP7* and *OBP11* on scaffold 7180000648638; and *GmmOBP20* and *OBP21* were on scaffold 7180000649017

Five *G. m. morsitans* OBPs, *GmmOBP1*, *OBP2B*, *OBP8A*, *OBP21*, and *OBP22* had high expression profiles; the rest either had lower expression profile data or had no supportive data at all. For instance, *GmmOBP16*, *OBP17*, *OBP19*, and *OBP25* had no supportive transcription data. An established classical example is the antennal specific OBPs implicated in binding of different pheromones. The homologs of the first mosquito OBP to be described, CquiOBP1 in *Cx. quinquefasciatus* Say (Ishida *et al.*, 2002) have been demonstrated to bind an oviposition pheromone exclusively expressed on antennal neurons (Mao *et al.*, 2010; Pelletier *et al.*, 2010; Deng *et al.*, 2013). In the *G. m. morsitans*, the putative CquiOBP1 homologs could be GmmOBP8A, OBP8B, OBP9, OBP10 and a distantly branched OBP13, as revealed from the cladogram tree analysis where they also clustered with *An. gambiae* homologs AgamOBP2, OBP3, OBP10, OBP15, and OBP17, supporting earlier report by (Liu *et al.*, 2010). Potentially it means the *G. m. morsitans* OBPs may have a role in detecting suitable larviposition sites, often places that have previously been used for the purpose (Saini *et al.*, 1996). Another OBP that mediate olfactory detection of pheromones is LUSH (*D. melanogaster* OBP76a) protein, crucial in optimizing male flies courting females. The functionality of LUSH protein involves a complex of homologs of *D.*

melanogaster receptor proteins DmelOr67d (Kurtovic *et al.*, 2007; Smith, 2007; Swarup *et al.*, 2011) and SNMPs (Benton *et al.*, 2007; Jin *et al.*, 2008; Vogt *et al.*, 2009). In the *D. melanogaster,* the LUSH protein tunes up the Or67d expressing neurons (exclusively trichoid, T1 sensillae) to endogenous optimum levels of the pheromone cVA (Laughlin *et al.*, 2008). In *G. m. morsitans,* a likely homolog of LUSH detected from reciprocal blasts was GmmOBP26, and the homologs of DmelOr67d are GmmOR41-46, as presented above. Whether these sets of proteins (OR41-46, OBP26 and SNMPs) mediate social and sexual behaviors in tsetse flies as is the case in other insects remains to be investigated. It is believed that tsetse does not produce pheromone odors, and it is contestable trying to link the functionality of these proteins to detection of cVA in tsetse fly (personal communication). The only known conspecific pheromone in *G. m. morsitans* is the female cuticular contact pheromone morsilure, (15, 19, 23-trimethyl-heptatriacontane) (den Otter, 1985).

Chemosensory-specific proteins, CSPs

In contrast to OBPs, the repertoire of *G. m. morsitans* CSPs (five) are comparable to those in fruit fly (four) (Vieira & Rozas, 2011) and honey bee (six) (Foret & Maleszka, 2006). The *G. m. morsitans* CSPs had conserved canonical four cysteine sites, a similar number contained in their homologs in fruit fly (Xu *et al.*, 2009; Murphy *et al.*, 2013). The *GmmCSP1, CSP2* and *CSP3* were previously confirmed to be expressed in the antennae of both *G. m. morsitans* sexes, with CSP2 having the highest expression in female *G. m. morsitans,* thus their likely involvement in olfaction (Liu *et al.*, 2012). In contrast, the expression level revealed here for CSP2 is much lower, to be linked to the crucial olfactory demand in a female pregnant *G. m. morsitans.* The *G. m. morsitans* CSP1 and CSP4 seemed divergent with no specific homolog from both the *D. melanogaster* and *An. gambiae,* but see (Liu *et al.*, 2012) for a similar analysis that included *Ae. aegypti* and *Cu. quinquefasciatus*. Functional classification revealed *GmmCSP3* and *CSP4* to be related to ejaculatory bulb protein type III; they also had high expression level. These functions are yet to be validated experimentally.

CD36-like genes

The *G. m. morsitans* CD36-like genes were most conserved across different insect lineages (Nichols & Vogt, 2008; see Table 6). The *Glossina* have similar number of genes encoding CD36-like proteins in the genome as are in other insects, two genes encoding SNMP specific proteins, SNMP1 and SNMP2. This means the CD36-like functional mechanisms may also be conserved in the *G. m. morsitans.* The *G. m. morsitans SNMPs* had ontologies similar to glycolipid membrane bound proteins with potential odorant signaling activities. The *Glossina* SNMPs had characteristic two trans-membrane domains with an extensive extracellular domain containing six cysteine sites with heavily N-glycosylated sites. The

expression levels of SNMPs were relatively low, accounting for slightly over 1% of total RNA-Seq data for all CD36-like genes (n = 611505 RPKM). The GmmSNMP1 and SNMP2 had 80% and 20% expression level respectively (n = 5238 RPKM). An interesting genomic arrangement was observed on scaffold 7180000652157, in which *GmmCD36-9* (on reverse strand), *CD36-10* and *CD36-11* (on forward strand) were flanked at the 5-prime end by *GmmCSP4* located on forward strand and at the 3-prime end by *GmmCSP1* and *CSP3* located on reverse strand. Additionally, *GmmCD36-3, CD36-4, CD36-5* and *CD36-6* were encoded on same scaffold (7180000648975), with gene CD36-6 having the only splice variant detectable amongst this class of genes. This suggests genomic clustering of these accessory chemosensory genes..

The *G. m. morsitans* SNMP2 had relatively more transcripts supporting its expression than SNMP1. The *GmmSNMP2* also had two transmembrane domains located closer to the C-terminal end, a similar observation that have been made in respect to SNMP1 homologs (Benton *et al.*, 2007). Further, *GmmSNMP2* also had many glycosylation sites spread across its sequence, and a single detectable S-palmitoylation site at C497 (see Figure 3.20), instead of the expected intracellular four sites immediately flanking the trans-membrane domains (Tao *et al.*, 1996). The glycosylations enhance solubility of lipoproteins for fast cellular internalization while palmitoylation promotes correct anchorage on the membrane. These suggest that these SNMPs may be expressed in different olfactory-related cells, ie neurons and accessory cells; they may also play different roles within the chemosensory media. Differential expressions of SNMP homologs were previously reported in Lepidoptera (Liu *et al.*, 2014). Additionally, the homologs of SNMP1 have been shown to participate in both male specific cVA sex pheromone detection and/ or activation of related olfactory neurons on which they are expressed (Benton *et al.*, 2007; Vogt *et al.*, 2009; Pelletier & Leal, 2011; Anderson *et al.*, 2011).

In conclusion, the chemosensory receptors ORs, GRs, and IRs may function directly by binding hydrophilic odors and tastes, or indirectly via the action of intermediate soluble proteins like OBPs and CSPs, and membrane proteins like SNMPs. These results suggest the *G. m. morsitans* OBPs, CSPs and SNMPs could form similar 3-D structures with potential for related functionalities as their homologs in other insects (Pelosi *et al.*, 2006; Jin *et al.*, 2008; Vogt *et al.*, 2009). Comparatively, more RNA-seq data (1,360,124 RPKM) supported expression of *GmmCSPs* compared to *GmmOBPs* and *SNMPs* at 865,561 and 5,238 RPKM respectively, probably an indication that the CSPs have a more ubiquitous role in the fly. However, lower or no reads support for majority of the non-receptor genes is no proof that they are artifacts in the genome. The data available here are insufficient to assess the differential expressions of these genes, since no experiments have been done to compare the expressions in say males and females, different pregnancy and nutrition states, and different

ages. Otherwise, the current data restricts estimation of expression profiles to a pregnant female fly, meaning the remaining lowly expressed genes including those without supportive reads may be transcribed say in males, or different nutrition states. These findings are interesting in that there has been no report on tsetse male aphrodisia cVA ligand. The specific ligands of these sexual proteins (GmmOR41-46, OBP26, and SNMP1 and SNMP2), remain to be determined, and establish if indeed they are functional.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

The *G. m. morsitans* chemoreceptor genes (ORs, GRs and IRs) numbers and those of non-receptor proteins OBPs are fewer, while gene numbers of iGluRs, mGluRs, CSPs and SNMPs are similar to those in *D. melanogaster* and *An. gambiae*. However, specific related chemoreceptor genes appeared to be encoded in multiple copies as species-specific genes, suggesting their important role. For instance, *GmmOR41-46* that are homologous to drosophilid Or67d implicated in pheromone cVA detection; though reports suggest that tsetse does not produce pheromone cVA (personal communication). Alternatively this expansion may be useful in regulating mating process in the *G. m. morsitans* sexual relationships. The *GmmGR1-4* which are homologous to CO_2 sensitive receptors in both *D. melanogaster* and *An. gambiae* could help the tsetse fly to detect and trace vertebrate hosts for blood-meal. Others, *GmmIR84aA-C* and *IR75a-d* gene families are also encoded in multiple copies in *G. m. morsitans* genome, probably their products perform prioritised roles in chemoreception. Some genes like *GmmOR15, GR1-4, IR64a, IR76b,* and *IR8a* have high expression profiles, again suggesting they may play an enhanced role in the tsetse fly. For instance, *GmmOR15* (and its related paralogs *OR14 and OR16*) a homolog of DmelOr45a that is associated with fly larvae escape responses in *D. melanogaster*, accounts for over 90% of all reads that support expression of ORs. This may be important for the survival of the pregnant fly and her larvae given that the RNA-seq reads were derived from pooled samples of pregnant flies. There was a single copy of the Orco in *G. m. morsitans,* as is in other insects examined, sharing over 70% sequence homology. Together with well conserved LUSH-like and SNMP-like genes, these sets of genes reveal prioritization and over-investment in the tsetse genome.

Other than encoding less chemoreceptor genes, the *G. m. morsitans* genome does not encode gustatory sweet taste receptors, corresponding to the report that tsetse flies derive their energy from the amino acids proline and alanine metabolic pathways. There were no homologs recovered for drosophila IR93a, IR41a and GluRIID, which are highly conserved in other insect genomes. The tsetse genome also lacks expansive species-specific gene lineages common in other insects; meaning tsetse flies could be interacting with a narrow range of common semio-chemicals from a select few hosts. Lack of divergent species-specific IRs may mean that the *G. m. morsitans* dedicate IRs to detecting important general vertebrate host volatiles probably with high sensitivity and specificity.

Presence of well conserved loci for iGluRs and mGluRs in *G. m. morsitans* correspond to their homolog roles in the drosophila CNS, where they help modulate overall fast behavioral dynamics while the fly navigates its ecology. The conservation of these receptors may also be linked to the fact that tsetse flies generally inhabit ecologies with plenty of vertebrate hosts as potential sources of blood meal, which may require a stringent modulation of all odor periceptions via the iGluRs and mGluRs so as to isolate just specific suitable hosts.

Structurally, majority of *G. m. morsitans* chemoreceptors and non-receptor proteins were encoded singly scattered across the scaffolds, with just a few specific gene groups clustered within their scaffolds, probably arising from recent duplication events. The *G. m. morsitans* ORs, GRs, and mGluRs were defined by conserved seven trans-membrane domains, while IRs and iGluRs had four trans-membranes in addition to their known ligand binding domains (LBD S1 and S2). The *G. m. morsitans* OBPs and CSPs had conserved canonical six and four cysteine site signatures respectively. In addition to definitive CD36 signature, the *G. m. morsitans* SNMPs had two trans-membrane domains with palmitoylation and glycosylation sites conserved in them. Suggestively, the *G. m. morsitans* OBPs, CSPs and SNMPs could form similar 3-D structures with potential for related functionalities as in other insects.

Evolutionarily, the *G. m. morsitans* glutamate gated and non-glutamate gated receptors reveal diverse relationships to other insects. For instance, the *G. m. morsitans* GR1-4 seem to have higher sequence similarity to house flies (*M. domestica)* and Mediterranean fruit flies (*C. capitata)* than to the fruit flies or to the facultative blood-feeding mosquitoes. Majority of the *G. m. morsitans* glutamate-gated receptors have homologies to other drosophilid species other than the model *D. melanogaster*. These suggest the *G. m. morsitans* chemoreception and chemo-modulation proteins have diverse relationships to different insect species. The *G. m. morsitans* OBPs clustered into three main classes corresponding to Minus C, Plus C and GOBP/PBP classifications, but lacked the atypical class of OBPs present in only in *An. gambiae*. The *G. m. morsitans* CSP1 and CSP4 seemed divergent from their homologs in *D. melanogaster* and *An. gambiae.* Amongst the CD36-like genes, which were highly conserved as in other insect lineages, the *G. m. morsitans SNMP2* seemed to be a prioritized co-expressed gene in probably detection of yet unknown ligand.

5.2 Conclusion

The *G. m. morsitans* genome has a contracted chemoreceptome, probably retaining only ancestrally conserved and crucial orthologous genes. Such highly conserved genes included the OR41-OR46, which are all homologous to drosophila Or67d,

in drosophila the homolog is implicated in detection of mating pheromone cVA; the GR1-GR4, which are homologous to CO_2 sensitive receptor in fruit flies and mosquitoes; and antennal expressed olfactory IRs including IR8a, IR21a, IR31a, IR25a, IR40a, IR64a, IR68a, IR75d, IR76b, and IR84a. The fly seems to prioritize and invest on these specific few chemoreceptor genes towards their ecological adaptive behaviors. This is by encoding more copies of such gene set or by increasing the transcription level of specific gene.

Chemoreceptors with broad lineage expansions in other insects are all unrecoverable in the *Glossina* genome. In those other insects, such receptors define their broad ecological host range, meaning the tsetse fly narrow host range can be defined at molecular level. Key gene lineages, IR93a, IR41a and GluRIID were not recovered in the genome, yet they are reportedly conserved in protostomes. *Glossina* lacked all sweet taste receptor homologs Gr5a and Gr64a, which are conserved and expanded in plant sap/juice sucking insects. One gene lineage (IR84a) is encoded in multiple copies in the *Glossina*, in contrast with its single copy homolog found only amongst drosophila, implicated in both social and sexual behaviors. Functionally validating this gene lineage will be important.

Overall, the *Glossina* genes whose homologs are known to be expressed in the CNS like iGluRs and mGluRs had higher putative expression levels corresponding to their ubiquitous neuronal modulation role. In contrast, the periceptors like ORs, GRs and IRs had lower expression levels except putative co-receptors that had higher expression levels. Amongst chemosensory-related proteins, CSPs seemed to have higher expression levels than OBPs and SNMPs, probably an indication that they have a more ubiquitous role in the tsetse fly.

While the annotated chemosensory-related genes of *D. melanogaster* were used to search for their homologs in the *G. m. morsitans* genome, reciprocal blast searches using the tsetse proteins onto non-redundant databases revealed a closer evolutionary relationships with not only fruit flies, but also other flies like house flies (*M. domestica)*, Mediterranean fruit flies (*C. capitata*), and Muscidae fly *C. irritans.*

5.3 Recommendations

The repertoires of genes reported here are based on the currently available genome assembly GMOY1.1 Yale strain. It is possible that more genes could be recovered with revised versions of the genome assembly in future. While the findings reveal putatively a contracted chemoreceptome in *G. m. morsitans*, there is need to perform experimental validations to identify functional receptor genes and pseudogenes. It requires in-depth functional studies to establish the differential

expressions of all the genes, including characterization of the receptor-odor space. Never-the-less, these findings are interesting in that there has been no report on tsetse male aphrodisiac cVA. It remains to be determined the specific ligands of these sexual proteins, GmmOR41-46, OBP26, and SNMP1 and 2, and if indeed they are functional.

Some genes had very low or no supportive RNA-seq data to estimate their expression levels. This is, however, no proof of them being artifacts in the genome. The data available here are insufficient to assess the differential expressions of these genes, since no experiments have been done to compare the expressions in say males and females, different pregnancy and nutritional states, and different ages. The current data restricts estimation of expression levels to a pregnant female fly, meaning the remaining lowly expressed genes including those without supportive reads may be transcribed say in males, or different feeding states.

There is need to characterize the regulatory constructs including transcription factor binding sites (TFBS) of all the genes so as to facilitate performance functional expression studies on any class of the genes. Never-the-less, the results presented and discussed here lay the foundation for any future comparative studies involving any species of tsetse flies and with other closely related insects. This, among other things, will establish clearly the evolutionary relationships between the tsetse flies and other insects with respect to underlying proteins responsible for ecological behaviors. Further, the results also open the door for functional and neurological studies to be carried out. It will be possible to establish the receptor-odor-sensilla-neuron pathway, and possibly make the *G. m. morsitans* a model neurobiology insect. In addition, the results form the basis for re-examining new approaches for improving tsetse control tools including trap and target olfactory lures and possible drug targets based on chemoreception.

CITED REFERENCES

Abuin, L., Bargeton, B., Ulbrich, M. H., Isacoff, Ehud Y, S. K., & Benton, R. 2011. Functional architecture of olfactory ionotropic glutamate receptors. *Neuron* **69**: 44–60.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., et al. 2000. The Genome Sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195. Retrieved March 20, 2014, from http://www.sciencemag.org/cgi/doi/10.1126/science.287.5461.2185

Ai, M., Blais, S., Park, J.-Y., Min, S., Neubert, T. a., & Suh, G. S. B. 2013. Ionotropic Glutamate Receptors IR64a and IR8a Form a Functional Odorant Receptor Complex In Vivo in Drosophila. *Journal of Neuroscience* **33**: 10741–10749. Retrieved June 27, 2013, from http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5419-12.2013

Aksoy, S., 2003. Control of tsetse flies and trypanosomes using molecular genetics. *Veterinary Parasitology*, 115(2), pp.125–145. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0304401703002036 [Accessed May 6, 2011].

Aksoy, S., Berriman, M., Hall, N., Hattori, M., Hide, W., & Lehane, M. J. 2005. A case for a Glossina genome project. *Trends Parasitol* **21**: 107–111. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15734656

Altschul, S. F., & Lipman, D. J. 1990. Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A* **87**: 5509–5513. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2196570

Alves-Silva, J., Ribeiro, J. M., Van Den Abbeele, J., Attardo, G., Hao, Z., Haines, L. R., Soares, M. B., Berriman, M., Aksoy, S., & Lehane, M. J. 2010. An insight into the sialome of *Glossina morsitans morsitans*. *BMC Genomics* **11**: 213. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20353571

Andersson, M. N., Grosse-Wilde, E., Keeling, C. I., Bengtsson, J. M., Yuen, M. M., Li, M., Hillbur, Y., Bohlmann, J., Hansson, B. S., & Schlyter, F. 2013. Antennal transcriptome analysis of the chemosensory gene families in the tree killing bark beetles, *Ips typographus* and *Dendroctonus ponderosae* (Coleoptera: Curculionidae: Scolytinae). *BMC genomics* **14**: 198. Retrieved March 24, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/23517120

Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnocci, A., Petacchi, R., & Pelosi, P. 1999. Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *Eur J Biochem* **262**: 745– 754.

Arensburger, P., Megy, K., Waterhouse, R., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., Campbell, C., Campbell, K., Casola, C., Castro, M., Chandramouliswaran, I., et al. 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science* **330**: 86–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20929810

Attardo, G. M., Ribeiro, J. M., Wu, Y., Berriman, M., & Aksoy, S. 2010. Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the tsetse fly (*Glossina morsitans morsitans*). *BMC genomics* **11**: 160. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2846916&tool=pmcentrez&rendertype=abstract

Attardo, G. M., Strickler-Dinglasan, P., Perkin, S. A. H., Caler, E., Bonaldo, M. F., Soares, M. B., El-Sayeed, N., & Aksoy, S. 2006. Analysis of fat body transcriptome from the adult tsetse fly, *Glossina morsitans morsitans*. *Insect molecular biology* **15**: 411–24. Retrieved August 1, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/16907828

Bailey, T. L. 2008. Discovering sequence motifs. *Methods in molecular biology* **452**: 231–51. Retrieved March 28, 2011, from http://www.springerlink.com/content/k712222066900072/

Bailey, T. L., & Elkan, C. 1994. MEME: "Fitting a mixture model by expectation maximization to discover motifs in biopolymers." *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*: 28–36.

Bailey, T. L., Boden, M., Buske, F. a, Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**: W202–8. Retrieved October 29, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2703892&tool=pmcentrez&rendertype=abstract

Bairoch, A., & Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic acids research* **28**: 45–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102476&tool=pmcentrez&rendertype=abstract

Baker, N., Alsford, S. & Horn, D., 2011. Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. *Molecular & Biochemical Parasitology*, 176(1), pp.55–57. Available at: http://dx.doi.org/10.1016/j.molbiopara.2010.11.010.

Barrett, M. P., Burchmore, R. J. S., Stich, A., Lazzari, J. O., Frasch, A. C., & Cazzulo, J. J. 2003. The trypanosomiases. *The Lancet* **362**: 1469–1480.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C., & Eddy, S. R. 2004. The Pfam protein families database. *Nucleic acids research* **32**: D138–41. Retrieved March 20, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=308855&tool=pmcentrez&rendertype=abstract

Baxevanis, A. D., & Ouellette, B. F. F. 2001. *BIOINFORMATICS A Practical Guide to the Analysis of Genes and Proteins*. New York: John Wiley & Sons, Inc. Retrieved from www.Wiley.com

Bellmann, D., Richardt, A., Freyberger, R., Nuwal, N., Schwärzel, M., Fiala, A., & Störtkuhl, K. F. 2010. Optogenetically Induced Olfactory Stimulation in Drosophila Larvae Reveals the Neuronal Basis of Odor-Aversion behavior. *Frontiers in behavioral neuroscience* **4**: 27. Retrieved March 25, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2889724&tool=pmcentrez&rendertype=abstract

Benoit, J. B., Attardo, G. M., Michalkova, V., Krause, T. B., Bohova, J., Zhang, Q., Baumann, A. A., Mireji, P. O., Takáč, P., Denlinger, D. L., Ribeiro, J. M., & Aksoy, S. 2014. A Novel Highly Divergent Protein Family Identified from a Viviparous Insect by RNA-seq Analysis: A Potential Target for Tsetse Fly-Specific Abortifacients. *PLOS Genetics.* **10**: e1003874.

Benton, R. 2006. On the ORigin of smell: odorant receptors in insects. *Cellular and molecular life sciences : CMLS***63**: 1579–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16786219

Benton, R., Sachse, S., Michnick, S. W., & Vosshall, L. B. 2006. Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS biology* **4**: e20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16402857

Benton, R., Vannice, K. S., & Vosshall, L. B. 2007. An essential role for a CD36-related receptor in pheromone detection in Drosophila. *Nature* **450**: 289–93. Retrieved October 8, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/17943085

Benton, R., Vannice, K. S., Gomez-Diaz, C., & Vosshall, L. B. 2009. Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell* **136**: 149–162. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2709536&tool=pmcentrez&rendertype=abstract

Benton, R., Vannice, K. S., Gomez-Diaz, C., & Vosshall, L. B. 2009. Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell* **136**: 149–62. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2709536&tool=pmcentrez&rendertype=abstract

Bionaz, M., & Loor, J. J. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. *Bioinformatics and biology insights* **5**: 83–98. Retrieved September 11, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3118679&tool=pmcentrez&rendertype=abstract

Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., Down, T., Eyras, E., Fernandez-Suarez, X. M., Gane, P., Gibbins, B., et al. 2004. An overview of Ensembl. *Genome research* **14**: 925–8. Retrieved March 19, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=479121&tool=pmcentrez&rendertype=abstract

Birney, E., Clamp, M., & Durbin, R. 2004. GeneWise and Genomewise. *Genome research* **14**: 988–95. Retrieved July 20, 2011, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=479130&tool=pmcentrez&rendertype=abstract

Bogdanik, L., Mohrmann, R., Ramaekers, A., Bockaert, J., Grau, Y., Broadie, K., & Parmentier, M.-L. 2004. The Drosophila metabotropic glutamate receptor DmGluRA regulates activity-dependent synaptic facilitation and fine synaptic

morphology. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24**: 9105–16. Retrieved March 30, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/15483129

Bogner, F. 1992. Response properties of C02-sensitive receptors in tsetse flies (Diptera :*Glossina palpalis*). *Physiological Entomology* **17**: 19–24. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3032.1992.tb00985.x/pdf

Bohbot, J., Pitts, R., Kwon, H., Rutzler, M., Robertson, H., & Zwiebel, L. 2007. Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Mol Biol*. **16**: 525–537. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17635615

Boratyn, G. M., Schaffer, A. a, Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. 2012. Domain enhanced lookup time accelerated BLAST. *Biology direct* **7**: 12. Retrieved July 13, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/22510480

Bouyer, J., Pruvot, M., Bengaly, Z., Guerin, P. M., & Lancelot, R. 2007. Learning influences host choice in tsetse. *Biology letters* **3**: 113–116.

Braga, M. F. M., Vassiliki Aroniadou-Andersjaska & He Li 2004. The Physiological Role of Kainate Receptors in the Amygdala. *Molecular Neurobiology* **30** (2): 127-141.

Broman, J., Hassel, B., Rinvik, E., & Ottersen O.E. 2000. *Biochemistry and anatomy of transmitter glutamate*.

Buck, L., & Axel, R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**: 175–187. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1840504

Burge, C., & Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**: 78–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9149143

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., & Yandell, M. 2008. MAKER : An easy-to-use annotation pipeline designed for emerging model organism genomes. **18**: 188–196. DOI:10.1101/gr.6743907.1

Carey, A. F., & Carlson, J. R. 2011. Insect olfaction from model systems to disease control. *PNAS* **108**: 12987–12995. Retrieved May 22, 2013, from /pmc/articles/PMC3156210/?report=abstract

Carver, T., Berriman, M., Tivey, A., Patel, C., Bohme, U., Barrell, B. G., Parkhill, J., & Rajandream, M. A. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**: 2672–2676. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18845581

Chun-Jen Lin, C., Summerville, J. B., Howlett, E., & Stern, M. 2011. The metabotropic glutamate receptor activates the lipid kinase PI3K in Drosophila motor neurons through the calcium/calmodulin-dependent protein kinase II and the nonreceptor tyrosine protein kinase DFak. *Genetics* **188**: 601–13. Retrieved March 9, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3176550&tool=pmcentrez&rendertype=abstract

CLC bio. 2012. *CLCGenomics Workbench: User Manual for CLC Genomics Workbench 5.1 Windows, Mac OS X and Linux*.

Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., Carlson, J. R., & Haven, N. 1999. A Novel Family of Divergent Seven-Transmembrane Proteins : Candidate Odorant Receptors in Drosophila.*Neuron* **22**: 327–338.

Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., Carlson, J. R., & Haven, N. 1999. A Novel Family of Divergent Seven-Transmembrane Proteins : Candidate Odorant Receptors in Drosophila.*Neuron* **22**: 327–338.

Cole, C., Barber, J. D., & Barton, G. J. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* **36**: W197–W201. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18463136

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)* **21**: 3674–6. Retrieved November 2, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/16081474

Croset, V., Rytz, R., Cummins, S. F., Budd, A., Brawand, D., Kaessmann, H., Gibson, T. J., & Benton, R. 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS genetics* **6**: e1001064. Retrieved March 16, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2924276&tool=pmcentrez&rendertype=abstract

Curwen, V., Eyras, E., Andrews, T. D., Clarke, L., Mongin, E., Searle, S. M. J., & Clamp, M. 2004. The Ensembl automatic gene annotation system. *Genome research* **14**: 942–50. Retrieved June 11, 2011, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=479124&tool=pmcentrez&rendertype=abstract

Dahanukar, A., & Ray, A. 2011. Courtship, aggression and avoidance: Pheromones, receptors and neurons for social behaviors in Drosophila. *Fly* **5**: 58–64. Retrieved November 29, 2012, from http://www.landesbioscience.com/journals/fly/article/13794/

Dahanukar, A., Hallem, E. a, & Carlson, J. R. 2005. Insect chemoreception. *Current opinion in neurobiology* **15**: 423–30. Retrieved December 8, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/16006118

Danysz, W., Parsons, C. G., Bresink, I., & Quack, G. 1995. Glutamate in CNS disorders. *DN&P* **8**: 261–278.

de Brito Sanchez, G., & Giurfa, M. 2011. A comparative analysis of neural taste processing in animals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **366**: 2171–80. Retrieved May 24, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3130363&tool=pmcentrez&rendertype=abstract

de Bruyne, M., & Baker, T. C. 2008. Odor detection in insects: volatile codes. *Journal of chemical ecology* **34**: 882–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18535862

Den Otter, C. J., & Naters, V. der G. van W. M. 1993. Responses of individual olfactory cells of tsetse flies (*Glossina m. morsitans*) to phenols from cattle urine. *Physiological Entomology* **18**: 43–49.

Den Otter, C. J., & van der Goes, W. van N. 1992. Single cell recordings from tsetse (*Glossina m. morsitans*) antennae reveal olfactory, mechano- and cold receptors. *Physiological Entomology* **17**: 33–42.

Deng, Y., Yan, H., Gu, J., Xu, J., Wu, K., Tu, Z., James, A. A., & Chen, X. 2013. Molecular and functional characterization of odorant-binding protein genes in an invasive vector mosquito, *Aedes albopictus*. (W. Blenau, Ed.). *PloS one* **8**: e68836. Retrieved March 26, 2014, from http://dx.plos.org/10.1371/journal.pone.0068836

Ebbs, M. L., & Amrein, H. 2007. Taste and pheromone perception in the fruit fly *Drosophila melanogaster*. *Pflugers Arch - Eur J Physiol* **454**: 735–747.

Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–7.

Engsontia, P., Sanderson, A. P., Cobb, M., Walden, K. K. O., Robertson, H. M., & Brown, S. 2008. The red flour beetle's large nose: An expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology* **38**: 387–397.

Fan, M., Zhang, B., & Li, M. Y. 2010. Mechanisms for stable co-existence in an insect community. *Math Biosci Eng* **7**: 603–622. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20578788

FAO, 1992. The Daiagnosis, Treatment and Prevention of Animal Trypanosomiasis under Field Conditions. In R. J. Connor, ed. *Programme for the control of African animal trypanosomiasis and related development: ecological and technical aspects*. Rome: Agriculture and Consumer Protection, p. 199. Available at: http://www.fao.org/docrep/004/t0599e/t0599e01.htm.

FAO, 1993. Training manual for tsetse control personnel. Vol. 5: Insecticides for tsetse and trypanosomiases control using attractive bait techniques. , 5(88).

FAO, 2008. *COLLECTION OF ENTOMOLOGICAL BASELINE DATA FOR TSETSE AREA-WIDE INTEGRATED PEST MANAGEMENT PROGRAMMES*. (John N. Pollock (FAO), Ed.) (pp. 1–8). FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS 2008, Rome Italy. Retrieved from ftp://ftp.fao.org/docrep/fao/011/i0535e/i0535e.pdf

FAO. 1992. *Training Manual for TSETSE CONTROL PERSONNEL* (J. Pollock, Ed.). Rome, Italy: FOOD AND

AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Retrieved from ftp://ftp.fao.org/docrep/fao/009/p5178e/p5178e00.pdf

Farine, J.-P., Ferveur, J.-F., & Everaerts, C. 2012. Volatile Drosophila cuticular pheromones are affected by social but not sexual experience. *PloS one* **7**: e40396. Retrieved October 26, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3394786&tool=pmcentrez&rendertype=abstract

Felsenstein, J. 1985. PHYLOGENIES AND THE COMPARATIVE METHOD. *The American Naturalist* **125**: 1–15.

Forêt, S., & Maleszka, R. 2006. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Research* **16**: 1404–1413.

Fox, A. N., Pitts, R. J., Robertson, H. M., Carlson, J. R., & Zwiebel, L. J. 2001. Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 14693–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=64743&tool=pmcentrez&rendertype=abstract

Fuss, S. H., & Ray, A. 2009. Mechanisms of odorant receptor gene choice in Drosophila and vertebrates. *Molecular and cellular neurosciences* **41**: 101–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19303443

Fuss, W. 2009. Does life originate from a single molecule? *Chirality* **21**: 299–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18537164

Galizia, C. G., & Rössler, W. 2010. Parallel olfactory systems in insects: anatomy and function. *Annual review of entomology* **55**: 399–420. Retrieved November 25, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/19737085

Gao, Q., & Chess, A. 1999. Identification of candidate Drosophila olfactory receptors from genomic DNA sequence. *Genomics* **60**: 31–9. Retrieved April 10, 2014, from http://www.sciencedirect.com/science/article/pii/S0888754399958949

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C., & Ritchie, M. G. 2008. Drosophila chemoreceptor gene evolution: selection, specialization and genome size. *Molecular ecology* **17**: 1648–57. Retrieved November 8, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/18371013.

Gegelashvili, L. G., Robinson, M., Davide, T., & Thomas, R. 2001. *Regulation of glutamate transporters in health and disease*.

Ghaninia, M., Ignell, R., & Hansson, B. S. 2007. Functional classification and central nervous projections of olfactory receptor neurons housed in antennal trichoid sensilla of female yellow fever mosquitoes, *Aedes aegypti*. *The European journal of neuroscience* **26**: 1611–23. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2121139&tool=pmcentrez&rendertype=abstract.

Gikonyo, N. K., Hassanali, A., Njagi, P. G. N., & Saini, R. K. 2003. Responses of *Glossina morsitans morsitans* to blends of electroantennographically active compounds in the odors of its preferred (buffalo and ox) and nonpreferred (waterbuck) hosts. *Journal of chemical ecology* **29**: 2331–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14682515.

Gikonyo, N. K., Hassanali, A., Njagi, P. G. N., & Saini, R. K. 2003. Responses of *Glossina morsitans morsitans* to blends of electroantennographically active compounds in the odors of its preferred (buffalo and ox) and nonpreferred (waterbuck) hosts. *Journal of chemical ecology* **29**: 2331–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14682515.

Gikonyo, N., Hassanali, A., Njagi, P., Gitu, P., & Midiwo, J. 2002. Odor composition of preferred (buffalo and ox) and non-preferred (waterbuck) hosts of some Savanna tsetse flies. *Journal of chemical ecology* **28**: 969–981.

Gladding, C. M., Fitzjohn, S. M., & Molna, E. 2009. Metabotropic Glutamate Receptor-Mediated Long- Term Depression : Molecular Mechanisms. *PHARMACOLOGICAL REVIEWS* **61**: 395–412.

Gong, D.-P., Zhang, H.-J., Zhao, P., Xia, Q.-Y., & Xiang, Z.-H. 2009. The odorant binding protein gene family from the genome of silkworm, *Bombyx mori*. *BMC genomics* **10**: 332. Retrieved March 26, 2014, from http://www.biomedcentral.com/1471-2164/10/332

Gooding, R. H., & Krafsur, E. S. (2005). Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. *Annual Review of Entomology*, **50**(69), 101–23. doi:10.1146/annurev.ento.50.071803.130443. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15355235.

Grosjean, Y., Rytz, R., Farine, J.P., Abuin, L., Cortot, J., Jefferis, G.S., Benton, R., 2011. An

olfactory receptor for food-derived odours promotes male courtship in

Drosophila. Nature 478, 236e240

Guindon, S., Delsuc, F., Dufayard, J.-F., & Gascuel, O. 2009. *Estimating maximum likelihood phylogenies with PhyML*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19378142

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**: 307–21. Retrieved March 20, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/20525638

Guo, S. & Kim, J. 2007. Molecular evolution of Drosophila odorant receptor genes. *Molecular biology and evolution* **24**: 1198–207. Retrieved November 16, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/17331958

Ha, T. S. & Smith, D. P. 2008. Insect odorant receptors: channeling scent. *Cell* **133**: 761–3. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18510917

Hallem, E.A. & Carlson, J. R. 2006. Coding of odors by a receptor repertoire. *Cell* **125**: 143–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16615896

Hallem, E.A., Dahanukar, A., & Carlson, J. R. 2006. Insect odor and taste receptors. *Annual review of entomology* **51**: 113–35. Retrieved August 3, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/16332206.

Hansson, B. S. 1995. Reviews: Olfaction in Lepidoptera. *Experentia* **51**: 1003–1027. doi:0014-4754/95/111003-2551.50Retrieved from http://www.staff.amu.edu.pl/~osiejuk/Dydaktyka/Sygnaly/konwersatoria/downloads/files/7.3.a.pdf

Hansson, B. S., & Stensmyr, M. C. 2011. Review: Evolution of Insect Olfaction. *Neuron* **72**: 698–711. Retrieved from http://dx.doi.org/10.1016/j.neuron.2011.11.003

Hargrove, J. W. 1976. Tsetse population dynamics. *Most*.

Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J., & Tanouye, M. A. 2002. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome research* **12**: 1357–69. Retrieved March 24, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=186648&tool=pmcentrez&rendertype=abstract

Hildebrand, J.G., 1995. Analysis of chemical signals. *Proc. Natl. Acad. Sci. USA*, 92(1), pp.67–74. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC42818/.

Hill, C. A., Fox, A. N., Pitts, R. J., Kent, L. B., Tan, P. L., Chrystal, M. Aa, Cravchik, A., Collins, F. H., Robertson, H. M., & Zwiebel, L. J. 2002. G protein-coupled receptors in *Anopheles gambiae*. *Science (New York, N.Y.)* **298**: 176–8. Retrieved January 12, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/12364795

Holt, R. A, Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science (New York, N.Y.)* **298**: 129–49. Retrieved July 20, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/12364791

Huerta-Cepas, J., Capella-Gutierrez, S., Pryszcz, L. P., Denisov, I., Kormes, D., Marcet-Houben, M., & Gabaldo, T. 2011. PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic acids research* **39**: D556–D560.

IGGI. 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science (New York, N.Y.)* **344**: 380–6. Retrieved April 28, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/24763584

ILRI, 2011. *African Bovine Trypanosomiasis: Trypanosomiasis Control*, Kenya: ILRI. Available at: http://blip.tv/ilri/why-livestock-4500548.

Ishimoto, H. & Tanimura, T., 2004. Molecular neurophysiology of taste in Drosophila. *Cellular and molecular life sciences : CMLS*, 61(1), pp.10–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14704850 [Accessed November 15, 2012].

Jiao, Y., Moon, S.J. & Montell C., 2007. A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. *Proceedings of the National Academy of Sciences of the United States of America* **104**(35): 14110-14115. DOI: 10.1073/pnas.0702421104.

Jin, X., Ha, T. S., & Smith, D. P. 2008. SNMP is a signaling component required for pheromone sensitivity in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 10996–1001. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2504837&tool=pmcentrez&rendertype=abstract

Jones, D. T., Taylor, W. R., & Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**: 275–282. Retrieved April 11, 2014, from http://bioinformatics.oxfordjournals.org/content/8/3/275.full.pdf+html

Jordan, A., 1995. Control of tsetse flies (Diptera: Glossinidae) with the aid of attractants. *J Am Mosq Control Assoc*, 11, pp.249–255.

Julio-pieper, M., Flor, P. J., Dinan, T. G., & Cryan, J. F. 2011. Exciting Times beyond the Brain : Metabotropic Glutamate Receptors in Peripheral and Non-Neural Tissues. *PHARMACOLOGICAL REVIEWS* **63**: 35–58.

Kamuanga, M., 2003. SOCIO-ECONOMIC AND CULTURAL FACTORS IN THE RESEARCH AND CONTROL OF TRYAPANOSOMIASIS. *FAO*, pp.1–71. Available at: ftp://ftp.fao.org/docrep/fao/005/y4619e/y4619e00.pdf.

Kaufman, T. C., Severson, D. W., & Robinson, G. E. 2002. The Anopheles genome and comparative insect genomics. *Science (New York, N.Y.)* **298**: 97–8. Retrieved November 22, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/12364783

Kent, L. B., Walden, K. K. O., & Robertson, H. M. 2008. The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chemical senses* **33**: 79–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17928357

Kioy, D., Jannin, J., & Mattock, N. (2004). Focus: Human African trypanosomiasis. *Nature Reviews Microbiology*, **2**(3), 186–187. doi:10.1038/nrmicro848

Krafsur, E. S. (2009). Tsetse flies: genetics, evolution, and role as vectors. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, **9***(*1), 124–41. doi:10.1016/j.meegid.2008.09.010

Kreher, S. A., Kwon, J. Y., & Carlson, J. R. 2005. The Molecular Basis of Odor Coding in the Drosophila Larva. *Neuron* **46**: 445–456. Retrieved September 21, 2013, from http://www.sciencedirect.com/science/article/pii/S089662730500320X

Kreher, S.A., Mathew, D., Kim J. & Carlson, J.R., 2008. Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**(1): 110-124. DOI: 10.1016/j.neuron.2008.06.010; PMID 18614033.

Krieger, J., Raming, K., Dewer, Y. M., Bette, S., Conzelmann, S., & Breer, H. 2002. A divergent gene family encoding candidate olfactory receptors of the moth *Heliothis virescens*. *Eur J Neurosci* **16**: 619–628. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12270037

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–580. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11152613

Kugaya, B. A., & Sanacora, G. 2005. Beyond Monoamines : Glutamatergic Function in Mood Disorders. *CNS Spectr.* **10**: 808–819.

Kurtovic, A., Widmer, A., & Dickson, B. J. 2007. A single class of olfactory neurons mediates behavioural responses to a Drosophila sex pheromone. *Nature* **446**: 542–6. Retrieved May 22, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/17392786

Laughlin, J. D., Ha, T. S., Jones, D. N. M., & Smith, D. P. 2008. Activation of pheromone-sensitive neurons is mediated by

conformational activation of pheromone-binding protein. *Cell* **133**: 1255–65. Retrieved March 24, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/18585358

Lawniczak, M. K. N., Emrich, S. J., Holloway, A.a. K., Regier, Aa. P., Olson, M., White, B., Redmond, S., Fulton, L., Appelbaum, E., Godfrey, J., Farmer, C., Chinwalla, a., Yang, S.-P., Minx, P., Nelson, J., et al. 2010. Widespread Divergence Between Incipient Anopheles gambiae Species Revealed by Whole Genome Sequences. *Science* **330**: 512–514. Retrieved October 22, 2010, from http://www.sciencemag.org/cgi/doi/10.1126/science.1195755

Lawson, D., Arensburger, P., Atkinson, P., Besansky, N. J., Bruggner, R. V, Butler, R., Campbell, K. S., Christophides, G. K., Christley, S., Dialynas, E., Hammond, M., Hill, C. a, Konopinski, N., Lobo, N. F., MacCallum, R. M., et al. 2009. VectorBase: a data resource for invertebrate vector genomics. *Nucleic acids research* **37**: D583–7. Retrieved November 7, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686483&tool=pmcentrez&rendertype=abstract

Le, S. Q., & Gascuel, O. 2008. An improved general amino acid replacement matrix. *Molecular biology and evolution* **25**: 1307–20. Retrieved March 26, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/18367465

Lehane, M. 2005. *The Biology of Blood-Sucking in Insects*. Cambridge, UK: CUP.

Lehmann, K., Steinecke, A., & Bolz, J. 2012. GABA through the ages: regulation of cortical function and plasticity by inhibitory interneurons. *Neural plasticity* **2012**: 892784. Retrieved March 27, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3390141&tool=pmcentrez&rendertype=abstract

Liu, C., Zhang, J., Wang, G. & Dong, S. 2014. Expression of SNMP1 and SNMP2 genes in antennal sensilla of *Spodoptera exigua* (HUBNER). *Arch. Insect. Biochem. Physiol.* **85**(2): 114-126. Doi:10.1002/arch.21150

Liu, R., He, X., Lehane, S., Lehane, M., Hertz-Fowler, C., Berriman, M., Field, L. M., & Zhou, J.-J. 2012. Expression of chemosensory proteins in the tsetse fly *Glossina morsitans morsitans* is related to female host-seeking behaviour. *Insect molecular biology* **21**: 41–8. Retrieved November 7, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/22074189

Liu, R., Lehane, S., He, X., Lehane, M., Hertz-Fowler, C., Berriman, M., Pickett, J. a, Field, L. M., & Zhou, J.-J. 2010. Characterisations of odorant-binding proteins in the tsetse fly *Glossina morsitans morsitans*. *Cellular and molecular life sciences : CMLS***67**: 919–29. Retrieved June 18, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/20012146

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. 2000. *Molecular Cell Biology* (W. H. Freeman, Ed.). New York.

Logan, J., & Birkett, M. 2007. Semiochemicals for biting fly control: their identification. *Pest Management Science* **63**: 647–657.

Maleszka, R., & Stange, G. 1997. Molecular cloning, by a novel approach, of a cDNA encoding a putative olfactory protein in the labial palps of the moth *Cactoblastis cactorum*. *Gene* **202**: 39–43. Retrieved April 13, 2014, from http://www.sciencedirect.com/science/article/pii/S0378111997004484

Manoharan, M., Ng Fuk Chong, M., Vaïtinadapoulé, A., Frumence, E., Sowdhamini, R., & Offmann, B. 2013. Comparative genomics of odorant binding proteins in *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*. *Genome biology and evolution* **5**: 163–80. Retrieved March 26, 2014, from http://gbe.oxfordjournals.org/content/5/1/163.full

Mao, Y., Xu, X., Xu, W., Ishida, Y., Leal, W. S., Ames, J. B., & Clardy, J. 2010. Crystal and solution structures of an odorant-binding protein from the southern house mosquito complexed with an oviposition pheromone. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 19102–7. Retrieved January 26, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2973904&tool=pmcentrez&rendertype=abstract

Marchler-Bauer, A., Lu, S., Anderson, J., Chitsaz, F., Derbyshire, M., Deweese-Scott, C., Fong, J., Geer, L., Geer, R., Gonzales, N., Gwadz, M., Hurwitz, D., Jackson, J., Ke, Z., Lanczycki, C., et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**: D225–9.

Martin, J. P., Beyerlein, A., Dacks, A. M., Reisenman, C. E., Riffell, J. Aa, Lei, H., & Hildebrand, J. G. 2011. The neurobiology of insect olfaction: sensory processing in a comparative context. *Progress in neurobiology* **95**: 427–47. Retrieved May 28, 2013, from http://dx.doi.org/10.1016/j.pneurobio.2011.09.007

Masiga, D. K., Okech, G., Irungu, P., Ouma, J., Wekesa, S., Ouma, B., Guya, S. O., & Ndung'u, J. M. 2002. Growth and mortality in sheep and goats under high tsetse challenge in Kenya. *Trop Anim Health Prod* **34**: 489–501. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12537387

Masse, N. Y., Turner, G. C., & Jefferis, G. S. X. E. 2009. Olfactory information processing in Drosophila. *Current biology : CB***19**: R700–13. Retrieved March 4, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/19706282

Mathew, D., Martelli, C., Kelley-Swift, E., Brusalis, C., Gershow, M., Samuel, A. D. T., Emonet, T., & Carlson, J. R. 2013. Functional diversity among sensory receptors in a Drosophila olfactory circuit. *Proceedings of the National Academy of Sciences of the United States of America* **110**: E2134–43. Retrieved April 7, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3677458&tool=pmcentrez&rendertype=abstract

McBride, C. S., Arguello, J. R., & O'Meara, B. C. 2007. Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**: 1395–416. Retrieved November 16, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2147975&tool=pmcentrez&rendertype=abstract

McGraw, L. A., Gibson, G., Clark, A. G., & Wolfner, M. F. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. *Current biology : CB***14**: 1509–14. Retrieved April 14, 2014, from http://www.sciencedirect.com/science/article/pii/S0960982204006098

McQuilton, P., Pierre, S. E. St., & Thurmond, J. 2012. FlyBase Consortium FlyBase 101 – the basics of navigating FlyBase. *Nucleic Acids Research* **40**: D706–14.

Megy, K., Emrich, S. J., Lawson, D., Campbell, D., Dialynas, E., Hughes, D. S. T., Koscielny, G., Louis, C., Maccallum, R. M., Redmond, S. N., Sheehan, A., & Topalis, P. 2012. VectorBase : improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res* **40**: 1–6. Retrieved from http://www.ncbi.nlm.nih.gov/corehtml/query/egifs/http:--highwire.stanford.edu-icons-externalservices-pubmed-custom-oxfordjournals_open_access.gif.

Miller, R., & Tu, Z. 2008. Odorant Receptor C-Terminal Motifs in Divergent Insect Species. *Journal of Insect Science* **8**: 1–10. Retrieved from http://www.bioone.org/doi/abs/10.1673/031.008.5301

Missbach, C., Dweck, H. K. M., Vogel, H., Vilcinskas, A., Stensmyr, M. C., Hansson, B. S., & Grosse-wilde, E. 2014. Evolution of insect olfactory receptors. *eLife* **3**: e02115.

Montell, C. 2010. A Taste of the Drosophila Gustatory Receptors. *Curr Opin Neurobiol* **19**: 345–353.

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 1–8.

Mugasa, C.M. et al., 2008. Detection of *Trypanosoma brucei* parasites in blood samples using real-time nucleic acid sequence-based amplification. *Diagn Microbiol Infect Dis*, 61(4), pp.440–445. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18486402.

Murphy, E. J., Booth, J. C., Davrazou, F., Port, A. M., & Jones, D. N. M. 2013. Interactions of *Anopheles gambiae* Odorant-binding Proteins with a Human-derived Repellent: IMPLICATIONS FOR THE MODE OF ACTION OF N,N-DIETHYL-3-METHYLBENZAMIDE (DEET). *The Journal of biological chemistry* **288**: 4475–85. Retrieved March 7, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/23261834.

Naur, P., Hansen, K., Kristensen, A., Dravid, S., Pickering, D., Olsen, L., Vestergaard, B., Egebjerg, J., Gajhede, M., & Traynelis, S. 2007. Ionotropic glutamate-like receptor delta 2 binds D-serine and glycine. *Proc Natl Acad Sci USA* **104**: 14116– 14121.

Nakagawa, T. & Vosshall, L. B. 2009. Controversy and Consensus: Non-Canonical Signalling Mechanisms in the Insect Olfactory System. *Curr Opin Neurobiol.* **19** (3): 284-292. doi:10.1016/j.conb.2009.07.015

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z. J., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E. M., Lobo, N. F., Campbell, K. S., Brown, S. E., et al. 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science (New York, N.Y.)* **316**: 1718–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17510324.

Nichols, Z., & Vogt, R. G. 2008. The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: *Drosophila melanogaster, D. pseudoobscura, Anopheles gambiae, Aedes aegypti, Apis mellifera,* and *Tribolium castaneum*. *Insect biochemistry and molecular biology* **38**: 398–415. Retrieved April 13, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/18342246

Niimura, Y., & Nei, M. 2005. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene* **346**: 13–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15716120.

Niswender, C. M., & Conn, P. J. 2010. Metabotropic Glutamate Receptors : Physiology , Pharmacology , and Disease. *Annu. Rev. Pharmacol. Toxicol.* **50**: 295–322.

Nozawa, M., & Nei, M. 2007. Evolutionary dynamics of olfactory receptor genes in Drosophila species. *PNAS* **104**: 7122–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1855360&tool=pmcentrez&rendertype=abstract.

Obiero, G. F. O., Mireji, P. O., Nyanjom, S. R. G., Christoffels, A., Robertson, H. M., & Masiga, D. K. 2014. Odorant and gustatory receptors in the tsetse fly *Glossina morsitans morsitans*. *PLoS Negl Trop Dis* **8**: e2663.

Olivier, V., Monsempes, C., François, M.-C., Poivet, E., & Jacquin-Joly, E. 2011. Candidate chemosensory ionotropic receptors in a Lepidoptera. *Insect molecular biology* **20**: 189–99. Retrieved June 2, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/21091811.

Omolo, M. O., Hassanali, A., Mpiana, S., Esterhuizen, J., Lindh, J., Lehane, M. J., Solano, P., Rayaisse, J. B., Vale, G. a, Torr, S. J., & Tirados, I. 2009. Prospects for developing odour baits to control *Glossina fuscipes* spp., the major vector of human African trypanosomiasis. *PLoS neglected tropical diseases* **3**: e435. Retrieved May 6, 2011, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2674566&tool=pmcentrez&rendertype=abstract

Otter, den C. J. 1991. Olfactory responses of tsetse flies to phenols from buffalo urine. *Physiological Entomology* **16**: 401–410.

Owaga, M., Hassanali, A., & McDowell, P. 1988. The role of 4-cresol and 3-n-propylphenol in the attraction of tsetse flies to buffalo urine. *Insect Science and its Application* **9**: 95–100.

Parra, G. 2000. GeneID in Drosophila. *Genome Research* **10**: 511–515. Retrieved March 30, 2014, from http://www.genome.org/cgi/doi/10.1101/gr.10.4.511

Parra, G., Bradnam, K., & Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)* **23**: 1061–7. Retrieved March 26, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/17332020

Pavlopoulos, E., Trifilieff, P., Chevaleyre, V., Fioriti, L., Zairis, S., & Pagano, A. 2011. Neuralized1 Activates CPEB3 : A Function for Nonproteolytic Ubiquitin in Synaptic Plasticity and Memory Storage. *Cell* **3**: 1369–1383.

Peakall, R., & Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics (Oxford, England)* **28**: 2537–9. Retrieved October 30, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3463245&tool=pmcentrez&rendertype=abstract

Pellegrino, M., & Nakagawa, T. 2009. Smelling the difference: controversial ideas in insect olfaction. *The Journal of experimental biology* **212**: 1973–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19525421

Pelletier, J., & Leal, W. S. 2011. Characterization of olfactory genes in the antennae of the Southern house mosquito, *Culex quinquefasciatus*. *Journal of insect physiology* **57**: 915–29. Retrieved April 15, 2014, from http://www.sciencedirect.com/science/article/pii/S0022191011000953

Pelletier, J., Hughes, D. T., Luetje, C. W., & Leal, W. S. 2010. An odorant receptor from the southern house mosquito *Culex pipiens quinquefasciatus* sensitive to oviposition attractants. *PloS one* **5**: e10090. Retrieved April 9, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2851645&tool=pmcentrez&rendertype=abstract

Pelosi, P., Calvello, M., & Ban, L. 2005. Diversity of Odorant-binding Proteins and Chemosensory Proteins in Insects. *Chem. Senses* **30**: 291–292.

Pelosi, P., Zhou, J., Ban, L. P., & Calvello, M. 2006. Soluble proteins in insect chemical communication. *Cell. Mol. Life Sci*

63: 1658–1676.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785–6. Retrieved March 20, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/21959131

Pittendrigh, B. R., Clark, J. M., Johnston, J. S., Lee, S. H., Romero-Severson, J., & Dasch, G. A. (2006). Sequencing of a new target genome: the *Pediculus humanus humanus* (Phthiraptera: Pediculidae) genome project. *J Med Entomol*, **43**(6), 1103–1111. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17162941

Ramaekers, A., Parmentier, M. L., Lasnier, C., Bockaert, J., & Grau, Y. 2001. Distribution of metabotropic glutamate receptor DmGlu-A in *Drosophila melanogaster* central nervous system. *The Journal of comparative neurology* **438**: 213–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11536189

Reese, M. G., Kulp, D., Tammana, H., & Haussler, D. 2000. Genie-Gene Finding in *Drosophila melanogaster*. *Genome Research* **10**: 529–538. Retrieved April 10, 2014, from http://www.genome.org/cgi/doi/10.1101/gr.10.4.529

Reinhard, J., 2004. Insect Chemical Communication Graham Bell, ed. *Chemo Sense*, 6(4), pp.1–6. Available at: http://www.inscent.com/docs/chemosept04v5.pdf.

Richards, S., Gibbs, R. a, Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Roth, S., Schröder, R., Tautz, D., et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–55. Retrieved January 14, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/18362917

Robertson, H. M. 2009. The Insect Chemoreceptor Superfamily in *Drosophila pseudoobscura* : Molecular Evolution of Ecologically-Relevant Genes Over 25 Million Years . *Journal of Insect Science* **9**: 1–14.

Robertson, H. M., & Wanner, K. W. 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome research* **16**: 1395–403. Retrieved March 22, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/17065611

Robertson, H. M., Warr, C. G., & Carlson, J. R. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **100**: 14537– 14542. Retrieved September 29, 2013, from http://www.pnas.org/content/100/suppl.2/14537.abstract

Rogers, D. J., Hay, S. I., & Packer, M. J. 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Annals of Tropical Medicine and Parasicology* **90**: 225–241. Retrieved July 5, 2011, from http://www.dfid.gov.uk/r4d/SearchResearchDatabase.asp?ProjectID=1306

Rogers, D. J., Hendrickx, G., & Slingenbergh, J. H. 1994. Tsetse flies and their control. *Rev. Sci. tech. Off. Int. Epis* **13**: 1075–1124. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7711306

Rytz, R., Croset, V., & Benton, R. (2013). Ionotropic Receptors: Chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochemistry and Molecular Biology* (In Press).

Saini, R. K., & Hassanali, A. 2007. A 4-alkyl-substituted analogue of guaiacol shows greater repellency to savannah tsetse (Glossina spp.). *Journal of chemical ecology* **33**: 985–95. Retrieved March 25, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/17404820

Saini, R. K., Hassanali, A., Andoke, J., Ahuya, P., & Ouma, W. P. 1996. Identification of major components of larviposition pheromone from larvae of tsetse flies *Glossina morsitans morsitans* Westwood and *Glossina morsitans centralis* Machado. *Journal of Chemical Ecology* **22**: 1211–1220. Retrieved from http://download.springer.com/static/.pdf

Salamov, A. A., & Solovyev, V. V. 2000. Ab initio gene finding in Drosophila genomic DNA. *Genome research* **10**: 516–22. Retrieved March 23, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=310882&tool=pmcentrez&rendertype=abstract

Sanacora, G., Carlos, A., Zarate, J., Krystal, J., & Manji, H. K. 2008. Therapeutics for Mood Disorders. *Nat Rev Drug Discov* **7**: 426–437.

Sánchez-Gracia, A., F, G. V., & Rozas, J. 2009. Molecular evolution of the major chemosensory gene families in insects.

Heredity **103**: 208–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19436326

Sato, K., Pellegrino, M., Nakagawa, T., Vosshall, L. B., & Touhara, K. 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**: 1002–1006. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18408712

Siju, K. P. 2009. Neuromodulation in the chemosensory system of mosquitoes - neuroanatomy and physiology. Doctoral Thesis-Swedish University of Agricultural Sciences Alnarp : 65. Retrieved May 28, 2013, from http://pub.epsilon.slu.se/1956/

Silbering, A. F., & Benton, R. 2010. *Ionotropic and metabotropic mechanisms in chemoreception: "chance or design"?* Retrieved November 13, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2838705&tool=pmcentrez&rendertype=abstract

Simarro, P.P., Jannin, J. & Cattand, P., 2008. Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med*, 5(2), p.e55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18303943.

Smith, D. P. 2007. Odor and pheromone detection in *Drosophila melanogaster*. *Pflügers Archiv : European journal of physiology* **454**: 749–58. Retrieved March 25, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/17205355

Solano, P., Ravel, S. & de Meeûs, T., 2010. How can tsetse population genetics contribute to African trypanosomiasis control? *Trends in parasitology*, (March). Available at: http://www.ncbi.nlm.nih.gov/pubmed/20202905.

Solano, T. & Philippe, S.J., 2010. Olfaction in Glossina-host interactions: a tale of two tsetse. In B. G. J. Willem Takken and Knols, ed. *Olfaction in vector-host interactions: Ecology and control of vector-borne diseases*. Netherlands: Wageningen Academic P u b l i s h e r s, pp. 265 – 290. Available at: http://www.wageningenacademic.com/_clientFiles/download/ecvd-02-e.pdf [Accessed September 22, 2013].

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18218656

Steverding, D., & Troscianko, T. (2004). On the role of blue shadows in the visual behaviour of tsetse flies. *Proceedings. Biological Sciences / The Royal Society*, *271 Suppl 3*(February), S16–7. doi:10.1098/rsbl.2003.0121

Steverding, D., 2008. The history of African trypanosomiasis. *Parasites & vectors*, 1(1), p.3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18275594 [Accessed September 25, 2013].

Stich, A., Abel, P.M. & Krishna, S., 2002. Clinical review: Human African trypanosomiasis. *BMJ*, **325**(27 July), pp.203–206.

Stocker, R. F. 1994. The organization of the chemosensory system in *Drosophila melanogaster*: a rewiew. *Cell and Tissue Research* **275**: 3–26. Retrieved March 21, 2014, from http://link.springer.com/10.1007/BF00305372

Swarup, S., Williams, T. I., & Anholt, R. R. H. 2011. Functional dissection of Odorant binding protein genes in *Drosophila melanogaster*. *Genes, brain, and behavior* **10**: 648–57. Retrieved March 21, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3150612&tool=pmcentrez&rendertype=abstract

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**: 2731–9. Retrieved October 26, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3203626&tool=pmcentrez&rendertype=abstract

Tikhonov, D. B., & Magazanik, L. G. 2009. Origin and Molecular Evolution of Ionotropic Glutamate Receptors EVOLUTIONARY ORIGIN OF GLUTAMATE. *Neuroscience and Behavioral Physiology* **39**: 763–772.

Traynelis, S. F., Wollmuth, L. P., Mcbain, C. J., Menniti, F. S., Vance, K. M., Ogden, K. K., Hansen, K. B., Yuan, H., Myers, S. J., & Dingledine, R. 2010. Glutamate Receptor Ion Channels : Structure , Regulation , and Function. *Pharmacological Reviews* **62**: 405–496.

Vale, G. 1980. Field studies of the response of tsetse flies (Glossinidae) and other Diptera to carbon dioxide, acetone and other chemicals. *Bulletin of Entomological Research* **70**: 563–570.

Vale, G., & Hall, D. 1985. The use of 1-octen-3-ol, acetone and carbon dioxide to improve baits for tsetse flies, Glossina spp. (Diptera: Glossinidae), to host odour. *Bulletin of Entomological Research* **75**: 219–231.

Vale, G., Hall, D., & Gough, A. 1988a. The olfactory responses of tsetse flies, Glossina spp. (Diptera: Glossinidae), to phenols and urine in the field. *Bulletin of Entomological Research* **78**: 293–300.

Vieira, F. G., & Rozas, J. 2011. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome biology and evolution* **3**: 476–90. Retrieved March 21, 2014, from http://gbe.oxfordjournals.org/content/early/2011/04/28/gbe.evr033.short?rss=1

Vieira, F. G., Sánchez-Gracia, A., & Rozas, J. 2007. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. *Genome biology* **8**: R235. Retrieved March 21, 2014, from http://genomebiology.com/2007/8/11/R235

Vogt, R. G., Miller, N. E., Litvack, R., Fandino, R. A., Sparks, J., Staples, J., Friedman, R., & Dickens, J. C. 2009. The insect SNMP gene family. *Insect Biochemistry and Molecular Biology* **39**: 448–456. Retrieved from http://dx.doi.org/10.1016/j.ibmb.2009.03.007

Voskamp, K. E., Natters, V. D. G. W. M., & Den Otter, C. 1999. Comparison of single cell sensitivities to attractants in the tsetse *Glossina fuscipes fuscipes*, *Glossina morsitans morsitans*, and *Glossina pallidipes*. *Medical and Veterinary Entomology* **13**: 460–462. Retrieved from http://doi.wiley.com/10.1046/j.1365-2915.1999.00186.x

Vosshall, L. B., & Stocker, R. F. 2007. Molecular Architecture of Smell and Taste in Drosophila. *Annu. Rev. Neurosci* **30**: 505–33.

Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, Aa, & Axel, R. 1999. A spatial map of olfactory receptor expression in the Drosophila antenna. *Cell* **96**: 725–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10089887

Vosshall, T. N. and Vosshall, L. B. 2009. Mechanisms in the Insect Olfactory System. *Curr Opin Neurobiol* **19**: 284–292.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19151095

Whelan, S., & Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution* **18**: 691–9. Retrieved September 25, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/11319253

WHO, 2004. *STRATEGIC REVIEW OF TRAPS AND TARGETS FOR TSETSE AND AFRICAN TRYPANOSOMIASIS CONTROL* TDR/IDE/TR. C. J. S. F.A.S. Fuzoe, ed., Rome, Italy: World Health Organization -on behalf of UNICEF/UNDP/WORLDBANK/WHO Special Programme for Research and Training on Tropical Diseases, 2004.

WHO. (2000). World Health Report 2000. Health Systems Improving Performance, Geneva, 2000. *World Health Report*.

WHO. (2002). Reducing Risks, Promoting Healthy Life. *The World Health Report 2002 Press Kit*, 1–230.

WHO. (2010). Working to overcome the global impact of neglected tropical diseases: First WHO report on neglected tropical diseases. *World Health Report*, 94–90.

WHO. 2011. Working to overcome the global impact of neglected tropical diseases – Summary. *Weekly epidemiological record / Health Section of the Secretariat of the League of Nations* **86**: 113–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21438440

Wicher, D., 2013. Sensory receptors-design principles revisited. *Frontiers in cellular neuroscience*, 7(1), p.1. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3553420&tool=pmcentrez&rendertype=abstract [Accessed May 26, 2013].

Wicher, D., Schäfer, R., Bauernfeind, R., Stensmyr, M. C., Heller, R., Heinemann, S. H., & Hansson, B. S. 2008. Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* **452**: 1007–11. Retrieved July 13, 2010, from http://www.ncbi.nlm.nih.gov/pubmed/18408711

Xu, X., Dong, Y., Abraham, E., Kocan, A., Srinivasan, P., Ghosh, A., Sinden, R., Ribeiro, J., Jacobs-Lorena, M., Kafatos, F., & Dimopoulos, G. 2005. Transcriptome analysis of *Anopheles stephensi-Plasmodium berghei* interactions. *Mol. Biochem. Parasitol.* **142**: 76–87.

Xu, Y.-L., He, P., Zhang, L., Fang, S.-Q., Dong, S.-L., Zhang, Y.-J., & Li, F. 2009. Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects. *BMC genomics* **10**: 632. Retrieved July 30, 2010, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2808328&tool=pmcentrez&rendertype=abstract

Zhou, J.-J., Huang, W., Zhang, G.-A., Pickett, J. A., & Field, L. M. 2004. "Plus-C" odorant-binding protein genes in two Drosophila species and the malaria mosquito *Anopheles gambiae*. *Gene* **327**: 117–29. Retrieved March 26, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/14960367

Zhou, X., Lin, Z., & Ma, H. 2010. Phylogenetic detection of numerous gene duplications shared by animals, fungi and plants. *Genome Biology* **11**: R38. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2884541&tool=pmcentrez&rendertype=abstract

Zhou, X., Slone, J. D., Rokas, A., Berger, S. L., Liebig, J., Ray, A., Reinberg, D., & Zwiebel, L. J. 2012. Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent Ant Species Reveals Sex-Specific Signatures of Odor Coding (N. A. Moran, Ed.). *PLoS Genetics* **8**: e1002930. Retrieved August 30, 2012, from http://dx.plos.org/10.1371/journal.pgen.1002930

Zwiebel, L. J., & Takken, W. 2004. Olfactory regulation of mosquito-host interactions. *Insect biochemistry and molecular biology* **34**: 645–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15242705

APENDICES

Appendix 1

Dataset S1. Annotated amino acid sequences of *G. m. morsitans* ORs and GRs.

>GmmOR1|Orco GMOY005610 (TMP007718) scf-648683 14760:27560 reverse
MFKQTMANDLQPGKYVGLVADLMPNIKLMKFSGLFMHAFTSGSAVGKKIYSSIHLALILL
QFLSILINMALNADEVNELSGNTITVLFFTHSITKFIYLACSQKNFYRTISIWNQVNTHP
LFAESDARYHSIALAKMRKLFFLVMLTTVTSAIAWITLTFFGESVKFAVDKENNSTMTVE
IPRLPIKSFYPWDASSGISYIVSFIYQAYFILFALSHANLCDVLFCSWLIFACEQLQHLK
SIMKPLMELSASLDTFRPDSGALFRSLSAHSKAELIENEEKEPPPSNGLDLSGIYSSKAD
WGAQFRAPSTLQTFNNNNPNGLTKKQEMMVRSAIKYWVERHKHVVRLVAAIGDTYGVALL
LHMLTATIKLTLLAYQATKITGVNVYALGVVGYLGYSLAQVFHFCIFGNRLIEESSSVME
AAYSCHWYDGSEEAKTFVQIVCQQCQKAMSISGAKFFTVSLDLFASQWSHISWCWCNLNE
VIASFAIKFSSSSSLRPSSLITKRNHGECSKISYLVYLYFV
>GmmOR2 GMOY005796 (TMP007907) scf-648756 29714:32260 reverse
MEYRAHLETNTAFRYHWLVWQLTGIIQPRYFSTSLYRLYTVLVNILVTFLFPLTFVINVF
FSKNLQQLCENLIITLCDITANLKFLNVFLVRKELQQIKIILECLDKRLNTEEEYRQLKR
AIRTAQLSFCIFLVISTTGTTLSLFMVLYSEERSLLFPAWYGLDWQKNDLAYAVCFTYQL
VALVVQAIQNVANDSYPPAYLIILTSHMRNLELRVRSVGLSPANNKENMFLSEKEQQLDF
KKFNDCIEDFIQINRLYDLIQKILSKACLAQFICTALVQCVVGLHILYLLDESDDYGAQI
LSFIFFLAVTMEVFIICYFGHYMSAQSLALIDAFYECGWIPQTLAFRRNLIITLMRTQRY
AILYAGSYIPVDLPTFLLLMKYAYSTFTLLLRFK
>GmmOR3 GMOY004772 (TMP006862) scf-648228 109437:110684 reverse
MVVTKIHTWKAFEYHWRLWTFFGLKPPPRNSVWFKPYVAYAILLNISVTFFFPCTLIINL
ILAKNMNEVCENLYHTITDVVCNIKFLNMFLVRKKLLQINQILKRLDMRAQTLEEITELQ
RGVNSARNCFKIVGRFFCVALITSQLVAYLSPERILMYPAWFPWDWRASKMNFLYAHSYQ
LYGLTLQTAQNLGSDTYPQAYLVVLIGHIKALSLRIKALGSEVTPFSTSDDEFKKKNENH
LYRELVDCIIDHETINDLFRIVQQGISSTCLAQFFCTGLAQCTMGVYTFYVGLDYSKLPN
LIAYFSAVTAEIFILCYYGDLFCQANEKLINSIYSCNWVDPHRFLLQRSQKINTVMAGNI
IPVSLSTFVTVKTESYRLHLHFKSHIM
>GmmOR4 TMP_Or4 scf-642438 111181:112437 forward
MFLDTLQFFRNNWLSWRLLGIVLPKGEQNHKLIKYLLWSTIVNVVATFLFPLHLLLGIFQ
EQPQSTRFESVAICVTSIATSLKFIIYARKLQHVKQMGKLFQRLDARISNDNERQFYQKH
IRRNVIRIQTMFIVVYISVGITVTVAFIFSQERRLFYPGWLPFDWHRSIKYYMAALGFQL
ISIFFQILQNFANDCFTPKALCLLSGHIELLYMRVANIGYVNRVAYQQNRNLRELNECVL
DQKHLYQLFDVIQSIISWPMFLQFLASTVNMCMAMVTLLFFVTDILERIYYVMYFAAMCL
QIFPTCYYGSDFEIKFERLHYAVFSSNWTEGTQCFKRHMMLFTERALRETRAMAGGVFRI
HLDTFFATIKGAYSLFAVVITMKN
>GmmOR5a GMOY012018-PA (TMP002414) scf-639717 1: 6297 MW: 51399.66
SELDKRVSTADECDYFIRQHQRKANSAIKIFFFVYLGTNSFAFMGAIVDQRLMAPAWFPF
DWKSSSTLYWSALLYQFIGLNMLIMQNLVNDTVGPMSLCLLSGHVHLLAMRVARIGYNKR
KSQKYHEDELKLCIEDHQKLLKAFTLLESSMSWLQFILFFTSGLNTCLGVVNFLFYSRIL
YDYIYYGCFLLALGVEVFPCYFYGSVLLEEFKHLPYAIFSSNWVEQSRNYRQNTTIFLEM
ALKSVTMLAGGIVETNLDSFFAIYKAAYSLFTVILTMKIGYNKRKSQKYHEDELKLCIED
HQKLLKAFTLLESSMSWLQFILFFTGGLNTCLGVVNFLFYSRILYDYIYYGCFLLAIGME
VFPCYFYGSVLLEEFKHLPYAIFSSNWVEQSRNYRQNTTIFLEMALKSVTMLAGGIVEIN
LDSFFAIYKAAYSLFTVILTMK
>GmmOr5b GMOY012018-PA (TMP002414) scf-639717 - 1: 6297 MW: 35691.12
SELDKRVSTADECDYFIRQHQRKANSAIKIFFFVYLGTNSFAFMGAIVDQRLMAPAWFPF
DWKSSSTLYWSALLYQFIGLNMLIMQNLVNDTVGPMSLCLLSGHVHLLAMRVARIGYNKR
KSQKYHEDELKLCIEDHQKLLKIGYNKRKSQKYHEDELKLCIEDHQKLLKAFTLLESSMS
WLQFILFFTGGLNTCLGVVNFLFYSRILYDYIYYGCFLLAIGMEVFPCYFYGSVLLEEFK
HLPYAIFSSNWVEQSRNYRQNTTIFLEMALKSVTMLAGGIVEINLDSFFAIYKAAYSLFT
VILTMK
>GmmOr5c GMOY012018-PC (TMP002414) scf-639717 - 5069: 6297 MW: 33056.258
MKIEWWSKLDSTVAYGRFWSCFRILGVADFRYKALSQFYVICLTLFVTVYCPIHLLIGLL
MLTDAGDFFKNFSMTLTSLVCSLKYFFLRLNLNKIHQLIKIYSELDKRVSTADECDYLIG
YNKRKSQKYHEDELKLCIEDHQKLLKAFTLLESSMSWLQFILFFTGGLNTCLGVVNFLFY
SRILYDYIYYGCFLLAIGMEVFPCYFYGSVLLEEFKHLPYAIFSSNWVEQSRNYRQNTTI
FLEMALKSVTMLAGGIVEINLDSFFAIYKAAYSLFTVILTMK
>GmmOr5d |GMOY012018-PD (TMP002414) scf-639717 - 1: 902 MW: 31461.035
SELDKRVSTADECDYFIRQHQRKANSAIKIFFFVYLGTNSFAFMGAIVDQRLMAPAWFPF
DWKSSSTLYWSALLYQFIGLNMLIMQNLVNDTVGPMSLCLLSGHVHLLAMRVARIGYNKR

KSQKYHEDELKLCIEDHQKLLNSMSWLQFILFFTSGLNTCLGVVNFLFYSRILYDYIYYG
CFLLALGVEVFPCYFYGSVLLEEFKHLPYAIFSSNWVEQSRNYRQNTTIFLEMALKSVTM
LAGGIVETNLDSFFAIYKAAYSLFTVILTMKIGYNKRKSQKYHEDELKLCIEDHQKLLKY
LKYSYI
>GmmOR6 GMOY009475 (TMP011670) scf-651846 662444:668878 reverse
MLLALSRGVGVKTQSSQASEYLFNVINFLAQDFHRKWTFGFFFQSFVANGTVVLFMPILF
NLSYLNDSMQFDLGQLFTSVQAAINICAIPIKFITMQLYMKNLHRINEMLNILDGRCKHP
EEFSKLRHCAITGNRIYVGAIIVYFSYSISTCMGFMLTGQAAYNIYIPGINWRNSLFEFV
IQGLVEFASMNLICLHQTVYDSYSGIYLYIIRIHIQILNERVKRLGSNLKNNEKQSYEEL
IQSILDHKLILREGMALKSSRDNQYPLVAAISPMISLTIFVQFAITAAMLATTMINISFF
SNVVGRIASIFYIILVFAQSSLCCYQATCLVSDADELPVSIFHCQWVEQGYRFRKMIIYF
MVNTQKPITFTAGKLFPVNMASVLSNY
>GmmOR7 TMP_Or7-RA scf-651846 655714:658243 reverse
MESVIPNTVEDQQSGERHSINKGRFDSQLNINYTNNVTNAFDRNARTMHGTMFLFNGFRF
LGVYMPEKNRLFYLIWSFLINSTVTVYLPVAFILSFVKISGDDLQIGNLLTSAQVAINVV
GCSVKIILMGFLLPKLLSTEKFLNRLDERCRTAEELASINKFTKQGNRFVVLFSVAYWSY
STSTCISAVAFGRLPYNIYNPFFDHHQSIGHFILSVLMEMTLVNIACFQQVVDDSYAVIY
VNILRTHVDLLLKRIKRLGLTVSMSHEQTYEELKLCIIDHKYIIQFTITATILGTTLINI
LLFATNFASIVASCFYVLAVVVEIFPLCYYTQYLMDESNLLAEVVFHSNWMEQNLPYQKM
LIFFMQRSQRVMEFTAGKLFPITLNSFLSVSES
>GmmOR8 TMP_Or8-RA scf-651846 - 655986:662305 reverse
MLVDTVADYKSPISSFEATTYLFRIYRVLGINKPQRHKNLYIFYTIILNGVVNFLQPIAF
TMNYFTNELVSKLGSLLTSIEIVINVYGCTAKFIIIAWLLPRSNATVQLLKKLDERCQAP
DELELLKSIKKFGRLVVVVLAVSYLSYSVSQFICSLIAGHPPYGLYIPYINWKRSFWEFL
VASSIEFLLMNVSCIVQASNDAYPIIYINILRTHIKILLKRLNRLGTGADKSQDEHLEEL
KLCIKDHKNLNRFIAAFSLFETIAPIISTTVFVQFMITAAIIALTLINMLFFTTNISAVA
NSCVYIADLVLEIFPICYYANCLIDDNDLLSMEIFHSAWPEQNKIYRKMLIVFMQRSQQS
MSLTAGKMFPINLNTFINLQYRFIYNFYNNICAIYNNCHHIGYDSY
>GmmOR9 TMP_Or9-RA scf-652157 136730:138889 reverse
MLTFRKFFKMAHFRGDGPPPKTRDAVLYLFRGITIIGLLTPASNKKCFYVWSLFINSLVT
LYMPVGFLLSFIMRLSSFSPSEFFTSLQISINCVGASLKLFVFSLMYKRLIKATKYMDRL
DKRPIDEQGEFTLRQAVAFSNCSSLLFTLLYLSYSSSTFLSSVINRRPPYQIYNPVIEWR
NSTRNFIIQSAIEYVMIDVHCYQQALLDAYPVIYIMIIRAHLHILSRRISQLGKDDKLKK
NQRYEELVQCVLDHKNIMQLYNLFSPVISGTMFVQFLIIGIILGVTTIHIVLFANFFAIV
ASLFYVASILAETFPCSYLANSLMDDSDELSLAIFHSGWFNEEPRYKQVVVFFLQHTQKT
LVLTAMKIFPITMNSNINVVVKFAFSVYTLISQMDLRQKLKDSVAGQEA
>GmmOR10 TMP_Or10 scf-644232 131824:133273 forward
MDGREHCRYLKTFYKHQCWVFRLFGLWKLSADMTRSHQLLHALYFNYILTIWVLSLDASC
IMQLVIKAGDLNEVIEVFSIFATALAVLAKFLSIKMKNHLYRQWCEIVESSTFRPNNERE
IQLFQQSERLARLIRNAYCVLSLIALNMLIFIVDDRGLPLSVYSPFDMNRKWGYLSTYGY
QYIAASICCYLNIAFDSLSASFLIHLKGQMDILCDRLKNFGKYSECNNDHKITEELKNCI
KYYEEILKVAHLIEDLISLPISIQIICSVLVMVANFFGMAFLAEPGDYAYFFKMLIYQLC
MLSQIYILCYFPNEVTDRSQTISRSLYSAEWFSWSQMNRKLTLLMMDRFDVPIRFRSIIP
TYSFNLAAFTSVFTTKRQQNNILLCFIKKHEQLKYFSANLKLKRNLQRTTKSLIGPKSNN
RETSTADVKERNSWTELDELDEYS
>GmmOR11 GMOY0110761 (TMP012981) scf-652156 1267206:1269387 reverse
MDNSAESVKMLYSKQILIFRLVSIWDLAANAGYYKRLAFAVYFWIVAIFLVALFALLLII
QIFCDINNISEVIRVMFNLASSLTVLGKFLTVKMKNRSFQQLFDLLHSTEYLPRDLKEDK
LFKRALQLSKTVLKVYGGMSLISINTTFLIQFAKDTTELPLPIYEPIDATVPWKYFIMYF
YEYLGFGLCCVMNIAYDSLGAAFFIHIKGQVDILSKRLEEIGHKSSSKQNHSHINEELKA
CAIYYDRILELTHMMEDLMCLPLSIQILFTLTSEKLSLAIYRAGWVDWNEKNRKLALQMM
GRLDLPIRIKTINRCYSFNLAAFTSIVNSSYSYFALLKNFN
>GmmOR12 GMOY009271 (TMP011461) scf-651831 454103:455258 reverse
MLHVPITFTFTTLMWLKVIFSSNLNDVTDVLHMAFTETGLVIKILSAWRFARLLQFVFLE
WQTNNLFSLKSANEQLIWNDQFKSFKITAFVYMSCSLSVVIFSFISVPLMKTYRLPFAFW
TPAGWEQSSYFWYICFYDFIGITFTCISNCTVDMYFCYLLKHITVCLRMISVRLVKLGYS
SEDVTKNLKEIITFHNKLKRMSRHCEKVISYPILGQILFSSMVLCFSIYRLQAISFIETP
FDFLSVFQYMWVMAAQIYLPCHYCNELTLQSGALHTAIYNCNWIAMTAAERKTILLFMLY
IKRPIILKAGHFFEIGLPIFTKTMNNAYSLLALVLNMSDS
>GmmOR13 GMOY003312 (TMP005369) scf-645812 24048:26362 reverse
MIDLFERQRFCLRVMGHNFVRDRSQQSRRWHNTRCTYKYILCLFLVVSAQVPMINYTIYY
LDHVELATASLSICFTNILTTVKIITFLLYKWEFAALMEKMEEMYYEVRKAETKAKLKRD
NDYAGDLTRMYWNSCCCTGAYFMTTPILKIIWSKLQGMDVPLELPMPMRFSFDFESTPGY
EFCYIYTGLTTLVVIAYAVAIDGLFIGFTINLKAHFIALQSIIETNNFLKSESELQRDLS
ACIGYHNKILSLAFKLRDIYRPIIFAQFLMTSLQVCVIVYQMVTVSASQYINAQHIFLKN
CLFLCSILLQLFIYCYGGEILKLESLRVGVSVEISHWYNLKPSHRRILCLLMLRSQREAI
IKAGFYEASLANFTAILKAAMSYITLIQSIE

>GmmOR14 GMOY001365 (TMP003388) scf-641298 67504:69789 forward
MFFKQWLPKADDALPSVNGLSRHFTVQQYTFAAIGLDPKSLQRPIFNKMLALVPMLGLFS
LVIPMIGYASLYKSDILKVTDALSPVWEGLLALAKFFYFIWNRQKVIQLLRKIWIKNLEV
SSNPEELTIIAEANHRDYLFSLTFCINVITTGVLALAAPLIIATFYTLQGEKFLNVLEPP
LKATYPFDFHTPNGLIALYVWDSLFVYFIIFGNLSIDGIFSWFTCNIAAHFRILRLRLKY
AGQENGGNITKTTVNGCIDLHRQTIELAEEFNKIFRVNVFIKFTISCLQIACLAFQLVRG
KEKVDQIFHFSFLTSVTLQFLLYCYGGQKIKDEVIMSNYGL
>GmmOR15 TMP_Or15-RA scf-651027 85495:88783 forward
MLLHTLQLKRYFWVTRRSFNLIGIDITALDYHDIVKYPMRCFLFTAFAAVLAWAMTIHVY
EYRDDFGEFADTSGMLLQTIIALWKTLVFLFKRKEICDLMQNVWQCNINVNPQEFHIILK
FNSQNFTISALYMVLVSSTVFGSFLVPLIYMFEFYRKNGEKIWLPPQKGGYFWDYSNAIG
YSVLYICHLLGIFFVAAFSIGVDTLCPWLVSNIVVQYHVMYYRLRDIAELSYEISTEKLN
AKIIECVKCHRQVLNLSNQLENFFAEIIFIKFVISGLLICSLAFRLVRAEGQFYILLYQL
VFLTTVSTQLLMYCYSGQRLKNESSQVASEIYSIFEWSHLSRNSQKLLLFAMMRSQRECH
LTGAFFMVDLSLFVWVRERERDVCMYVKQKKKKKKRTNIFNFWFCLKKKSYHFVLLLHVF
LCYLLKVLRTGGSLIAMLKTLDEKQV
>GmmOR16 TMP_Or16-RA scf-649095 5203:6553 reverse
MPDKKCAKFLSIQQRNLAVLGFDLNAAERRYLVEKPLKLLFVLSCNFYWTYGLMNFAIYN
IENFDEITGSLSVFNQDILLFFKMFIFFAKADKYLNLIKSMNKLADKAKNQEYDEWMSEN
RLAELIAKVYSYTCRVAVAFSATVPLIYSVYVTYTAGELKLKLPVKARFTLDGKLSYFAF
NYITFVIHLNSMASLTVGLDSLYFWFIYNISAHFRILRKKLEIMAVNLKTNLQYSIQDDL
GAIVVYHLEVIRFIKSMNEIFGEILWAEVTMSCLQMCFATHALMSDVDVSNAPFNIVVVF
AVIIQLAIYCFGGEKIREESISLCSDVYLLFPWHKMSMHQRRLMLLPLLRAQKEAYLKGL
FFQVDRNLFVFVSKRIRSVTLIPVGFA
>GmmOR17 GMOY005386 (TMP007492) scf-648564 107291:115929 forward
MSFRLGNVSFNLTRKIIVAYGLIETHLFKYHESNIPKICLNDLIRAITIHEMDYKFKKLE
PHNYIKNGSDELNHIIVWLMKAYSDNILNSVETGLKDKCLEMILVATMWKIKQSLMNHFP
SDPRKGQIGSIELNLWLAQMSGVSVMRLRQNSSDINIFIFLYTALITILVTFVYTACEFY
DLALNWYDLNILTQNSCISLTHLAGLAKIINALYKLKDIQKVLDKLKYGLRTYTISKEQR
QTFLDGELENKLLLSVYVAIVATTGLLGMFVLFLSIQVLLHPVKMAGKTFPYRAVVPSWV
PFPLQLLYMSLSVLIFAMQIVAIDYLNINLLNQLRYQLNILNLAFDKLSFEKEQKNPAPH
SRRLNAIIEHHCLLKEVCQDIESIFSVSILLQFFTSLIIFAMTGFQATVQSDASNEAILI
YFYCGCICCELFLYCCFGNAVTEQSKTLPIRSFNSSWYDYDGPYRKSLLVFLTNAQQPFV
FTAGRFMDLSLPSFLGILSKSYSYIALLRQLYGPLNLGAQAPGTVVFMYFLAAFPLSSIM
S
>GmmOR18 TMP_Or18-RA: scf-648792 266134:274630 forward
MAANLREELERSRLSHQKILKPLKVISLAVGVDIRQPTRFWNPVKILAIALIVNGFVGFI
ADYNFIRNNLNESINVYADSLLISMQIIISNIKLIHLILKQHEFYKVIKMAENSEILLNL
REFELTTNEHKNLMQKIKDILKESWQDIHRQLSFFIYSCCAIVGWYWIVALAQNIHDLYK
RSEHFTVTTTHPVKYPVWINKGPKFIIDPLQYINSAACNHISGFGAVCYDGIYVVLVVHC
AALVHVLRVLVEHSTTNEIPQARRVQYLLGCIRFYQKIFEFYSSIDGLFRIVNLIQYCIN
AIILCMIIFQASIGLEAEASLVVKMFLYLLAIGFQNIMYCYNGEKLITQSNSLPIAWYSC
CWYNESAEFKFLVRMMILRTNRALYMKMSDFSTMSLMTMLTVIISDKLLLFFTLEKRQCG
SMKAYAVQL
>GmmOR19 GMOY012322 (TMP008029) scf-648792 271039:283655 forward
MSNMLIEVLKNEQKCNNKTLKNLYKISFMTGVNIKYKTDFKDPVKLVNLFLITISLMGLC
AQYCLVWHNRHESFVESADAICTANQAWISIFKMIYLVFVQHKFYDLLHAATDGNLLYDL
GIFDLAINCKEQLLKEIKDILRDSWLDIRRQLNFFIFSCMMACGFYMFSCLFVNYYYMHI
NPQNFTLQLRAVCFDGIFIVVVVHCSALFQILHQLLQHATDEDIPRSERVKYLLCCVQLH
DRIYNYYAKINSMYKNPSLAQCLLSMLVLCVVMFMANVGLEEDITLFFKMLCFLMAAGFQ
IVIYCYNGQKIITQSGMSPSSWYNCSWYNESPQFKYIINMMVLRANRTLYLQVSGFTTMS
HMTLLSIVQTSGSYFLLLRNLSGMD
>GmmOR20-pseu scf-652157 - 1785914: 1788436 MW: 36658.492
MVLISEMSYVIMATMNTEHFVKVIVNLSYIRVVGVETKTRRLYKFYPKRNTNQKAYQVML
YPRNYRRTTMFYTFIHESLILTQFAYKNDKLDSSILKLRSVERGPTSFICTAKFSWTHWH
VATISQWFAFAPDSFTTDYAFSRIKENLIPDGHYERHKNTLKDIIRYPQSLLRLSPFTCI
TEWHHIVSVAKFSPRQTNDSHSITMFASLCVINLKSTVQSIIYRKINYSIGNFFQPNVFQ
TSKLIKLLLFPGCCLVQMFLIWASFKQFHLSADIRYKKMLLVLIEHFQKPPVLKATSLLD
FTFFLLMQFF
>GmmOR21 GMOY011399 (TMP013635) scf-652170 3479799:3481562 reverse
MASNDYHGGKANILYSIRDSERLDNLFILMIKYLKLTGQIPLDLSSYIPTFLSPIERWFS
RLYGFLVLFSVLHVALFLAKNTFDILETGELEQITDSLVLTMIYLFASFSSAYWMFRQKR
LMELFQQINQLHRHHSLAGVTFVSYRCSYNLAHKAVKYWNLWCIIGVVFWALAPLCMGSH
TLPLPCWYPFDALLPFIYELVYITQFWAQFNMGLLFGNGSALFVAIIIIVLGQFDVLYCS
LKNLDSHALLMSGQTLKAIKDSQSLQDDNEREVNQYLYSTEYLSDLENLSQQILKKGDKS
FKRRLHEALIECVHLHQFILKACDTLEELFNPFCLIKSLQVTFQLCLLVFVGVEGERSTV
RIVNFLQYLCLTLVEFFLFAYFGEMLRRHSVRVGDALWRSRWWLNAHLIKRDVFIFLINS

ERAVKITAGKLFTMDLDRLRSVVTQAFSFLTLLQNMAEKQENEVA
>GmmOr22-pseu scf-641538 58384:59452 MW: 35047.4
MTIKIMQSFLFPNMVLKTIGYEFAHKPTSKWLLCLKRVLTHFYCSSYSIKQMYDIARNGM
PNVPLFLRLTSSLMYTISGDVKLLNFVIHLKQSKAMFKRFQYIYPQSLEEPHSYQVNQHF
WSSWDYGILYFYLLSTFFIIVSPVTNSSIVISLSSAESWMVWRFMLRRSLYDGYGHGSRI
IDIYSLSTHSCRTILCGLLLWSATDYFGNCYNLLTTYYFKFFFQSKNIGNSAYNHRWYIA
SSSYKKSILIILIRAQKEADLNANSMQTISFEAFKILLYYRYWVRFIALLLYLELC
>GmmOR23 TMP_Or23-RA scf-652170 23283182:23286809 forward
MHLRTIEDVPLYQTSLRIMKFWSFLLQDNWRRYVCLIPYVIINTTQFLDIYHSEESIDAI
VRNAYIAVLFFNTILRAVLLCINRLDYEKFMDNIRLLYIDLMDSEDNVIRALVRDTTIAA
KYIAKINLLMGVCSCVGFVTYPIFATSRVLPFGMYVPGIDKHESPFYEIFFVLQVIITPM
GCCMYIPFTNIIVAFILFGILMCKVMQHKLKQLHNIEDEKAREVIIWCIKYQLGLINYVK
TINELTTYTYLVEFLAFGAMLCAMLFLLIIVGPNGDYRHLHIHDILTECYYSLFVAEAAY
DSDWFDFNVTTQKILHLLILRAQKPCAILVGNIYPMNLEMLQSLLNTTYSYFTLLKRLYG
>GmmOR24 GMOY010839 (TMP013060) scf-652157 888049:890329 reverse
MKLYKFEDFVRLANLFYTSLGIEPYALGQSTKWQIFCRYLIFSFQIINLSSMVVCEVTYV
FLAFRNDNNFLEATMIMSYIGFVLVGIFKMLSIWRQRSLLTTFVQELLRIFPQTPEQQRL
YNLDIYVRQSTRVTVCFSLLYMLLIWTYNLFAILQYVIYERWLMWRVVGKQLPYTMYILW
DWHDHWSYYPLYALECIAGFTSAAGQISCDLLLCAFATQLIMHYDYVSRSLAMYEVKMRQ
KFREPRKAMAEDMKFLTNIIAYHANVLSLSELMNEVFGVALFFNFMASSFVICFVGFQMT
MGADPDTLFKLFLFLFTSASQVYLISHYGQQLIDAQWYNADIGYKKMLVLIAARAQKPVE
LQATKFVLISRGTMTDVSVIKKYQKYFD
>GmmOR25 GMOY012357 (TMP008534) scf-649009 35111:37791 forward
MSADQSTSSTNFNSFVKGPKNMFKLLGIKLQKQGNLISTMYTSFVVISATIVAILQLCHI
LTSDTSDPHRVQNVVYITYFTVGLGKILNILYRRSILIECLSELEGIYPDKEMERKYSLQ
HYLKRYKRAEYFFWNFAMFLISVYNIETIVRSLLQLYLKGRYGYLLLVRIYAPFPYDGKL
LVYLCYFALGSITGAWIAIIIAAADLYLLGCVLQLCLHFELLTKQLMELDTRILTESEAI
KRLNAIAKYQSELIRLSRKVNRTFSGSMIVSLTAASFIICFLLFQLLGDVQLVDIAMVLI
LLLNESKQVLLICYCGDKLLNSSQRFNESLYMHNWVDGSVRYQHHLLFMLLYARNPIMLN
LLDITDITLVTLKEVLVVKTLGKRI
>GmmOR26 TMP_Or26-RA scf-652170 7565092:7567386 forward
MLTVREVTSFPFRFYNFVGIKLFQWDDNDTLTKREKYTLLVTLIIFVMNFVCKFSCFVLR
KYEDAQELTKLISYFGFACNGVFKMLSVWIGRKTLHAVIKDLAKNFPRTASECHEYKFYE
QYAFLKRHMYLLSLIHWSIAITFMLFPIVQSIFEYLVNFNDNGKFIYRFPYIMIYPFDHH
TPAGFIFAYITQLIGGITIHSYFCGSDCLLLATVHLVNMQFVSLAVRIKKFKPQTYEKDL
KQLRKILKLHISAHQNAKLVNDVFSISIFLNYLISIAVLVMIGVQVISGSEFWEFSKFVG
FLIASASQVYYVCLYGSLLLDYEWYFADNRYQRMVILAIARSQRPAHLTAYKFFMISMES
FANVRLNHLTRNCFVLMTTSYQFFTLLKARMEEQN
>GmmOR27 GMOY008038 (TMP010200) scf-650866 8500:9930 reverse
MDEEVIVTIKEFFDIPLKFCLTLGIRLYKWSETERTTILQVCLLSNLLLHTSVYPFFLVI
YQIKIDPNDLLGRTSSLAISLFCVNAVSKILFVARHYKDLRKIIHKLIKYFPTTSGGQKN
FNLHYEFKTMRRVSSIMLWTHLITAVLFDFTPPITFGIEYMNSGGKKEFNFILPYGIWYP
WDHEDSAIMFVFTYMTQLLGSYVAVVSFVVPDLLLISVVALANMNFRYISKLIREFHPTG
TNEDFKSLGQILHYHDDILNMIDEVNDVFSFSVLLSFFGFGGLLCLVAFNAVVGSSMLDI
FSQTLFILSILLIMYYLSAYGTEMIRLVNGNGNEIEMKFKTRQKQTIFQSTDVSVALTDH
PWYDGSIHYQRMLPFPIARAQRPAQLLGYKFFVVSMESFQSLMTTSYQLFTLIKARYDED
N
>GmmOr28-pseu scf-650866 82839: 84829 MW: 29955.6
MNGLNYSNVPLNIIEMSHLTKTQRVSTISLAFSITVIYLMVYSIPGITMNSVCILLHFWL
KCVCMGGGEVGYLAVISFVNPDLLFISIVALAYMNFKHVSKLIQELRPEGTDEDFISLST
IHARYFNLSSVYIVQSIDNVLLVCAYGTEMIRLVNGNEIEMKFKTRQKQTIFQSTDVSVA
LTDYPWYDGSIYYQRMLPFPIARAQRPAQLLGYKFFVVSMECFQSVSRYRHRLRFRYLFI
YSLFIIPVATDNVLSIIRFD
>GmmOR29 TMP_Or29-RA scf-652141 245471:248227 reverse
MVRNSGAKPMVVLDFFKVPFLILQKFGVCVYRTSSKERLSVKEAISFAISMVQIGAYSIL
VPLFFIKTTPPTTAEFSDACIPLILFFTSIVRLTSILVNPKRIRTLVDIFQKYFPQNMEE
QKNFKVERNYKELIRVTKALAICLSLGCVLFSLAPLFNFAMAYYTIGDEAKFDYRMPYPI
WYPFKVNTPGMYAIMCSTQAFAAFSCVCAYFLPNMVLITSMMLINMNFKHLAKTVRNLTP
TNTDDDMKNLSKILGHHQDTFLLVDVTNEIFNISVFISFFSSIALLCSIGMNVLGESQPY
HIIKQSLLLVTSLFDLYYTCKYADDMKTSVRNIWLNSVYYRVLTTLLVHFAARGANEVRIR
EHATRNYISKPKTKYSLDMSDALAEHPWYDGSVYYQRMLLFPMARAQRPAQLMAYKFFAV
SMESVLSVNTNIAYLTSI
>GmmOR30 TMP_Or30-RA scf-652141 215509:218352 reverse
MVRNSGAQVMVVLDFFEGPFWMFEKFGVCLYRTSSKERLSLKEAISFAISMVQIGAYSIL
VPLFFIKTTPPTTAEFSDACIPLILFFTSIVRLTSILVNPKRIRTLVDIFQKYFPQNMEE
QKNFKVERNYKELIRVTKALAICLSLGCVLFSLAPLFNFAMAYYTIDDEAKFDYRMPYPI
WYPFKVNTPGMYAIMCSTQAFAAFSCVCAYFLPNMVLITSMLLINMNFNLVDVTNEIFNI

SVFISFFSSIALLCSIGMNVLGESQPYHIIKQSLLLVSSLCDLYYTCKYADDMKTSVRNI
WLNSVYRVLTTLLVHFAARGANEVRIREHATRNYISKPKTKYVLTNSYQLLTLVKARFDV
E
>GmmOR31 TMP_Or31-RA scf-650238 26161:28277 reverse
MIPQFLRDDYPLEKHLFLIPKFALRLIGFYPESKLNTPLLCWAFFNFLLLGYGSYAEFTY
GIHYLTIDMQTALDALCPVLSSIMSFIKIFFIWWHRSEYKHLIEEVRRLTAAQSSRKNVH
IKRKFFTIATRLTALVLFFGFNTSTAYTVRLVISNTILYLNQQPMPIYPLTCILSHWHGY
ITVAGFVGADGFFLGFCFYFATLFKMLQQDLSDALAVNNCKSVNATMYLALLDKNSQAIR
CEADMVSNLTDIIRRHNEIAQLMKKFSSIMVGIVLSHFITSSLIIGTSFSGYAILVYIVH
TCAVIAEISLYCLGGTAVMECVIPSSTKSSPGIMPFQNTETTNLTHTNFQNSSSSKRLAN
EELALQAYCSQWYDYSVRIQKMILLIMVRSQQTITVKVPFLTPALPMLAAVCIVIVYPAI
LRFAGSVITLFKTTI
>GmmOR32 GMOY005084 (TMP007180) scf-648410 219323:223133 reverse
MFNPVPANDKRFHWPRQCVWLKLNGSWPLKSTKEFQSEFYTTENYSSFLYILWSWYVILS
VGVTIVYQTAFLITNFGDIMMTTENCCTTFMGALNFVRLLHMRLNQNQFRRVIEQFVNNI
WINRREHHPQVAHECENRMSTFRIMTILLSCLIAMYCLLPLIILFVDVGWNAPEKPFPYK
MLFPYDAHNGWRYAFTYVFTSYAGICVVTTLFAEDSIFGFFITYTCGKFRILHQRIDNII
SDSIEVTSTRQNDNNVQRIFEKKLNEIAYDHNKLIERFSNRLEHFFNPILLVNFLISSVL
ICMVGFQLVTGQEMFIGDYVKFIVYISSSLSQLYILCWNGDDLIQHVSRSTETAKHLYGC
NWEGTTLNIRNAKKFPKWHRAEEFHLTSIHPTNKEFRQNLQIMILCSQRPVKITALKFSI
LSLQSFTAVILSTSMSYFTLLQTVYNADQEEDHNIN
>GmmOR33 GMOY005479 (TMP007587) scf-648614 40250:49045 forward
MFEEIPLIYMNVKILKFWSFLFEKNWRRYLCLPATTFLVFTQFVYMFKTKEGIDSIIRNS
YMLVLWFNTILRAYIMIYDNDKYQQVIHELVELFYDLKKSKDDYVKDLLLEVNSKGIIMA
RGNLFLGLLTCIGFGVYPSLAVDRVLPFGSIIPGINEYSSPFYELWYIFEMCITPIGCCM
YIPYTSLIVSFILFGVVMSKTLQHRLRTLHRIANQPKLIHAEIISCIKYQKRIINFIEAV
NGLCTFIFLLEFLAFGALLCALLFLLIIVNSSGQAIIVCAYIIMILAQISALYWYANELR
QENFAIAAAAYETDWFTYDTIVQKDILLVILRAQKPCSVFIMFFKSYNEILFL
>GmmOR34 GMOY011902 (TMP014141) scf-652170 25006662:25007967 forward
MTIKIVQSLNQLKQFDDYIWFPNAVLKTIGYEFAHKPTSKWLLCLKRALTLLEISAHFYC
SSYSIKQMYDIARNGMPNVPLFLRLTSSLMYTISGDVKLLNFVIHLKQSKTMFKRFQDIY
PQYFEEPHSYRVNQHFWPSWVNVILYFYLLSTFFIILLTVGYAEAHFFYLKCDELPYSFD
HPMKLLLSLLFIAFYVWRFMLPRSLYDGYGHGSRIIDIYSLSTHSCRTILCGLLLWSATD
YFGNCYNLLTTYYVKFFFQSKNIGNSAYNHRWYIASSSYKKSILIILIRAQKEADLNANS
MQAISFEAFKIVLGAVYRAFAVFRTMLK
>GmmOR35 TMP_Or35-RA scf-648722 141939:145983 reverse
MSSHSLKLEDNPMLEINVKVWKYLSVIFPDREHAWRVYVFVLPVCVMNIMQFVYLLRMWG
DLAPFILNTFFAAAIFDALLRTCLVIINRDKFEAFLLELAEMYRDIEESKDDTYGRELLR
AATATVRKISIFNLTASFFDIIGALIYPLLCEGRVHPFGVAIPGVDMTASPIYEIFYILQ
FPTPIILTTMYMPFVSLFASFAIFAKTALKVLQHRMNSIFLYNYRTEEHQFAALTACITY
YSRIARVPFIWGDYMFLAILHEYCKCSCCYCCLINFLILTKLYFKIDSSTQIVSIVMYIL
TMLYVLFTYYWHANEVIMEVGHRLMFSLNRKRTTFYCCRVLKCLRQLMGYRGIIVVNAFI
MVGNVYPMTLATFQSLLNTSYTYFTMLRGLYG
>GmmOR36 TMP_Or36-RA scf-648722 191876:198790 reverse
MYYRLPLFSVNVKGWLYFGYIGKRNQGIKSLLIVNALLTLAAEILLYIKTEDVSNVIRDI
FKTAILFNSLVRILYVMRREEDFIQFMKGIESWYQEFKDDNDHMALAILNRLPKYTKLVT
AFGLSFGSIGAVASTITALLWHTHIFPIYVPGFDAFQSPLYEIINLWQSITMVSFVMSAY
ILFTNLFISWLIFGIGLLEILCKKFEQMSSANDAERLRNLKYLIRYHKRIIRYGEELEDL
VALISMVELILFTVMLCVLLVFFLITENFVDQIATVIYIFSIFYVLFISYWHTNAFSAES
LKIPDAAYRIDWAESGPETRKCLLILISRSQTALQISESNRYT
>GmmOR37 TMP_Or37-RA scf-648373 4213:5709 reverse
MNLLYRPKLSDGKLVRLSWPIAMFRLTHIVCWPLEDDAPRWAYVFDRFCWFLAFIVFVLT
NDAELRYLRVNMQNLDELLNGVPTYLVLIEAHIRGFTLGYRKNKFKNLLRKFYTDIYVDE
RQHPSFYKKIQPRFWPLYVFSSMYVATLTNFIVTPLALLLTRGSRELTFKMIPLFDYRYF
PIYLPCLLSNIWVGFLVVSLFSAEPNILGLVVLHLHSRYLIMNENLRKKTENLLKNPSNT
EIARRFRKIVVETIDENKRLNLFAQEIQNEFSFRIFILFSFAAMCLCAVASKVYMVSISI
GRKSLSLSNSIFFFHFNDVSQNPLGSFAYIFWMFGKIQELMIIGDLGSTIIATTDEVSTM
YYNSNWESVIARSSDSCENVRLMKLLTMAIALNRKPFYLTGLNFFTVSLNTVIKILQGAG
SYFTCLISFR
>GmmOR38 TMP_Or38-RA scf-648495 88188:91340 forward
MIGILPREWSEEHVYYSLAHTLLIAMGSMFIVTVACDLYEARQNLTLLGDDIVVIIGGSL
IMLKIYFHGQHSSEIHPIVDKMNDLHKMFAEYNGRSRLTIKRLQCSFYLFEVFAFSFYIF
LIVLFASAVMVPPLLTHHGLPYRAHFPILRWENSDQHPIGFVIAYIFQVIWTFHALLSIV
CMDSLACGIFLQTALNLKILCIRLREISAQNVEERESLNELKSLIKIHQYIINLIEKINA
CYYLNYVAQMGASTFMICLTAFEALLAQDRPMIAIKFEIYMLSAFLQLLYWCCAGNLVFF
ESLNVADAAYEIPLWYTRSKEFKMTVAFLIRRAQKPLQFRPSPLYGFNIKTFTSILSTSY
SYFALLCTMNK

>GmmOR39 GMOY004392 (TMP006475) scf-648080 333109:334447 forward
MHFYKIDEFFLILHPLQKYTRWIYFDFRRRDSSVPFEFQSMTKLKISLAFVLFELICNFI
KFAIEIRADRLSEAKQIAAVTSIALLCMIRGISLYTDRNRMLAICNDLDKIFPNTFYLQR
RMRVQKLAAFFKVRFRVLRSYLYLGLPAFASIPLIRYFLFYDRENGGRLLDEYHQHASWA
PFQLKQNNRAYPYVYVYETFLTILGFTCILTWDHIFTVTVSQLTMHFEFVNTELESLNVR
DTTRSMTSKFYWRRLKEIIVYHQHVYRLAKKLNETFNLTIFLTDIGCAGSICFHLYLIAN
SDSVLTIVTFFFPCFILIAFTFDYCQQGSRLAEASARLQTVLYNQEWYDASPTYRRLMLS
LLQYAHKPFTLNGFKLFDLDMLHFQSIMTIAYRLFAFVQTQGK
>GmmOR40 GMOY012356 TMP_Or40-RA scf-649009 40802:42730 forward
MVEIKDLQIRPEVCQNPLIVIHLKSMLLYGFIVSTEQKHKRFSLRRGAVFTISFVISCAL
IVSRGFESLAAGATSCATAFLYLSTSITIVNAFFQRARVVRMCTFLHDDINKLMELADER
EKKMFAATVKYLRYVTAILWTPSVFAGFIAYADCFYRTIFMPETVFNIPQVLRGEAEPIL
LFQLFPFGEVYDNFLVGYLGACYALFLGITIIPCWHTFITCLMKYIVIKFQIINKRLEAM
DISKLSPKFSLQPDMVDNLNEKDLNYWRLKMCEFCVQEQTKIKW
>GmmOR41 GMOY006480 (TMP008603) scf-649048 4425:5787 reverse
MPCLDRFRKFIGAASLIARFCACDMFDETYTMFNPLFMVLLGFLAIYTVCFVYTIYDGFV
TNQDWTIILQCLTIGAMALQAISKYACYYIHRTTLRKVVKYLDEIYGEYQDRGARYEYAL
KKSLQKIMKCSKITGIIYFTSLAGLIVSPFIYMLITGKRELTLIALIPGIDPKETVGYFI
HTGFHSAIISFAGVGFFASEIVMFSIFLHVLLHRDILMEKFHDLNEATAKPDVPAARTRD
LLNDIIAWHQKYNLFVKYIEELFSGVIFVHVTTSCASICSTLFCIVLKVWPFAPVLLLLN
TSMYAYCGFGQLLATTNDDVTRMIYEVCIWYRLEVKEQKMILLMLRKSQYAVELTVGKIM
PLNFNTALQLSKGIYTYLMVLINFLE
>GmmOR42 GMOY006479 (TMP008602) scf-649048 15:1464 reverse
MPPLDRFRKLIKHLRVVVRFCACDMFDETYTMFNPFFMVLLLCLAIYTVCFINTIYDGFV
INNDWTIILQCLAVGAMAVQAISKYLCYYSHRITLRKMVKHLDNIYEEYQSRGTRYEYAL
KKSLQKIVKCLKVAGGIYFTTVIGLIVLPIIYMLITGKRELSLTALIPGIDPKETVGYFI
HTGFHSAIILFSGVGFFASDTIMFSMLLHVLLHRDIMVAKFYDLNEITEQDDKSAERSRP
SLNDIIAWHQKYNLFMEYIEELFSGIIFVHITTSCVSICSTLFCIVLKVWPFAPLLLISN
TSMYAYCGFGQLLATTNDDVTRMIYEVCIWYRLEVKEQKMILLMLRKSQYAVELTVGKIM
PLNFNTALQLSKGIYTYLMVLINFLE
>GmmOR43 TMP_Or43-RA scf-651490 579:1910 forward
MSPLDHFRKLLKHMRVMVRFCGCDMFDETYNILNPVFMVVLLCLAIYTACFINTIYDGFV
INNDWTIVLQCLAVGSLAVQVIIIAKIYIYIYIYIYTYNIYMIKMVKHLDTIYDEYQSRG
TQYEYALKKSLQKIVKCSKITGIIYFTSLAGLIVSPIIYMLITGKRELTLTALIPGIDPK
ETVGYFIHTGFHCAIILFSGVGFFASDTIMFSMLLHVLLHRDIMVAKFHDLNEITEQEDK
PAERSRPLLNDIIAWHQKYNLRTYAGGYENGRSQVFPCRFVEYIEELFSGVIFVHVTTSC
ASICSTLFCIVLKVWPFAPVLLLLNTSMYGYCGFGQLLVTANEDVTRMIYEVCIWYRLKV
KEQKMILLMLRKSQNAAKLTVGKIMPLNF
>GmmOR44 GMOY006265 (TMP008385) scf-648928 37939:39536 reverse
MEHELTCIQRFQKIIKVLRICAKMCGCDILNPDYKMNFITWLLIVGVNGFFLCTIYTIYK
GMTIDHDWTVIPVCMCIIGSGIQGLAKILLVLKYRKVIVEQQYFLENVYTVYQQKTERYR
HVLNRWLEYTVKTYRICASMYVFVGTIIIGFPYIYWYIYGIRTMIMQFEIPFVDPNTHTG
YIIHTLYHFPMIAWGCLGHFMTDIYMFMFIINVPLLKDLLEQKFLDLNEILEETNESEKV
LPLLKDIFQWHLKYNEFIAGVDKIYRTIIFIEITTCGLSICCTIFSIVLDVWPAAYGYII
YLIFCLYSFCIMGTLIEISNDNVIEIIYTISHWYKLKLPEQKLILIMLIKSQKPIELTVG
QILPLSVSTALQVTKVVYSFLMLLLNLLDK
>GmmOR45 GMOY007896 (TMP010054) scf-650705 207676:209075 forward
MKLYSERYQEILNVSVTLLKLCGINIFAKNFRMNLITWFIIGIIGLYFILFTYTVYVGIQ
NDWSIVLKVITMSTTATQGLIKLQNFLCNPKTLRKLAEELEEVYLEYERRTNQKYPKYLQ
RSIALIKRINYAVLAFCLIALSTVIVWPLIFSYVKGEKLLVGSMRIPCVDDKNGWGFIIH
FALQSVMLIIGAYGNFAADSYCFLLLAHTSLFKDLLYCKFQDLNEILQQYPRNSLRSKPL
LQDILKWHQKHVLFIETASSIYFWVILVQISTAVTGIVTTLVCQFLGIYPTAVGYLIYCA
VLFYIYCGIGNLYERVNEEVINIICDSLWYELTIDEQKLILFMLHKSQTVEGLTIGNLIP
LSLNTALKLTRFMYTLSMMLLQFLD
>GmmOR46 GMOY003305 (TMP005362) scf-645804 155235:156578 reverse
MIVACVDHFRKLVRIFRFCAGICGADFGDRNYNPWNPLVFVTLSSIVFFTVCTVYTIYAG
LVYDNDWTVILQCLCLASSAFQGYPDKNRECSDDRYEARRFPLHYNSYIACAISAPLSIN
FGKRLFIMRFFIPGLDPNSTMGYFLHMLIQFMCIFFGAFGNFSGDMFFIVLVLHVPLLKD
IITEKLNELNRTQLKKRNKLLKDTFEWHQRCNLFIEDISELYSEIIFIEIFLCCLGICCT
TFSIVLSLSAWPAAPAYLLFTLVAMYVYCVLGYTIEEASDTLLRNIYTECLWYELAVQQQ
QMILLMLIKAQSPVVLNIGKVMPLSLSSALQLTKAIYSFLMMLLNFLE
>GmmGR1 GMOY007472 (TMP009619) scf-650411 214655:216066 MW: 48775.35
MAFWAAATGRTSPSKVVPVLNPNQKQFLQDEIRYREKMNILARTDAGNLTDYYVRKQETI
DDPELLDKHDSFYHTTKSLLVLFQIMGVMPIHRSPPKTNLPRTGYSWTSKQVMWACVVFA
IQTTLVVLVLRERVNKFVTNSDKRFDEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKN
MWTNYQLKFLKVTGTPIVFPNLYPLTWGLCIFSWLLSIVINLSQYFLQPDFELWYTLAYY
PIIAMLNCFCSLWYINCNAFGTASRALSDSLQTTLKGDKPAQKLTEYRHLWVDLSHMMQQ

LGRAYSNMYGMYCLVVFFTTIVATYGSFSEIIDHGATYKEVGLFVIVLYCMTLLYIICNE
AHYASQSVGLDFQTKLLNVNLTAVDSATQKEVEMFLVAIAKNPPIMNLDGYANINRELIT
SVGIT
>GmmGR2 GMOY011510 (TMP013746) scf-652170 18996209:18997720 MW: 49157.684
MDIITEKEFDSTLDSSVLTDDMEKLLKSEKRKKFIKNSNHNPDLMERLTMRAKRAKLKSR
HGQTAELYDQFYRDHKLLLKLFVILGVMPLARQAGRVTFSWKSYQMYYALVFYILTTLVV
LKVGSERIKLLHTTNEYDEYIYAIIFILFLLPHFWIPYVGWGVANQVAHYKSMWGTFQVR
YYRVTGTTLVFPTLKIAIIILFIGCVLLSILFLFSLSQLLDDFLLWHTTAYFHIITMINM
NAALFYINSKGTKIASKSLSDCFKKDMRIECNYIMIKQYRLLWLNLSEILQALGNAYART
YCTYCLFMSINITIGSYGALSEILEFGFSYKVLGLVVLTIYCASLLFIICDCAQSATSVM
AEGVQSTLLSLNLLQLDVQTQKEIDHFISAIELNPPVVNLRGYALINRELLTAVNFYLHY
IHAFI
>GmmGR3 TMP-Gr5 scf-652170 18992721:18996270 MW: 59057.336
MPSSINIYLPLDNIFKMYSVLEEKVVAENINFDLLRSDLKVNLKVDNINKHKINSMDLIV
QEKEYERFERSRLNSSDGDAVEIHDQFYRDHKLLLILFRVLAVMPILRSEPGRITFSWRS
IATVYAIIFWFFMTVVVIIVGKERIHILQTTTQFDEYIYAIIFVIYLVPHFWIPFVGWGV
ATEVAKYKTSWGAFQLRYFRVTGTSLQFPRLKSLIVVISVGCLLCAILFLLCLSFLLEGY
PLWHTLAYYHIITMINMNCALWYINSRAIKTASYSMALCFRKEVMKDCSASIISKYRFLW
LNLSELLQSLGNAYARTYSTLCIFIFVNIVIAVYGAFAEIVDHKHGYTISYKEAGLIVDV
VYCCTLLFIFCDCSHNATLGVAKGIQKTLLSINLLKADLEAKNEIDLFIVAIEMNPAIVS
LKGYVTVNRELLSGSISMITVYLLVLIQFKFSLDKSTTNDKRNGFVLTFLETNDQLMLSF
VSLGNNISKLMKNTYGYYNGKRIRLDSGFVCVNR
>GmmGR4 GMOY008001 (TMP010160) scf-650833 228463:237409 MW: 55967.164
MLNNYNRKKKHDTVFLNVKPAANGIGVRKYSNGILDRIDSGFGINYDNEKRASRHSVGTI
DSMNQQFVPNIFFRNLAPVKWFLSVIGVLPMTRPGPGSAKFTIGSLAFVYSILTFTLLTL
YVVYVANNRIRIVTSLSGPFEEAVIAYLFLVNILPIFLIPLLWWETRKICVLLNDWDDFE
ILYYQISGHSVPLHLRRTAVTVTVALPLLSVISVIITHVTMADFQLSQVIPYCILDNLTA
MLGAWWYLICEALSRTANILAERFQKALRHIGPAAMVADYRALWLRLSKLTRDTGTATCY
TFTFLNVYLFFIITLSIYGLMSQLSEGFGIKDIGLAITAVWNVCLLFFICDQAHYASFNV
RTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSNINCGGFFNVNRNLFKGLLTTMVT
YLVVLLQFQISIPTEIHNTNVNMTLDDDDYDDDLLTNAIVTTTTTIASIKGRKDPGGGHG
IATVAISRSLFMVRKL
>GmmGR5 GMOY004207 (TMP006286) scf-647997 53981:55829 MW: 53912.25
MSQQHQQSCSYVLLHFSLIFYLCKFLGIYPQNLSAYRNKHSLEASKNGSFLVVIVMIWII
VCYNTFILSQSEGKDHREANHNALNFVIGIFLTQIKIIVMATDQLSSICNQGRLAEFYNR
VYSIDNRLIKEGCLLNNGSLALHCYVMVVLTFLFEIIIILLTFVNLINYTQWVLVMWIFS
CLPTLYDSLDKIWFTTSLHALRIRLSALNKTLTNLKGVKEKSIFKRFYEEECNELSLLKI
DYVNGLEYLKRELNAGRVVKFKNRIRSFDLNIFEQKSFAKDFINVEKVKERLTNICQLHD
EICELVQTLHKMWTCPILALMAYGFLIFTAQLYFMYCVTQKQPIPILFRSAESLLISILF
LSYTGVKCIYPIFLSWKTVLEAKCTGIALHEYAINLNDKAVYEIVSVNHLSLKLLNHAFD
FANCGFLSLSMRTLCGLSGVITSYTIILLQFNFAAQQSKEADLMGME
>GmmGR6 GMOY011615 (TMP013853) scf-652170 14942509:14944789 MW: 50882.254
MTFWRDFIKPSDSYAAEQTLLWTTFLMGLTPIRISSSERNGERNIYISRAGYIISIVQVS
FFLYSFIYSFLMDESIVGYFFKTEISQVGDVLQKFIGMSGMLILFGVSLYKTNDLMELYH
VVAKIDNQFLGVGVEFNYRFIMKFRHSKLMFIISICSIYIGGSFWMLFHNNVWPTFQAIV
AFFVPHIFLLSVVTLNVAFIMRFTQHYDLINKVLKNLRHQWETRSIKTISHKQKSLQCLD
SFSMYTMVSKNPCDVVQESMEIHQLICEAASTANKYFTYQLLTIISIAFLIIVFDAYYVL
ETLLGKSKRETKFKTIEFVTFFSCQMMLYLVAIISIVEGSNRAIKKSEKTASIVHALLNK
AKTPDVKEKLQQFSMQLLHLKINFTAAGLFNIDRSLYFTISGALTTYLIILLQFTTNSSP
PSSSDCTFNSPVTSPIPMNSTST
>GmmGR7 GMOY006209 (TMP008329) scf-648889 6876:8857 MW: 46324.344
MDVENQNNLPLARPAYRNRWRSLFSAKGLYESLQPLFLLLYWHGLMVYSIKANADGKKEL
RHSLWSHLNVFFHLTLYIVCYIMTVMNNFESVAGYFFQSKISRFGDFMQILSGFIGVTVI
YITAIIPKHYVQNCLQFIQEIDQCLLRVGVRIMYSKILRYSCIFIVAMILIDILYTAASF
QILKSANEEPSLYLHITFILQHTVVLTAIAMFSCFTKLIEARFTMLHKKSLQCFDSFSTQ
TVASKSPCEIVQESMEIHQLICEASSMANKYFSCQLLTIISTAFLIIVFDAYYVLETLLG
KSRRESKFKTIEFVTFFSCQMILYVIAIISVVEGSNRAIKKSEKTGSIVHALLNRAKTLD
VKEKLQQFSMQLLHVKINFTAAGLFNIDRSLYFTVMICETIH
>GmmGR8 TMP-Gr4 scf-651593 207733:209113 MW: 47166.812
MYIKVRSLKSQAVRKWSGKVKHQSIRQENDVILNDLRIIMFPLKIISLIPLYGSASSYEL
GPPTKILYTIVMRFLILVIPVYNLYNLYLINSTETSNEPEWIDSSLNFLNYIITISFCTR
HYQLLEKIINGLLKFYEEMKEYGKCVDVTEARFSITFTIIMLTLQCCVVILKIVAIDVPL
AFNISIESVLYAFQNIVPDLYIIFISSLMRTIATQFSNLNNIATASTCGIYEGTIKNQIS
SNTLVLGFLLQMYRKLLRIHKFISDYCTFILLFYVGYAFFSITSKTYYIFVWIISPQEIS
QSEVSLSFIILVMHMLLLILLCRCFNLVARKVFQTTNVKYVKIPRSKFPKGYMPYQGWIR
VLVVLTTLSCTPVPNWNFKICNRTLIFIYSRTLKQFGAEHPKRTGSL
>GmmGR9 GMOY011903 (TMP014142) scf-652170 23328288:23329768 MW: 39397.14

MDIVESISVLQTVYQLANLSPWILDKKTWSFLQSRVLELYTTIVIIGSSALMLYSLFTDD
ILVKASDNEIGQTVDSVQMVGIRMAHIASVLESLIRKNDQKKFYDDVKEIDRIFESSLGI
TINNRILIIKERLNELVAALDEVNLTAEPSTPMRTRMAQNHSGHVFSPKRQQESNIENSP
LTKLVVIRNIYDRLHSQSVVINHCFGVSMLINVGNDFISITSNCYWIFINFKDFSSTVID
FLQIACSAIWSVPHLLNVLVLAFICERTVHTSVDMALILHRINSNFNNDKYGSAVTQFSL
QLLHQKLSLTAAGFFTIDSTLLYTIVGATTTYLIILIQFHLNEMKPNT
>GmmGR10 GMOY003231 (TMP005286) scf-645661 235649:237209 rev
MEFSNCDLLYDHISRLTLPSTVYHELMFASVMDKGLLILSWIYGVSPWNISRRSYCCNIL
DYGQMLMMLFFYLTCYALVNFDFFGIYQAPDCNGFCRLGNQLFLHLGCLLYLTTQLLAMK
SRKKFCKKFEQTLINMDRSLETCQLDERGEMSMKRNQSLHFLYKTAMTLGLLMLLIYEVR
QLTYFYGHWFFVPLMVTTYPYSAATVMLLQFAYYVYKISDRHQIINQFLEQINQDIVKTP
KEITPEIFNVESEIKITSSGNVQLNERHVKQTTHRREENDYNEDSMKFLESSTKRYATTT
SLTKLFKMFDSVLSLSVLTNSEYGPHCVPYMSACFVITIFCIFFQIKIYHVVGGKSRLLD
YIVFIYFIWSFVTMLVTYIVLRLCCNASGEAKQTAVIVQEIMRKRPAFMFGVDGNYNEMK
SFLLQSLHWEEYFQYNAIGLFSLDYTLIFTVRKKRSRK
>GmmGR11 TMP-Gr3 scf-652146 202807:205545 MW: 52777.453
MLQRKRNKKTFVCRATNKITDFSNESNQNRMKSLELQFHRALRPLLIISEILVCAPVGLQ
NLHNQSNYCTFCRKCLQIVWGLIAYVAVAHGSYNEYFFLSAYLPTQELPFYLSEFAFYLL
HVLLIMLASYFGRKTFTFSLTFILDFDSKLLKHFKIRMHYIDLRKFLRNHLSLNFAFFLS
AVMLGFIQRRSSLLGILTVNTSYTLPNVITQTSLIQYYALVYVVNKRMQLLHQLIEQTFQ
QSLKGNIFNVQQRLRVLRGLYADMDAYTKHLNDAFAVPLLLFFMASLKSLSFYIFTLYKW
IDQWTDLSYTVISYAIIEICWQFSRAFLILHFNQAIQNQRKQAAILFTSFSSVAERLEPT
IVLIISDQPFYCAIVYGSSELYYLWNHTVKYACSDDGKRYILGKETMCVCVLQFQNNFPF
TDVLSSLSIIHIFGAIRYHLRSFNECRSKT
>GmmGR12 TMP-Gr2 scf-650947 5268:6453 MW: 44579.047
MFEKKRKNKNFFSKATNNKILDFSKKPNQKRTNSLESKFYKALAPLLMISIILSCAPIGL
QKSIKRINHYYVYQKFLQILWGLTVYISIVLGAYHEYFLWTGVLPVVQLPFYLSEIVLYL
LHVLQIMIVSYFGRNIFIYLLESIVDFDDKLKQLKIHLHYNDLRKFLRNHLLLNFAFFLC
ATTLGYIQRSTTFLGFLSINSSYTLPNVIIQTSLIQYYAMTYVINKRTQLLYQLIDQVLK
QSPTKNVFNVQQKLRFLRRLYSDMDEYTKRLNETFSISILLYFMSSVNGISFYIFCVYKY
AIEWAETEFIWMAYAIVEICVQFARVFLLLHYNQGVQNQKKRATVLFTSFRHNNERLEPT
VSTFQFFSFANIFLN
>GmmGR13 TMP-Gr1 scf-640662 25103:26585 MW: 53170.637
MIKLVKIYFNINRAVGLFNLHYDQKKGQFISHHKPTILYCALVDTAILTLLPFITIAFLK
SNYYCESIKSIMINTSLLIITDHTTIIVIMLTTWWKRQAICRWCNAFAKMLKAYLGVSDT
YEYYDIVQRTRKQLLQKLGLSSFTIVLLYIYLYGTGKGNFCPTTNDISYVLKTLFCGLTE
LIIVLIDYNLFLGLTLMHMILQILSKKLKEISHDIQLMQRMQEIRYEMRESSKTFNPHMW
LKQLQGDVDAVAAESAKLKKMIYELVRILQVPYLCLLLNAFQSLTAIMFHILRFTQQRHL
DIITVIFYLIVLSANITNLMKSFQVCEYLCEDFKIFNEKVYALILNDNLRQAVNDNGADN
SLVLSRLFLENSSDNDSRKKDEIWFSFLTKYEKNGNLSHIPIINTILQVKLIVLETFSGK
VRNIIAIDHVKNELSTCVCMQDETMNDEFTKAKDMKI
>GmmGR14 TMP-Gr6 scf-652170 14938990:14942365 MW: 35515.863
MSTKFQRLRQHFISHQVFEALQPLFFITFLYGLTPFRVTTNKGGEKSIKMSAFGFINIAL
YILLYGSCYIISLLREESVAGYFFRNKISDVGNTMQICNGLITGTVIYISAVTQRSKMLR
VIATLHNLDKNFANIGIKVKYSRIYRFSIVMLTFKSLVIGVYFVGVYRLLMSMNITPAFT
VCVTFFLQHSVLFLAICLFCFVARSFERRLVILNKVSIGELVLCIVAELIDVLIPRLFST
MPTDTWPIREMLPSFPELITVCFSHFRRRASSEDNVLMTYIKPLQSMSFVQFFESNNLSS
FPLKINLKL

Dataset S2: Annotated amino acid sequences of *G. m. morsitans* glutamate-gated receptors

>GmmIR8a GMOY012127 (TMP005604) scf-646432 9948: 18262 MW: 110055.02
MQTMDIETILSHIHGNYNVQRIRSLHHRNILNRCRRSNWTFVHRLLLKQTRSIWIDRVQS
DFYNDLSNALKDIEDLHLETEISDSILEIEQEYETVSNGVCEILASEGASVVIDFSYFVW
QPGLDYIRFHHIPYLRVDRLLRPFLQIFAEFLKHKDANDVVIILQNERDEMEALLQITEG
YPFRTLLLNGGNGMDFLKRLRDLRPRPSYYGIFAGGTNMNSIFEKINKAKLFARPPEWHF
VFLDPKDRVFKYKKVAENAAKFAVNARSVCKSLRMKDAYCLSGFNLQRALLLDIFRSLIA
VRQSNLDWLEPIQAECNETEHSDNSELEIRRNFDILDHFPLTNFMTLMTPDNPMQFEDEY
DQIIPSLTFMVNISINFYSTAHEAVTDLAAWQNGELRKINQTISPAKRFFRIGTAEAIPW
SYQRRDPRTGKISLDYLGKPIWEGFCNDMIAMLSQKMNFEYELISPTKGRFGERNPETGE
WDGIVGDLVSGETDFAVSALKMYSEREEVIDFVAPYFEQTGITIVMRKPVREASLFKFMT
VLRLEVWMSIIAALIVTAVMIWLMDKYSPYSSRNNRQAYPYPCREFTLRESFWFALTSFT
PQGGGEAPKAISGRILVAAYWLFVVLMLATFTANLAAFLTVERMQTPVQSLEQLARQSRI
NYTVVEGSDTHQYFINMKFAEDTLYRMWKELALNASRDFHKFRVWDYPIKEQYGHILLAI
NSSMPVKSAEDGFHKVNERESADFAFIHDSAEIKYEITRNCNLTEVGEVFAEQPYAVAIQ
QGSHLSDELSYALLELQKDRYFEELKDKYWNRSRSNCPLTEDQEGITLESLGGVFIATLF
GLGLAMITLVIEVVYYKNKPFNIPENAAQANIIEVKAAENSVKPVSEAWHTDIERKKITP
PPSFEVAAFRGREIPSSITLGNEFKPRKRLLGSAIRASDGLLTAEDELPPYVE
>GmmIR40a GMOY004663 (TMP006751):CDS scf-648181 66980:73328 forward
MGCLVLTITLFSVFESIETFPMQMEDNSGTEVSIALSEIIKSLKPIQLAILRKPRLPFNS
SVEEDKEIDDDLYSDNNEKHVNDFIYQLHKLYFKSVVFYDVELFFQYIESSIIGSIESTN
LIFCKPQDLMDRIYERKLAHRLSLFIFYWGAKYPPKRSEIQFSEPLRAVVVTRPRNQAFR
IYYNQAVSNNNGQLSLVSWYDGENLGLSKEPLLPLASSVYANFHGRVFRVPSPPWFWVNY
KNDTHDKGSQLEYDNHENHRREESLEFQEVQVIGGRDHRLLELLANHMNFEYVYVEAPGR
TQGSLRTDNEDNDTFTGGIGLLQNGLADFFLGDVSLSWERRKVVEFSFFTLADSGAFATH
APRRLNEALAILRPFKADVWPYLILTIIFSGPVFYFIIAIPFKWHPKNDIKKQPQFYMAY
IREITHVECRVDGNQQGTQKPDAEIPKNLFDKCLWFTVHLFLKQSGTYVLADVYSAQLTS
FFARPAREPPINTLSKLQKAMLEDGYQLFVEKESSSLEMLENGTEIFRQLYALMKLQSPL
EDGYLIESVEAGIHLIADGQGNKAVLGGRETLYFNIKQYGSKNFQLSQKLYTRYSAVAVQ
IGCPFLDSLNDVLMHLFEGGILDKMTTAEYEIQSRFLNKDKAYKMILQMIDENDDLDNAA
EDSSNSKGETLKSSGANEVKKPQDGSTIIQPLNLRMLQGAFIVLMVGYTLAGRS*
>GmmIR76a GMOY008789 (TMP010969):CDS scf-651574 92956:98839 forward
MIHTKFTSIVQLYLHETTICIIWNPKVEFQLTTSHINNYAVINIDLKNLELTFDEDIVNF
SEQRQQFSKDDVKFNALTEKLTMSIKKSHCESFLAFENDVLPFVEAFLNASIYSVWRSKR
NRFVFGIQNENLCKHKFFNEQSAVLLVEQSIIDLQIFYLKTNKFQGLHSKDSASFYELDT
FNASSGKFINENDLFPDKIKNLKGREIVVTGYEYLPYSSLKYVSDGNNSYDLAFGSNSSG
AALIDGTEIILILTFCELYNCRVLIDSSEFRCSNWVFPYFLIASTSDIITGKDIDYTFLD
MSMYLARCGVTLLIPAPSAYVLYKLTCDMYAGVSVIIRQYQQYNKNSL*
>GmmIR31a GMOY012048 (TMP005356) scf-645803 241794:246174 reverse
MLNDCDSQIIARQLFDWNHFVKIYDVRNNFTLKSISERDRSRAKTAIILNCECKATQNLI
FEASQFRYFNKTYQWLLWDFHNNNCFSHLQQPGLNYLGPNAQIVVLKYEEPHYLLWWDVH
SKGRHLKADLEMHLVAKAIYSNETLQFVTKNDINDIQGIRYRGDFGGLKLRGATVIDQDN
ITDNEQIEQILSRSSKDPGVSAFTKYHYELFNILRHRLNFTVKFRNARGWAGHLDNSSYR
LGFLGIMQRNEADIGASAAYPRLNRFAEFDSLHQGWKFHTAFLYLYTPDLHAQSKKGNFL
APFEEHVWLGSLIILNMVGLVWIVEEHVKSYLNPVPSGASARVITVTSFLFSLVMYNYYT
SSVVGGLLSATDKGPSTVDEIISSPLRLSFEDIGYYKVKFRENKAPSIAKLLNKKVLPPR
DPYDLPVYTDLETAIPYIQKGGYAFHCEEVDAYPELAKIFTDSEICDLRAVSGLLGSALM
NWIVHKNSQYTEIFRHIVTQAQEVGLVQRLLRQRRPKKPPCQNLYTVYPVNLSGTLSAFV
LLAGTRIP*
>GmmIR68a GMOY005753 (TMP007862) scf648742 51688:56389 forward
MTDLRQLLADLLVASQVERCFTIVTDSWYQTLYDKYFFEYSRQPLSYLYVHVKASEDLLS
PNYQTVRVLKQIKAFNCDLHFVTLLNGLQVKRFLMFVEKYRILNMSRKFVFMHDNRLIEE
DMLQVWSKILSSIFIQTRDFTREHQYIGLEVEIMKALGKAMNFNPQLYETSDAEHERWGR
LLRNGSYTGLIGEINQGRALLAMGDLHLFSAYRDVLDFSMPHTYECLTFLTPESSQDNSW
KTFIQPFSLGMWIGVFLSLFLVGILFYLLSFLHALLVRKSVQSRKFFKLLRKRKSVAITN
FRDVRFRIYLNRIRIPLKNEDLFDNFSNCILLTYSMLMYVSLPKIPKNWPLRVLTGWYWI
YCILVTVAYRASFTAILANPAPRITIDTLEELKNSQLILTVGSEDNKHLFNNAFDKTGNI
GEGTHAYYDNEYFLRHLRLRGVQSDGDREVILHIMKQCVVNMPVVLGMARNSPLKKSIDK
YIRRLSESGLIYKWLQDVVKHFPAEEDLPQEALIDMKKFWSSFVPLLIGYFIGVLIVLGE
YWHFRQVVCKHPLYDKHNLKLYYNFLRKFSDN*
>GmmIR21a GMOY006751 (TMP008882) scf-649289 31568:38862 forward
MVRYTCRFQTYYLSEKSQDLTQFTVVCSEFNKKGVLFRCIGDKFIERYGFSTEILNSCQN

NLGSKSERLNRLLSYTGNPSKNKTQARTDILMKRFNLIDNNDRQRISLIKLINKIAKEYL
SGCSAIIYYDSFAQNTDYELLRMLFKSIPLAYQHVELDGQHQPRELIDSLEGNCNNFILF
LTEPRMARKVIGPQLESKVVIISRSSQWKIRDFLASKSSADLLNLLVIGESQTYKGTEER
PFVLYTHNLYVDGLGSNTPHVLTSWLKGALSRPHLNLFPIKFQNGFAGHRFRVFAASQPP
FIFRSRLLDSRGEEQFSWEGIEYRLLNIFGKKLNFSLEILDTRLKEHIKSPVENLQLNVA
ERSADIGMAGIYVTTERLIHTDMSVGHSKDCAVFVTLASKALPKYRAIMGPFQWPVWITL
TFVYLGAVFPIAFSDRLSLSHLLGNWGEIENMFWYVFGMFTNAFSCSSRYAWSNTRTLST
RILIGSYWIFTIILTSCYTGSIIAFVTLPAFPNTVDSVKDLLGLFFKVGTLDSGGWENWF
RNSSHEPTAKLYEKMEFVSSLEEGIGNVTKSFFWNFAFLGSKAQLQYLVQSNFSNENLSK
RSALHLSEECFALFQVGFLYPRHAIYRPKIDDLILLAQQSGLILKLENEVKWTMQRSASG
KLLQASSSNPLREVIQEERQLTTADIAGMLLLMGIGYSMGFIALVSEIVGGITNKCRQIM
QKNRLSLSSASTRSSADSVKQILTIGEKSKIYREYEINVTRKNASKNQMTFLSGINFKEL
HSNPHNLQADVNPKKAKNMLKEHENHEELVPKRYINSIIDQFLNAEEFYTFTGASNERVR
QVAVAAGERDPNTSMQQLYQEQKHIMGSFIKSDDYDEKNLSNLKLFDDEEFEKANTAGQL
TKSQLHKEVARTE*
>GmmIR75d GMOY007825 (TMP009983) scf-650671 1561776:1564788 reverse
MAQGFPVHVWSGRDYLQDTPAAPIHRPSNSFRTDNDTRRPLKFQMKALSFKIGVFLANFN
SECGLNVLRWSAASEHNYFTTNRFWLLVTQDTADVAALEDPHIFIPPDSEVRIVVVLPLK
SFILSDIFKVASFKASKQLTISRNFSHSWQMIGDLQKFGSAISYRENLENITFNTGLVIA
FPDMFTNIEDLSLRHIDTISKVNNRLTLELGNKLNMHFNTHQMDNYGWRQPNGSFDGLMG
RFQRHELDFGQMAIFMRLDRIALCDFIAETYRIRAGIMFRQPSLSAVANIFAMPFENDVW
IALVILIVFTILIFILELLFSPYKHNMDYWDCAVFVWGAMCQQGFYLNVTNRSCRIVIFT
TFVANLFLFTSFSANIVALLQSPSEAIKSLRDLTQSPLEVGVQDTVYNKVYFNESTDPVT
NLLYHKKIASKGENIYIRPMVGMEKVRTGLFAYQVELQAGYQIISDTFNEPEKCGLKELE
AFQLPIIAIPTRKNFPYKELFRRQLRWQRETGLMNREERKWFPQKPKCEGGVVGFISVGI
TECRYALVMFTMGAGTSNAAVHF*
>GmmIR25a GMOY001810 (TMP003842) scf-652157 37924:45333 forward
MHFCLTGVPSETVKSFTQALGLPTISASYGQEGDLRQWRDIDANKQKYLLQIMPPADIIP
EVVRSIVRKMNITNAAILYDETFIMDHKYKSLLQNIQTRHVIVRITQGDRERTEEIEKLR
NLDINNFFLLGSLRTIGPVLEAVKPAYFQRNFAWHVITQDEGEIASKRDNATLMFLKPVV
YAQRRERLGQLRTTYNLNEPPQIMTVFYFDLALRAFLTVRDMIQANAWPKDMIYMGCDDY
QGENTPIRNLDLKEYFIKVNEPTSYGQFELVSEQGKPFNGYSYVKFEMDIDMVEIRGGNS
VNSKPIGRWTSGLDTPLMIKDETAINNLTADTVYRVYTVLQAPFIMRDPKAPKGYKGYCI
DLMNEIAAIVHFDYTIQEVEDGKFGNMDENGHWNGIVKKLIDKQADIGLGSMSVMAEREI
VIDYTVPYYDLVGITIMMQRPSAPSSLFKFLTVLETNVWLCILAAYFFTSFLMWIFDRWS
PYSYQNNREKYKDDDEKREFNLKECLWFCMTSLTPQGGGEAPKNLSGRLVAATWWLFGFI
IIASYTANLAAFLTVSRLDTPVESLDDLAKQYKILYAPLNGSSAMTYFERMANIEQMFYE
IWKDLSLNDSLTPLERSKLAVWDYPVSDKYTKMWQAMQEAQLPANLEEAVARVRNSTSAT
GFAFLGDATDIRYLVMTNCDLQVVGEEFSRKPYAIAVQQGSHLKDQFNNAILTLLNKRQL
EKLKEKWWKNDEIQAKCDKPEDQSDGISIENIGGVFIVIFVGIGMACITLLFEYWWYKYR
KNPRIIDVAEAVPSGKRDKIPEGIILGRTEKHGNVKSNIVLRSRFNQYPSKTPKPRF*
>GmmIR76b GMOY009750 (TMP011947) scf-651871 192990:204228 forward
MATEGASAVAVAEVEAATHANLQALQMFHQRCGRRITLSNCNRTATRSVRDFSHALVFSD
EPLIDDVLFENHSWGGSIEIGVTSESPDKLELPSCATVMRNGTWVMSGIDVRKDGMCLTE
FYGTDLEMLNEMDRVGVMRTSNHELVFYVNGESQGVAARNMPKTVWALVDLYGRCVQVSI
CPTHRSGSGDFSDSGFQNTQQIVQNIDVPLCVDVTVTNAGGNNSHNQTISLTESSGLSNA
AIMPSLPVGAGSLNNMSVTNANTTASSISSSSVGDYYDMNDRLRFHTRCGSLVKLSPNCR
SAERRRPLDEFNNGVVMTHRPLKDNELFEIRIDKLVDKWSGSIEVGVTTHNPSVLHFPAT
MTNMRSGTIMMSGCGILTNGKGTRRQYGEFNLDELREGDRVGMMRKLNGNLHYYINGQDQ
GVAATRVAPTLWGVIDLYGMTIKVTIVDRDEREQQNLVIRRNNLISSAAMVGGLQGVNNS
PINGNGNSAVPTPVLSLLSPESEVAAMADSALTATAANPRNDDRLTFHPLCGSHATVTHS
GRTALRPNASDDFNNGVILTRRPLRPNELFQVRLERVVTKWAGSVEIGVTTYSAEELDFP
FTMTNVRSGTWMMTGNGVMHNGITVIEQYGENLDRLQVGDRVGVVRKDTGTLHFWVNGVD
QGPSATNVPEKVYGVIDLYGQAAQASIIDTSECGSPDTGNSTISNTTLYSETPLRFHAIH
GKNASISNGGLTASRPNSLAEFNDAIVFSNRPLRQRELFEVVLETMVRHWSGNIEIGVTG
TRPDDIQLAANATEMDAMDTIILCGPMIFHNRKKIRSNVLIDLDTLGSGTKVGVMRNGDF
IHFFIDSIDQGPACECHSPSVWAVIDLYGQCAQVSLTQSSNNDMRAPYATSENSQSCQAT
SVIQPATLETKHRWTCISGNVTLSQNWTMASRLTGSSAALSRCVVFSEHPLSVGSPFEIK
MVTHNPLFAGCLNVGITDLNLSDDNVRKNIPLSIKRIPANVWYVTSNEVRYNSQLLQRAA
ASLEWLRVGDRIALELTPARTLRILLNSEDMNIQFHNVPNDVYAVVELQGSTMAVQVTSS
QGPTSPLRPCSLRLQDSLEFGADPLNKQDSMLESIDSEGLNYEFSEICGKNVKLLEENRS
AVRVQSYNQGLVYVTKPLCKGESISIKVDSINCKWKGNIGLGVVSISPSQLTNALTLSSS
IVHSKRPCWVAVDDAININGQIIPSKYGEALENIQAGTVITMTLTHAGVLIITIGSNNLQ
DLATGLPNHVYPVFDLYGKCERITILTGTDAGRNGTPIIEEPSALEHDSLNDQDSNVPQC
EKADLEVHEKETESQPSSSNVATASIPASASSMNRSVLESVSENLLLNLSIKNRTLEQNF
SEPSTSRSACCLRESLQLQHNTNLNIQRSQGSQRFRSALMAYALSPAPDNQALATLEEPT
KSETENKQLLREFFENNQEADGYFDDEAIDVNEANGASGDVGDDMSLLNNNNPDMDNYEE

DGAVGVISLHDELVANEGKDPANDNEETNLALQQQQQNQQQHQQQENRDVIIVADEENAL
QNNVTLEEGGDSSESDTDGLESNMDLYADRLVQLLELQRRQDRQDLEGIRFSQFCSDGVS
SSASSISNHFDSLQSIASIERKDCGYLKLVQQFRLSLLLPDTFFRPHHEPICFCAHCNAQ
ASEKLHGWVYFKLNQQTVNSNIAAVSQNTFDSDNLHDDWLPLYYMTRVDKIRSILDRGQP
LPMDSASDTHCHGSGGNQKDEPGARLELHFTPNSAAVIPINGQYKYLRNGYFYTINTAFE
VYVRRQSLCSAGKSKNSGRTKLQPIERRASTSEVDVASASPAPSNHCQDLGVGPSMKMKG
IELMLAALCLTCDFKADNYPNDFLVMDDNHFTNDLDIEDRKVLTNAKDIVSSDKHEQLEK
LRNWINGRELQIATLEDYPLSYTEVLGNGTRVGLGVAFELIDFLKQKFNFTYKVVVPGFN
IIGGTVDYTNSLIELLNGSQVHMAAAFLPLLSDQRGYIFYSTVTLDEGEWIMVMQRPHES
ASGSGLLAPFDYRVWTLILVSFLAVGPIIYILIILRNRLTGNTEQKPYSLGHCAWFVYGA
LLKQGSVLSPIAGLSVFFLLAFFFSFVNLLADSTRLLFATWWIFITILTSFYTANLTAFL
TLSQFTLPFNTVDDILHKNKYFVTIRGGGVEYAIKTVNESLYMLHYMAENHRAVFSLDIN
DTVNLRTYVEEQGFVFIRDRPAITHTLYTDYRQRRATSMNDEKVHCPFAQAKYPILKKKR
SFAYPIGSNLSHLFDRELLNLVESGIIKYLAVKNLPNAEICPQNLGSTERQLRNGDLIMT
YYITMVGFVISLVVFISELLFRCINQLFTPNNVADAVANQKAIVSRANNQDLILTMKGTP
PPPYQSIFTGGSERNAENKWKHVFGSKTQQLSREYPGFTSTTGLNSIGSHGIRRLINGRE
YIMYRAPNGQTQLVPVRAPSAALFQYIYVD*
>GmmIR84a-A GMOY002585 (TMP004633) scf-644004 3101:6575 forward
MQAFRDFLLNQHLKHAIIIRQRQQQQQQQQGQRDYLIEIRDLAMHCHIRFYRVPAKEDIE
NLFYKSSPKVGIYMNISDPEESLHVLQDFAVKDKFKNSYAWFITMQAPADGLLDNFKIIK
NIFDPLRLHINADVTVAIRVNGSHFHLYDVYKILPDFPVTIEIKGNWSLQKGLIIDHKFN
YGFMARRTNFQNISINVATVLHDKPKHFANYTFLSNNQDWKELDPMPRQVYGLMRPMEDL
YNFRANFILRHPRHVSIRNVYLAPLAVRVWWCVLALIILTIVLLVIQIRQEYKRRSEHHE
QSQQLEKHVDFAMLVAAEALLMQGPPSEIFHLISSRTLIATVCVFVFILMEFYNGYIVGS
LLAESPRTLTTLEALYNSNLELGMEDIQYNYNIFKNSSSQLMHIIYKERILGANGKSHTN
ILPLEQGINRIAQGGFAFHMSIDRAYRLLEIDVVGFSFDSRY*
>GmmIR75a GMOY008540 (TMP010712) scf-651418 21600:23559 reverse 539aa
MQIGFFLLLKSHLNLFKHYHHWLIYDGSNELIAFQNTFQHLNLSVDTDINYVIKNTTTTL
VEPSSSSFVVYDVYNNGFQIGGSLNITVDYEVECNSLACLKCKYLSKLHMRSLYGNRERL
SDIRMRVAVVVTRFDLKAPHSEINAFLSSQRDFYIDPLARFGFQVLLLLKDSLYTKVSLQ
YFDKWSDADNTGGVVGALVNHTVDLTSAPLVMSPLRFHFITPIAPTGYFRSVCMFRTPRN
SGIKGSVFVEPFSMQVWIYFGVILALAALLLWFTFLVEYYKMGKHSTFIPSFLTTCLISI
GSACAQSSSLLPGSIGGRFLFITLMLISFIMLNYYTSVVVSILLGSPVKSNIKTMSDLTD
SNLDVGFEPLPYTYAYLNTSTLPEVKRFVRHKVAPKIHTNNLWFSAQEGIMRMRDRPGFV
FVFEVSTGYHLIERTYEATEICDLNEVQFRPDTQLVTHLHRNSSYREIVRLKLHRRQWVR
KHLNCLTNNLIVHVGLEYTGSLFIMLLVAYILVLLIFIAELLWNRYNIKITNPCTRNLF*
>GmmIR75b TMP_IRNew4 scf-651418 - 17212..18523 rev MW: 35543.48 301aa
MINFALLNFIVHNFMQNQVQMAAIFNCWNIKKMNSIVLQDFNELSFSLFSFFFFIPFLFL
FVAFFVSAQLQLRRMTSNYGIYMQNINVNYAIADQDFIKSYFHHQRSSMGVFMDFNCWQS
GALLHKILIFSTFSRNCFSLPFTLGEKSSKIHFAKKKSVKNFKKTRANDLKLFDHHYRWL
IYDEMSDIMKFKIFVSNFNFFIDTDLTYAMLSDEHINSSKNSTFLLYDVYNKGQKWGGRL
NMTADHAVSCTHSNCHLQKYLSTLHQRSKSGNRERLSDLTLSVCTVVSLYCTCMYIKYFE
N
>GmmIR75c TMP_IRNew3 scf-651418 - 14041..15961 rev MW: 61845.0 541aa
MRDDLKGIQNIEMHCRNVSSLQYAVLLTIQDHKNTPPCENICPYAFPLKIIYQNLPLLLH
LSKTLSHIDCNFPAGIFSPQLNIQAKHFYNECVLHTFEKLQGTCSLYKVIFLSEKFALIS
SLNTNTKSVTRKPINLPEEVLIGYLMSQNNTHIDGLPRFGFQVLMVLKDILGSNFSYQFK
ETWSVTETTGGCIGAVGYEHSADLLSTPFLFTEKRSLYSRPILRNGHFRSICIFRTPRKA
DMKTGVFMEPFSVAVWILFGIVLGLIGVFLWVIFLVEHNKMKISLTYVPSLLTTCLMSFA
TACVQSSQLVPSSFGGRLTFICLSLLTFIMYNYYTSIVVSTLLGAPVKSDIKTMGQLADS
NLEVAMEPLPYNAAYLNFSKLPEIKRFVKRKIESKSNSGSVWLPAEEGILRVRDEPGFVY
IFEAFSFGLVERYYDAQEICDLNEVLFRPPADLHTHVNRNSSFTEILRWRLVRIFETGIY
SKFDRHWINSRLNCYLGNNALIQVGLEYTAPLFAMLFCSYILVLIVILLEILWKHFERRE
A
>GmmIR64a GMOY000804 (TMP002813) scf7180000640481
MYHWLLMGDYAFNRQTETDDANDVKTMLNSSNNGSSSSSSSSSSSSSCSSSSSSGNCNNI
NTIKENKKEVVSNNELLVKVGKKSNPSSSDVGGASTSGRTDTIGSGGGGGGNQYMAGVDA
EVGALHADDTETIEKFLEKLNININTELTLAKRTRVTRYIASSTPKPQYNDDYYSTLNEM
DYYKLYDVWNPGLQYGGELNVSAIGNFTLVHGLQLADWYQLSPPVTRRMNMHLARVRCLV
VIIHKNRSDTLEHYLNTHYDTHLDSMNRFNFALLSNVRDLFNFSFILSKTSSWGYLKNGK
FDGMIGALVRKQADIGGSPIFFRIERAKVIDYTTRTWIASIVGTLLMEKPKTIRTLRDLI
HSSLEIGIEDIAYNRDYFLRTKDPVAQELYAKKVTAVPASENGFTDVPQDGVPPTMAVEL
SDAAKAKAYRDILHSHETGAHAKTSEASNWYEPDFGVNKIKKGRFAFHVDVATAYKIIGD
TFTEKEICDLTEIQLFPPQKMVSIVQKGSPLRKVITYGLRRVAEVGITDYQRKVWHFHKP
RCIKQLHTDDLKVDMQTFTSALLVLIFGYAVSLLILGLEIMYHKLWQRYATT*
>GmmIR10a-like GMOY004578 (TMP006664) scf-648151 63913:65236 reverse
MRRYPLRVQIFKSTFARPQIDAVTKKVRYIYGVDGRVADLLQEYMNFTMDLQEPNKFYFG

EKDANGSYNGVLGSIIRNEVDLCLTAFFVKDYMVDEMEFSVSVYDDKICICTPKAKRIPD
SILPLLSIRSDLWLAFIISGFLCSTIWTSIRIINLKMKMPLTASAAKKDYLAKSIWQQYG
RIFNDTWVALVRVNIVRYPPFANECIFIASLCLISMVFGAVFESSLATAYIRHLYYKDIQ
TLQELDNSGLVLMYKYTSMVDNLFDEETLLFSNLKRKLRKADINAKLLNDIAFKGGKAAV
SRHSKLMLHSYHYIANKQIWIVPQIPKSYTLSYVWPKNAPWRERINQLLLRFQATGLLPK
FIKDMQIDTNIRDMTKYKNEEQQDPAGLTITDLQLAFYVIFFGSVMAGMALIVEQVCFSL
TS
>GmmIR84a-B GMOY008188 (TMP010354) scf-650995 8901:12374 forward
MQAFRDFLLNQHLKHAIIIRQRQQQQQQPQGQRDYLIEIRDLAVHCHIRFYRVPAKEDIE
NLFYKSSPKVGIYMNISDPEESLHVLQDFAVKDKFKNSYAWFITMQAPADGLLDNFKIIK
NIFDPLRLHINADVTVAIRVNGSHFHLYDVYKILPDFPVTIEIKGNWSLQKGLIIDHKFN
YGFMARRTNFQNISINVATVLHDKPKHFANYTFLSNNQDWKELDPMPRQVYGLMRPMEDL
YNFSLVGIQIRQEYKRRSEHHEQSQQLEKHVDFAMLVAAEALLMQGPPSEIFHLISSRTL
IATVCVFVFILMEFYNGYIVGSLLAESPRTLTTLEALYNSNLELGMEDIQYNYNIFKNSS
SQLMHIIYKERILGANGKSHTNILPLEQGINRIAQGGFAFHMSIDRAYRLLEIDVVGFSF
DSRY
>GmmIR84a-C GMOY004518 (TMP006602) scf-648134 49775:56259 reverse
MFFVFILWLLLANYCSTAAIEMQAFRDFLLNQHLKHAIIIRQRQQQQQQQQGQRDYLIEI
RDLAMHCHIRFYRVPAKEDIENLFYKSSPKVGIYMNISDPEESLHVLQDFAVKDKFKNSY
AWFITMQAPADGLLDNFKIIKNIFDPLRLHINADVTVAIRVNGSHFHLYDVYKILPDFPV
TIEIKGNWSLQKGLIIDHKFNYGFMARRTNFQNISINVATVLHDKPKHFANYTFLSNNQD
WKELDPMPRQVYGLMRPMEDLYNFRANFILRHPRHVSIRNVYLAPLAVRVWWCVLALIIL
TIVLLVIQIRQEYKRRSEHHEQSKQLEKHVDFAMLVAAEALLMQGPPSEIFHLISSRTLI
ATVCVFVFILMEFYNGYIVGSLLAESPRTLTTLEALYNSNLELGMEDIQYNYNIFKNSSS
QLMHIIYKERILGANGKSHTNILPLEQGINRIAQGGFAFHMSIDRAYRLLENKLNERQFC
ELQEIRYISGYSTGLILAKSTPYREYVAQVTLKLRESSLMQYNNKLWELHKIDCRLIKGN
EIIVDMEHFAAALVFLGCAIVLSLIVLGIEIFYKKCKKLQK
>GmmIR56d-like TMP_IRNew5 scf-652049 MW 71331.016 rev
MRLTRGATYLVVTLMASLCIAEAESQAFNSSLAIGIAKHLHKAYKFHNFVFFLSENLLSD
TDVMTDFLPQFWKKFPRMPYVIVTRQHYRMQGFLDRRSLIYIFITGYDDPILPLAALNLR
RIRYYPVIFVFLPRRIYDEPFTADSEAYANFTDNLVNIFDWIWKRQFLRTFLLSVDNNIY
KHDLFPVPRIINKTDNWSVKDLFVKVGNNFKGHIIKTPICYDLPRVFRLPNNELSGVSGK
LFKAFVHFINASLIDDATCYSIKKPYNLTKILLDVSEARLQISVHSYTEMLNSPAGSSYP
IGINDFCFIVPFRRRSPEHLFLQRTLQNSTWLLVVFAVFYVTFAIWWLTTEERRDFSLAF
LQSISSFLYVPPLQILYLHTNYMRFFSILLFILGFCLTNLFQTKMSSYLTASLPDSQINT
ISDLLATNLKIVVMEHELPILKAIGYDDAFLKRFLPVKKNFMDSHRDRLNTTYAYSSQTD
RWNFLKNQQLQLKYPVFHLSQICIGPYYLVYPIREDSVFYKPLKYFILYTHQYGLQIHWN
RQAFSEALALKYVNMMDVNEEIKPLSMNFFRSIWLIWVIGITLSGAVFIVEMKETVFKTI
KARAANQ
>GmmIR56b-like TMP_IRNew6 scf-652049 20381: 23107 MW: 48509.07
MIKVNSVYHIYIYIYMRVCVCVAISIRLSFDCKDLPRAFFYRHKSGFEELTGTHAKVATE
FAIHQKYKFRVENMVNFSIAEIHGLIEDGYFNWSMRPNTYGDSSANIDYTYPLEHSRYCV
IVPTRPRLARFWYVVWPFDRYIWLLIILAIVYAAALMTIQRHPTVSFTRNLLYSFALMHD
SPNANINKNNKCARVQIFFLLIFGLGFILSVYYITFLASYFIYPISQPHIDSLDAIIDAG
IDVIITPRTYEILRSGNFSNFHEFEKVMRIVPINSFMEIINSHHAHILTQEDWDLIDRMQ
THLILPKFKYSEICFGDYYTALPLKMDSSFAKNLSDYLLRIKESGLWLHWEDESFYVALK
TKLVELLVDDYPAEPLDMQFFLIAWIVLVGGWTLSALCFIAEILFIKYSKKLH
>GmmKaiR1A.1 (var 1) GMOY001514.RA (TMP003539) scf-641481 6802: 19237 MW: 95269.445
MLFPKLKCKLKNIENMCLLLFLSFSLFSIEIINAQKTNVGLICENDNTDVERIFQLAINK
VNPDMEDMYLNGVSISIEPGNAFETSKKLCKMLRQNLVAVFGPTSSMAAKHAMSICDAKE
LPFIDTRGDFVVELPTINLHPHPTQLAVILKDLVTMLEWEAFTIIYECGEYLPVVNDLLQ
MYGPAGPTIVVRRYELDLNGDYRNVLRRIRNSGETSFIVVGSIETLPELLKQAQQVGLMT
GAYRYIISDLDLQTIDLEPYQHSDANITGIRLVSPENELVLEVAKALYEEEDPYLSITGL
SGDIKFDYEGLRTDFLLEIIELTVSGLQKIGTWTGDEGASLDRPPSLISAEPDQRSLVNR
SFIVITAISEPYGMLKDSPAKLEGNDQFEGFGIELIEELGRKLGFSYTFRLQEDNKYGSL
NPVTKEWNGMLLEIMEGRADLGITDLTMTSDRESAVDFTIPFMNLGIAILFRKPMKEPPK
LFSFMSPFSGEVWLWLGISCVSVSLSMFILGRISPAEWDNPYPCIEEPTELENQFSLTNC
FWFSIGALLQQGSELAPKGFSTRSVASFWSFFILIMVSSYTANLAAFLTVESLSTPIENV
EDLAANKGGVNYGAKLGGSTFNFFQGAKFPTYQKMYEFMDSHPEYMTSTNAEGVTRVENE
NYAFLMESTTIEYITERRCTLTQVGTLLDEKGYGIAMKKNSPYRDVLSQAVLQLQEQGVL
VKMKTKWWKEKRGGGACTQKASSDSAAELGMANVGGIYAVLLAGSIIAVFCGIVEWICGL
YSTAQANKIPFKTELMEEIRFVLKCSGNTRPVKYPKPSRSASSRSKFSVSSKDSSSVTVD
SLATDDQQLQKRKK
>GmmKaiR1A.2 (var2) - 6802: 19237 MW: 95440.71
MLFPKLKCKLKNIENMCLLLFLSFSLFSIEIINAQKTNVGLICENDNTDVERIFQLAINK
VNPDMEDMYLNGVSISIEPGNAFETSKKLCKMLRQNLVAVFGPTSSMAAKHAMSICDAKE
LPFIDTRGDFVVELPTINLHPHPTQLAVILKDLVTMLEWEAFTIIYECGEYLPVVNDLLQ

MYGPAGPTIVVRRYELDLNGDYRNVLRRIRNSGETSFIVVGSIETLPELLKQAQQVGLMT
GAYRYIISDLDLQTIDLEPYQHSDANITGIRLVSPENELVLEVAKALYEEEDPYLSITGL
SGDIKFDYEGLRTDFLLEIIELTVSGLQKIGTWTGDEGASLDRPPSLISAEPDQRSLVNR
SFIVITAISEPYGMLKDSPAKLEGNDQFEGFGIELIEELGRKLGFSYTFRLQEDNKYGSL
NPVTKEWNGMLLEIMEGRADLGITDLTMTSDRESAVDFTIPFMNLGIAILFRKPMKEPPK
LFSFMSPFSGEVWLWLGISCVSVSLSMFILGRISPAEWDNPYPCIEEPTELENQFSLTNC
FWFSIGALLQQGSELAPKGFSTRSVASFWSFFILIMVSSYTANLAAFLTVESLSTPIENV
EDLAANKGGVNYGAKLGGSTFNFFQGAKFPTYQKMYEFMDSHPEYMTSTNAEGVTRVENE
NYAFLMESTTIEYITERRCTLTQVGTLLDEKGYGIAMKKNSPYRDVLSQAVLQLQEQGVL
VKMKTKWWKEKRGGGACTGVESDSGALALEISNLGGVFLVLIAGSFFGLFVSIVEMICGV
KQRCDENKIPFKTELMEEIRFVLKCSGNTRPVKYPKPSRSASSRSKFSVSSKDSSSVTVD
SLATDDQQLQKRKK
>GmmKaiRIA (var3) - 6802: 18019 MW: 92118.19
MLFPKLKCKLKNIENMCLLLFLSFSLFSIEIINAQKTNVGLICENDNTDVERIFQLAINK
VNPDMEDMYLNGVSISIEPGNAFETSKKLCKMLRQNLVAVFGPTSSMAAKHAMSICDAKE
LPFIDTRGDFVVELPTINLHPHPTQLAVILKDLVTMLEWEAFTIIYECGEYLPVVNDLLQ
MYGPAGPTIVVRRYELDLNGDYRNVLRRIRNSGETSFIVVGSIETLPELLKQAQQVGLMT
GAYRYIISDLDLQTIDLEPYQHSDANITGIRLVSPENELVLEVAKALYEEEDPYLSITGL
SGDIKFDYEGLRTDFLLEIIELTVSGLQKIGTWTGDEGASLDRPPSLISAEPDQRSLVNR
SFIVITAISEPYGMLKDSPAKLEGNDQFEGFGIELIEELGRKLGFSYTFRLQEDNKYGSL
NPVTKEWNGMLLEIMEGRADLGITDLTMTSDRESAVDFTIPFMNLGIAILFRKPMKEPPK
LFSFMSPFSGEVWLWLGISCVSVSLSMFILGRISPAEWDNPYPCIEEPTELENQFSLTNC
FWFSIGALLQQGSELAPKGFSTRSVASFWSFFILIMVSSYTANLAAFLTVESLSTPIENV
EDLAANKGGVNYGAKLGGSTFNFFQGAKFPTYQKMYEFMDSHPEYMTSTNAEGVTRVENE
NYAFLMESTTIEYITERRCTLTQVGTLLDEKGYGIAMKKNSPYRDVLSQAVLQLQEQGVL
VKMKTKWWKEKRGGGACTGVESDSGALALEISNLGGVFLVLIAGSFFGLFVSIVEMICGV
KQRCDENKVLQKMSFVIPKSPIDFTPCIFFSNVCHLFLRTIGTI
>GmmClumsy GMOY006490 (TMP008613) scf-649055 60746:70312 reverse
MSQDMFLLDCPTYEDYLDTFITRNDYRFIRNIRFCRMLVELGYRSSTEIYTPEQFVLHKA
AAQESLWPTKKSTIFFSDHLKSLDPVLRELALRERPNIQKMLSTIIFLKHRLKSGFEISG
YIDYEHSLRRANLHAEDHIDWGGVFEERVVLKPKRSHLSYFDWHKGHVYYNNSDNYAVVH
DVEYGLIFMHKALFQLGAIFETGSDELAIAFHTAIERANIYERSFELLPIVVYVDTEDSF
LMEKTACNLISEEGVIAIISPVSGGSDIIASIANTLDIPHLEYDWSPSEALDKKQHMAMT
LNVHPDNLDFARGLADIVQSFGWRSYTIAYESYTGKPITKLISINPSNELFFATELQQLQ
DILQIGERDSNPTTMFQLADDFDYKPMLKSIKMSTDNCLILHCSTEKIVEVLKQAKELKM
LGEYQSIFIPNLDTHTIDMPDILTVGANISTVRIMDPTDFHVKNIVHDWEEHEKREKRYF
RVDPDRAKTNMILLNDAIWIFVKGLAELLQEEELQVPKVECRRNRSWPMGRRIIEFMKAR
SDEVATGRMDFNQYGQRSFFTLRFLGLTQIGFQELATWDPVNGLLSKESDDMSDKLLGQK
MQNKTFIITSRIGAPFLMLREAEEGEILQGNARYEGYSMDLIDAIAKKLEFKYEFQLAPD
GRYGSYNKDTKQWDGLVRQILDGNADLGICDLTMTSSRRTAVDFTPPFMTLGISILYAKP
EVPPANYFSFLSPFSVDVWMYMATGFLAVSLLLFILARMAPADWENPHPCKEPEELENVW
TMTNTTWLAIGSIMGQGCDLLPKAASTRLATGIWWFFALMMLNSYTANLAAFLTMSRMGS
SIKSAEDLAAQTKIKYGAVLGGSTMGFFRDSNFTTYQRMWAAMDSNPTVFTKTNNEGAER
VLKGKRLYAFLMESTTLEYIVERQCDLMQVGGWLDYKTYGIAMPFNSPYRKQISGAVLKL
GEAGTLSALKRKWWKEMHGGGNCTVAEAASSSTPELKLVNVGGVFLVLGIGLLCSIVIGL
VEFLWNLKTVAIEEKISLKDALKLETMFALKVWITTKPVRGSSENSSSSSSSSSSSPSI
HSKKSRGLSISHSMRSLKSSIHETMQTAVGAKMRKIGSMFSLKSVHSGKPEIEIEPPSEV
DDKVDKEEKATQIEMENSQEHRHRHHHKQRNHLHPDDIEQDIQRDHRRDSKASKAHLNV
>GmmGluRIIA.1 (splice variant 1) GMOY012165 (TMP008614) scf-649055 71321:79880 forward
MYMCREIFFITLLKISIEFMDERKTEVKVGAIFFKNEETLELSFDAAFQEINNIKLFELH
FVTFKRYIPIDDSYVLQQLTCELISNGVAAIFGPTSKASNDIVALIANSTGIPHLQYDWN
IETTIEQLQLNHRMSVNVAPTLTAISKAYWDIIKINYDWKTFTIFYQSEQGLTRLQDLMG
IHTVNKDAIKLKRITDYSQDIRVLWKEAAEALHEQRFILDCDPKSLVDLLNTAKDFKLLG
PFKQWFLTHLDSHSSSLKLIYNEQFKANITGVRLKLNDLNPYERRKTRITLIDQIFGNQT
MLPILMYDAVILYANAARNIIISLKEYQHPYKRCDLGRYGRSSWLVGRLIVKEMKELSED
DVEPLFKTENMKIDENGQRSVFNLEIYKPTVNEPLAIWKSDGLVTPIRVKQEFQATAVVP
DFSQVRRTYIVVTHFEEPYFMLKADHENFRGMEKYEGYAVDLIQKLSELMDFEYDFMIVG
GNGKFNQITKEWDGIIRKLIDHQAQIGISDLTITQLRRKYVDFTVPFMQLGISILFYKRD
PEPKNMFAFLQPFAKEVWIYLILTQLIITLMFVFMARISNHEWENPHPTNLEPIELENKW
YTSNSAWLMIGSIMQQGCDILPRGGPMRMLVGMWWFFALMILSTYTANLAAFLTSNKMQS
TIKSLKDLIEQDYVKFGTFHGGSTSQFFSESNNTDYHKAWNQMKSFQPSAFTSSNKEGVE
RVRKGQGNYAFLMETTSLSYNIERDCHLQQIGHQIGEKHYGLAVPLGADYRTNLSVSLLQ
LSEKGELYKLKNKWWKNHNNTCNDNKEADLDGDELSIIELGGVFMVLGGGIVIAIVIGCC
EFLWNVQRVAVEERVTPCEAFKAELIFALKFWIKRKPIRISSGSSTSQRSSRSTSRRSSG
QSKQSKRSKISKRSKHSKRSHSRSHSVPKFSERN
>GmmGluRIIA.2 (splice variant 2) GMOY012165 (TMP008614) scf-649055 - 71321: 79880 MW: 76078.47

MYMCREIFFITLLKISIEFMDERKTEVKVGAIFFKNEETLELSFDAAFQEINNIKLFELH
FVTFKRYIPIDDSYVLQQLTCELISNGVAAIFGPTSKASNDIVALIANSTGIPHLQYDWN
IETTIEQLQLNHRMSVNVAPTLTAISKAYWDIIKINYDWKTFTIFYQSEQGLTRLQDLMG
IHTVNKDAIKLKRITDYSQDIRVLWKEAAEALHEQRFILDCDPKSLVDLLNTAKDFKLLG
PFKQWFLTHLDSHSSSLKLIYNEQFKANITGVRLKLNDLNPYERRKTRITLIDQIFGNQT
MLPILMYDAVILYANAARNIIISLKEYQHPYKRCDLGRYGRSSWLVGRLIVKEMKELSED
DVEPLFKTENMKIDENGQRSVFNLEIYKPTVNEPLAIWKSDGLVTPIRVKQEFQATAVVP
DFSQVRRTYIVVTHFEEPYFMLKADHENFRGMEKYEGYAVDLIQKLSELMDFEYDFMIVG
GNGKFNQITKEWDGIIRKLIDHLSEKGELYKLKNKWWKNHNNTCNDNKEADLDGDELSII
ELGGVFMVLGGGIVIAIVIGCCEFLWNVQRVAVEERVTPCEAFKAELIFALKFWIKRKPI
RISSGSSTSQRSSRSTSRRSSGQSKQSKRSKISKRSKHSKRSHSRSHSVPKFSERN
>GmmGluRIIB TMP_iGluRNew1 scf-659055 80691:85374 forward
MEAFDSAVTDINALDTELRFKVVKHFLAEDDSVILQNLACDLLEQGVLAIFGPSSKSSSD
IVAVLCNQTGIPHLQFDWSSKAKEGDMRYYQFTINVAPSESTLSQALWEILRSREYDWKS
FTIVYEVESNLSRYQHLLSWKPFHRTGLKMVKFRRGDDYRILWKILSNMKEKFILLDCPP
DILKEVVNISIYYNMTGGFNNLFLANLDTHASGLASLYSKKFTAKVTAIRLRRYIPPGHQ
GEYDIFQQEPEESVVLTKSQSLYDAVILYYEGLRMVIKSGSFRMPGKTKCGLQSWSIGEE
IVKHIKQFKTRGGNTSAFKTQKMFLSSTGLRDDFNLEIYNPLVERIVYIWNKGHGLVDYE
KLQDSSTQEDKRQKLSPGKEDFSLKRVRYTIATRMGEPYFMWRPEPEGVHYEGNERFEGY
VVDLIYKLAEECKFDFIFEPNAQLGICDLTITQARRSVVDFTVPFMQLGVSILHHKPKPA
EKRLFAFLKPFSLDVWLCLLVALLLMALLLTLMVRLNPSEWQEQKIDDCLVIIENKWFLR
NSLWLYIGSILGVSCDLLPKASSLRLCTAFWWLFALLMSRTYTAKLASFITASKLEGSIK
NLHDLVEQNKVQFGMVKGGSTSLFFSESNESDYRVAWNKMLAMKPEAFTANNREGVDRVR
RSKGRYAFLMETPNIQYYAQRNCELMQIGPTFSEKHYGIAVPLNAPFRSNLSVGILKLSE
KGILYNLKNKWFNNNETKCIPTDLINFEQSTQFNINSVGGLFIVLLAGLLVAIVLGIVEF
LWNTQKIAIKEKVTNEVFFNSFIVF
>GmmNMDAR1 GMOY007988 (TMP010147) scf-650827 1035:15119 reverse
MFVCFVNIVVNNSKVLVFLASKYRNQTFSRIYNLHVRIELGLSLRPCVKMMLRSAIVLTI
LIVFNKVTSFSNNSDDAEKYNVGGVLSDFESEKHFRLTISHLNFDQNYKTSRNMTYYGKT
IRMDKNPIKTVFNVCDKLIDKRVYAVVVSHEQTSGDLSPAAVSYTSGFFQIPVIGISSRD
AAFSDKNIHVSFLRTVPPYYHQADVWLEMLNHFGYTKVIIIHSSDTDGRAILGRFQTTSQ
NYYDDIDVRATVELIVEFEPKLESFMQHLDDMKTAQSRVYLLYASTEDAHVIFRDAGNYN
MTESGHAWFVTEQALSANNTPDGVLGLQLEHANSDNRHIRDSVYVLASAIKEMMTNETIT
EPPKNCDDSGVNWESGKRLFHYLKTRNITGETGQVAFDDNGDRIYAGYDVINIRAKQKKL
PVGKFYYDSGKATMRLKINDSQLIWPGKQKKKPKGIMIPTHLKVLTIEEKPFVYVRHLTD
DEVQCMEDEVSCPLFNTSNGIENDFCCRGYCIDLLKALSQRINFTYNLALSPDGQFGHYI
LKNSSSSSLPPSPPRKEWTGLIGELVNERADMIVAPLTINPERAEYIEFSKPFKYQGITI
LEKKPSRSSTLVSFLQPFSNTLWILVMVSVHVVALVLYLLDRFSPFGRFKLSHADSHEEK
PLNLSSAVWFAWGVLLNSGIGEGTPRSFSARVLGMVWAGFAMIIVASYTANLAAFLVLER
PKTKLSGINDARLRNTMENLTCATVKGSSVDMYFRRQVELSNMYRTMEANNYDTAEQAIE
DVKKGKLMAFIWDSSRLEYEASKDCELVTAGELFGRSGYGIGLQKGSPWTDAVTLAILEF
HESGFMEALDKAWIFHGNAQQCEQFEKTPNTLGLKNMAGVFILVAAGVAIGVGLIIIEVI
YKRHQVKKQKRLDIARHAAGKWRGTIEKRKTIRASLAMQRQYNVGLNSHTGTISYAVDKR
RYHRLGGHRGPENAWTGDTDALRSRRCLEDTAKLKHSPKIHAPLPLLNKNRQPINLLPPR
YSPGYTSDVSHLVV
>GmmGluRIIC GMOY012186 (TMP010998) scf-651593 231237:241716 reverse
MNGNMSYHTFEIEAKESYCIGKNRRRDFSPLKLRSEYNIMPDKAIFADSDYNSCSECSYC
KGVDDGIKTAGLKIFAMIMCHAWRKRREQLNELRRTMEELKQGSIKAKGQLQIYNQLLRV
EQKRNDELSEQVREALNTLKNSRASYEFLSSSIVNLKTDKSLLEEELKSKYEEYDALHAV
LSQTKTDLFRCMMDQRNLQVQLCKEQRSVQSLENQKNKLINENKLFYNCRTIFCASGKNS
DDTMNVVTHCFLFFKLLSFVCYSNQLQQLNIGAFFYENEFEYQRAFESAVLRINSENKSN
FKLLPIIKKLSSYDGPIALKREACNLIGDGVEAIFGPSSQAESDIVAVICNNTGIPHILF
DYWIDENHSRKYKHQMTLNVFPSPSILSKAYADIVSSFGWKKFTIVYDADDVQAYTKLQD
LFQLHSLHSDIVRIRKFKRDEDYRILWKSIKGERRIVLDCSPNILLDLLKAAVPFDLTGQ
FNHLFLTNLDTHSSAIKSLRDNPTYAMNVTAVRLKLNDDFYKQSLAHKRDFANIKFHRQP
LKVTLIYDALMLYGIALKNITTNYPTIQRYVSCEYNEYNSTYSSPAGSYVYNSMTNISPL
DGVHFRSGSLLFDDNGQRTGFEIELYEPLNDYGTAFWNTKGQITQQHTGANVLKKNCVPS
CNPHRGTVLYGKVLSTAKTENVTHNGRYMGYAVDLIEALSKEIGFEYIFVPVADNGYGKY
NKETKQWNGIIGELINNDAHMGICDLTITQARKTVVDFTVPFMQLGISILAYQKPTESKA
WSAFLDPFTVDVWIYVMISIFIIALLFIFMARDEWENPHPCNKDPICLENKWNISNSFWL
TMGSIMTAGSDILPCSASMRTFNAMWWIFAVIIANSYTANLAAFLTNSKMEGSISGIQDL
AEQTKVKFGTMEGGSTFTFFSESNETTYRLAYNLMQNNDPPAYTKDNKEGVDRVLKNNGS
YMFLMETTSLEYNTERNCKLKMLGEKFGEKHYAIAVPFGAEYRSNLSVAILKLSERGELF
QLKNKWWKNHNNTCHEGVQMDASETPDMTFSEVRGIFYTLGVGVIIAYTVGIFEFLIHTQ
KVAAAEKLTFKKAFLKEVIFVLCVWNNKKPLKLASSNGSKATSPNAP
>GmmGluRIB|AMPA GMOY012136 (TMP012409) scf652090 220218:284541 forward
MPSGLRFFVFLWITFANSWISYTTITNSRSSSLGISSGISAAANTRHFGLMMGVQAQPSL

TEKIPLGAIFEQGTDEVQSAFKYAMLNHNLNVSSRRFELQAYVDVINTADAFKLSRLICN
QFSRGVYSMLGAVSPDSFDTLHSYSNTFQMPFVTPWFPEKVLTPSSGFLDFAISMRPDYH
QAIIDTIQFYGWRKIIYLYDSHDGLLRLQQIYQGLKPGNESFQVEMVKRITNVTMAIDFL
HTLEDFGRFTNKYIVLDCPTEMAKEILIQHVRDISLGRRTYHYLLSGLVMDDRWESEIIE
FGAINITGFRLVDTNRRLIKEFYDSWKRLDPNTSVGAGRESISAQAALMYDAVFVLVEAF
NKILRKKPDQFRNNIRRGQTAMAAASTSANTTGLTSNGMLGGGIGSNLMGGSSSGGNGNG
IGGGNSGGVGGSNSNNNNAPRALDCNTSKGWVNPWEHGDKISRYLRKVEIEGLTGDIKFN
DDGRRVNYTLHVVEMTVNSAMVKVAEWSDDAGLQPLSAKYVRLKPHAEIEKNRTYIVTTL
LEEPYIMLKRPVVGETLDSNDRFEGYCKDLADLLAKKLGINFEMRLVKDGTYGSENPNVR
GGWDGMVGELVRREADIAIAAMTITAERERVIDFSKPFMSLGISIMIKKPVKQTPGVFSF
MNPLSQEIWRYIQLHWCEYCALLCFTFLAVRMACGAIHMHIYDQLSTQQPPGIIGGVPMP
GPLTTTSAGGVIGGGGAGGVGGGGGVVGIGTGSNIGGSGALPTSATVAVNDFSILNSFWF
SLAAFMQQGCDISPRSISGRIVGAVWWFFTLILISSYTANLAAFLTVERMVTPINSPEDL
AMQTEVQYGTLLHGSTWDFFRRSQIGLHNKMWEYMNSRKHVFVNTYDEGIRRVRTSKGKY
ALLVESPKNEYVNAREPCDTMKVGRNLDTKGFGIATPIGSPLKDPINLAVLSLKENGELI
KLRNKWWYDKTECNLNKDNQETSHNELSLSNVAGIFYILIGGLLVAVFVAIIEFCFRSKS
STAAKANGSMLSSTSSSVHQRNSLSDAMHSKAKLTIQASREYDNGRVGYLNCASLQYYPT
GQLSAGGAPANTPPNDPETLHMNAHSQV
>GmmGluRIA-a (ampa) GMOY01262 (TMP011168) scf-651742 15248:22242 forward
MWIESKIYTTLRKSESMWKFPKLLRKRKVPRLRENRQLLVSREAEECNERKHGKKCHRES
QCACFLLIFFPNFKNRKLSYYPLNGYDGTLNNIRLGLISDSNADNLRRTFDYAIEVVNSD
LTVPLTGHQETMDYGNIMQGMNRLCKLTKLGVGGIFAPSAENTASYLMHMCDSKDIPFVY
SHLSRRYDAFNLHPHPLDIAKAIHAIIEEFEWSRFIYLYENSEFLSILDALMSLYVSTGP
VINIVHYDLNLNGNFKSVLRRVRKSVDNRILIVGSTQSVAELLKQAQQIGIMNEDYKYII
GNLDFHTFDLEEYKYSEANITGMRMFSAEQSEVRNLMLNLGYFDETTENEIIRNGSCPIT
MEMALTYDAVVAFAETTKHLQYTPHALNCSDFSIDALEDGTTFKNYMRSILLDRNTITGH
IYFEGSVRKGSQFEIIELQPSGLIKVGTWQEHKNFTFKRPIQMKPIADNTDNSMINKTLR
VLIAVPNKPYASLVESHKKLDGNSQYEGYGIDLIKELADKLGFNFTFINGGNDYGSFNKT
TNVTTGMLKEMVEGRADLAVTDLTITSEREEVIDFSIPFMNLGIGILYLKPQKSNPTTFS
FMDPFSKEVWIYLGLAYLGVSLCFFILGRLSPTEWDNPYPCVEEPDELENQFTINNALWF
TTGALLQQGSEVAPKALSTRTVAAIWCFFTLIMVSSYTANLAAFLTIENPTSLLENVQDL
SENKGGVQYGAKRTGSTRNFFLTSEEEIYKKMNEYMLAHPEYLTETNQEGLERVKSSDVQ
NGKTYAFLMESTSIEYNTQRECKLTKIGEALDEKGYGIAMVKNWPYRDKFNNALLELQEQ
GVLAQLKNKWWNEIGAGVCNVS
>GmmGluRIIE-a TMP_iGluRNew2 scf-651742 24207:29724 forward
MDSWPCSYATVYLQHLVLIYFQSKSDGGVSPLAFSNLEGIYYVLIVGCAISMVFGIINWC
FEVARKAQNYNVPFSAALKEEFKAVTDFANNERLLKGANSIYSRSRNSSIDSTIDSFETD
SNDESLLTDRNHDQMTLVAEYAMEVANIDLQTPLVLKQEEVPFGNSFQGYGKLCKMMQVG
IGAVLGPSSKHTASHLMSICDAKDVPYFYGHMWQSAEAFNLHPHQDDISKALHSLLTEFQ
WSRFIFLYESTEYLNILNTIMALYGPNGPIITVLRYDINLNGNFKSVLRRIRKSIDNHIV
VVGSSDTMPEFLRQAQQVGIINEDYKYIIGNLDFHSFDLEEYKYSEANITSLRMFSPEKL
RVKELMLKLGYTYALDEFQNGLLKGSCPITVEMALTYDALQLIAETTKHISLKAEPLNCT
DRSDGVTEDGSTFKNYVRSLNIHKRTLTGQIYFEGNVRKGYTFDVIELQPSGVVKVGTWN
ELSNYTSQRLKPTSAIFENFDNTLVNRTFIILLSVPNKPYASLVESYKKLDGNNQFEGYG
VELIKELANKLGFNYTLINGGNDYGSYNKTTNTSTGMLKEIREGRADLAVTDLTITSERE
EAVDFSIPFMSLGIAILYVKPQQAPPATFSFMDPFSEEVWWYLGLAFLGVSLSFFILGRL
SPKEWDNPCPCIEEPIELENQFTISNSLWFTTGALLQQGSEIAPKALSVRTVASIWWFFT
LIMVSSYTANLAAFLTIETPTTLIENVNDLAENKGGVVYGAKRTGSTRNFFMTSEDERYK
KMNEFMTKNPQYLTETNQEGFERVKNSKDHTYAFLMESTSIEYNTMRECSLKKIGDALDE
KGYGIAMRKNWPYRDKFNNALLELQEQGVLAKMKNKWWNEVGAGICTTKAEQSDAKSLSM
ENLEGIYVVLLAGSGLALLHDIISWICFVIRKASNHKVSLKDAFTEEFKFVIDFSTYTRE
LKTSASIYSRSRNSSISIDTLEANSTQNI
>GmmNMDAR2.1 (var1) GMOY012037 (TMP010682) scf-651403 10436: 21305 MW: 106087.984
MPLTQIYIYIYIYIYLAILSTLCKEFLQFNVSAILYMMNNEQFGHSTASAQYFLQLAGYL
GIPVISWNADNSGLERRASQSTLQLQLAPSIEHQSAAMLSILERYKWHQFSVVTSQIAGH
DDFVQAVRERVAEMQDHFKFTILNAIVVTRTSDLMELVNSEARVMLLYATQTEAITILRA
AEDLKLTGENYVWVVSQSVIEKKDAHPQFPIGMLGVHFDTSSAALMSEISNAIKIYGYGV
EAYVSDPANRGKKLNTQSLSCEDEGRGRWDNGELYLRNVSIEGDLNKPNIEFTADGDLKS
AELKIMNLRPSANNKNLVWEEIGVWKSWETQKLDIRDIAWPGNSHAPPQGVPEKFHLKIT
FLEEAPYINLSPADPVSGKCLMDRGVLCRVAADHEMADIDVGQAHRNGSFYQCCSGFCID
LLEKFAEELGFTYELVRVEDGKWGTLENGKWNGLIADLVNRKTDMVLTSLMINTEREAVV
DFSEPFMETGIAIVVAKRTGIISPTAFLEPFDTASWMLVGIVAIQAATFMIFLFEWLSPS
GYDMKLYLQNTNVTPYRFSLFRTYWLVWAVLFQAAVHVDSPRGFTSRFMTNVWALFAVVF
LAIYTANLAAFMITREEFHEFTGLNDSRLVHPYSHKPSFKFGTIPYSHTDSTINKYFKEM
HYYMKQHNKSSVADGVADVLSGSLDAFIYDGTVLDYLVAQDEDCRLMTVGSWYAMTGYGL
AFSRNSKYVQMFNKRLLEFRANGDLERLRRYWMTGTCRPGKQEHKSSDPLALEQFLSAFL
LLMAGILLAALLLLLEHVYFKYIRKRIAKKDGGHCCALISLSMGKSLTFRGAVYEATEIL

KKHRCNDPICDTHLWKVKHELDMTRLRVRQLEKALDKHGIKTPQLRLASSSDLLNHHHLK
ERPPLLGNLSLAASAQDLYRWSYKTEIAEMETVL
>GmmNMDAR2.2 (var2) GMOY012037 (TMP010682) scf-651403 10436: 21305 MW: 99891.05
MPLTQIYIYIYIYIYLAILSTLCKEFLQFNVSAILYMMNNEQFGHSTASAQYFLQLAGYL
GIPVISWNADNSGLERRASQSTLQLQLAPSIEHQSAAMLSILERYKWHQFSVVTSQIAGH
DDFVQAVRERVAEMQDHFKFTILNAIVVTRTSDLMELVNSEARVMLLYATQTEAITILRA
AEDLKLTGENYVWVVSQSVIEKKDAHPQFPIGMLGVHFDTSSAALMSEISNAIKIYGYGV
EAYVSDPANRGKKLNTQSLSCEDEGRGRWDNGELYLRNVSIEGDLNKPNIEFTADGDLKS
AELKIMNLRPSANNKNLVWEEIGVWKSWETQKLDIRDIAWPGNSHAPPQGVPEKFHLKIT
FLEEAPYINLSPADPVSGKCLMDRGVLCRVAADHEMADIDVGQAHRNGSFYQCCSGFCID
LLEKFAEELGFTYELVRVEDGKWGTLENGKWNGLIADLVNRKTDMVLTSLMINTEREAVV
DFSEPFMETGIAIVVAKRTGIISPTAFLEPFDTASWMLVGIVAIQAATFMIFLFEWLSPS
GYDMKLYLQNTNVTPYRFSLFRTYWLVWAVLFQAAVHVDSPRGFTSRFMTNVWALFAVVF
LAIYTANLAAFMITRHNKSSVADGVADVLSGSLDAFIYDGTVLDYLVAQDEDCRLMTVGS
WYAMTGYGLAFSRNSKYVQMFNKRLLEFRANGDLERLRRYWMTGTCRPGKQEHKSSDPLA
LEQFLSAFLLLMAGILLAALLLLLEHVYFKYIRKRIAKKDGGHCCALISLSMGKSLTFRG
AVYEATEILKKHRCNDPICDTHLWKVKHELDMTRLRVRQLEKALDKHGIKTPQLRLASSS
DLLNHHHLKERPPLLGNLSLAASAQDLYRWSYKTEIAEMETVL
>GmmGluR1A-b (partial ampa-like) GMOY006890 (TMP009025) scf-649517 6035:7620 reverse
MGGFDINDKLEARDSDYMLRTQKISTDATGAARPLALYPHPYSRSSAAFLTIENPTSLLE
NVQDLSENKGGVQYGTKRTGSTRNFFLTSEEEIYKKMNEYMLAHPEYLTETNQEGLERVK
SSDVQNGETYAFLMESTSIEYNTQRECKLTKIGEALDQKGYGIAMVKNWPYRDKFNNALL
ELQEQGVLAQLKNKWWNEIGAGVCNVS
>GmmGluRIIE-b partial GMOY009209 (TMP011398) scf-651796 5803:6650 forward
MTKNPQYLTETNQEGFERVKNSKDHTYAFLMESTSIEYNTMRECSLKKIGDALDEKGYGI
AMRKKDKFNNALLELQEQGVLAKMKNKWWNEVGAGICTVSLKDAFTEEFKFVIDFSTYTR
ELKTSASIYSRSRNSSISIDTLEASSTQNI
>GmmKaiR2-like-d1 GMOY012113 (TMP005606) scf-646489 14412: 22378 MW: 110151.25
MYVKREKISLIIRSITIKIVPVGLHIMFISELCQILKVLQQQCKQCKHLMMLMNCQSEQR
KKHKHRTPGVMTPIIYLVYLSVFISFRIGDVIALPPIIQLGAVFTEDQRSSNIESAFKYA
VYRINKDKNILPDTQLVYDIKYAPRDDTFRTTKQVCRQLEHGVQVLFGPTDPLLAGHIQS
ICESFDLPHIETRIDLDTTIKEFSINLYPSQYHLNLAYRDLMVYLNWTKVAVIYEEDYGL
FKQQDLMYTTADLRTEMYIRQASPDTYRQILRAIRQKEIYKIIVDTNPINIKSFFRSILQ
LQMNDYRYHYMFTTFDLETFDLEDFKYNSVNITAFRLVDVQSQLYTDIVEQMQKFPHSGL
DIVEGHPYIQVPTSQPTRLFSFMNPLAMEIWLYVLAAYILVSFALFVMARFSPYEWSIPY
SCQKDSDIVENQFSISNSFWFITGTFLRQSSGLNPKRSDRAQTRFFSFMNPLAVEIWIYI
AFAYILVSLTIWIVARLSPIEWVPENPDVCDHEDCGDDASPEELIELQERPKRDPHNSNQ
VNSENPNSNQNDKNDNEKDDDKNAKQSSQLNDGHATWFSRKPFRHVIEDDVEQQQHHHHH
HSDSECGETELDFWFAIGALMQQGSDLYPRVGEKILNDWEMLLQFKKKNNKKKDATSTRI
VGGIWWFFTLIIISSYTANLAAFLTVERMITPIESASDLAEQTDISYGTLEGGSTMTFFR
DSKIGIYQKMWRYMENRKSSVFVKTYEEGIKRVMEGDYAFLMESTMLDYAVQRDCNLTQI
GGLLDSKGYGIATPKGSPWRDPMSLAILELQEKGIIQMLYDKWWKNTGDVCNRDDKSKES
KANALGVENIGGVFVVLLCGLALAVVVAILEFCYNSKKTVQTENQSLCSEMAEELRFAMH
CHTSKQRPSMKHNCAKCMPASTYVPQGPTTSASVTSGLAGIPHHLSGVQYNYLN
>GmmKaiR2-like-d2 GMOY012113.RB (TMP005606) scf-646489 14412: 22378 MW: 86957.61
MYVKREKISLIIRSITIKIVPVGLHIMFISELCQILKVLQQQCKQCKHLMMLMNCQSEQR
KKHKHRTPGVMTPIIYLVYLSVFISFRIGDVIALPPIIQLGAVFTEDQRSSNIESAFKYA
VYRINKDKNILPDTQLVYDIKYAPRDDTFRTTKQVCRQLEHGVQVLFGPTDPLLAGHIQS
ICESFDLPHIETRIDLDTTIKEFSINLYPSQYHLNLAYRDLMVYLNWTKVAVIYEEDYGL
FKQQDLMYTTADLRTEMYIRQASPDTYRQILRAIRQKEIYKIIVDTNPINIKSFFRSILQ
LQMNDYRYHYMFTTFDLETFDLEDFKYNSVNITAFRLVDVQSQLYTDIVEQMQKFPHSGL
DIVEGHPYIQVPTSQPTRLFSFMNPLAMEIWLYVLAAYILVSFALFVMARFSPYEWSIPY
SCQKDSDIVENQFSISNSFWFITGTFLRQSSGLNPKATSTRIVGGIWWFFTLIIISSYTA
NLAAFLTVERMITPIESASDLAEQTDISYGTLEGGSTMTFFRDSKIGIYQKMWRYMENRK
SSVFVKTYEEGIKRVMEGDYAFLMESTMLDYAVQRDCNLTQIGGLLDSKGYGIATPKGSP
WRDPMSLAILELQEKGIIQMLYDKWWKNTGDVCNRDDKSKESKANALGVENIGGVFVVLL
CGLALAVVVAILEFCYNSKKTVQTENQSLCSEMAEELRFAMHCHTSKQRPSMKHNCAKCM
PASTYVPQGPTTSASVTSGLAGIPHHLSGVQYNYLN
>GmmKaiR2-like-c GMOY004959 (TMP007054) scf-648346 2363:91226 forward
MRDSAPAKYNVLTLCNQIKTGVHAILGPSNGLLSCHINSICDALDIPHIEIRADAEHTAR
EFSINLHPFQDIINDAFLDIIHYLNWTKVAILHENQHGLVQLRRLMQTSLEVHVRYVNPL
TYVQYLNEMKEMEMHNLILDTKTDNVHIILKTILKLQMNEYKYHYFITSFDIEKFDLEDF
KYNFANITSFCLVDVNDVGVKKTLKELEIYKNLQKPNNEDLQFFRKLHTVETIPALIYDS
VHIFVIGLQSLQQSHFLRVANVSCESPQTWKGGLSLINYINSVEWKGLTGPIRFKEGRRT
QFKLDLVKLKQHSIVKIGEWSPHMRLNITEPSLFFDTGSMNVTLVVITILEKPYVMMRYG
KNYTGNNRFYGFCVDVLRLVARDVGFDYILDLVPDRKYGAKDPLTGQWNGMVEQLMKYKA
DLAVGSMTITYARESVIDFTKPFMNLGISILFKVPSTPSTKLFSFMNPLAFDVWLYVLAA

YFCVSITIHAMAKISAIERSNFNRSSKYFVKYLDKFTLRNSFWFAVGTLMQQSSNLKPRA
ISVRIVSAVWWFFSLIIIASYTANLAAFLTVERMTTPIENAEDLSSQTEISYGTLESGST
MTFFRDSLIETYRKMWRNMENIKKHSTFTSTYEEGIKRVKESNYAFLMESTMLDYVVQRD
CNLTQIGGLLDTKGYGIATPKGSPWRDKISLAILELQEKGDIQMLYDKWWKSVGETCFPR
SNTKQSKANALGFDNIGKLNKQVKMKYRQAESMFPEPDIPGICTTGETLSCGDITEEELK
NSAASGIPNEAVKNRTSIQAKKKDDWANQEILELFACHLQRLILPHNKKNAFADVLEDMV
VKNIIESKKPSFL
>GmmKaiR2 GMOY004222 (TMP006301) scf-648002 16220:27274 forward
MCTCLLTYLLVSYCLSQIQALPDIIKIGGLFHPVDDNQELAFRQAVERINEDRMILPRSK
LVAQIERISPFDSFHAGKRVCGLLNVGVAAIFGPQSSHTASHVQSICDNMEIPHLENRWE
YRLRRESCLVNLYPHPNILSKAYVDIVKYWGWRTFTIIYESNDGIVRLQELLKAHGQTPY
PITVRQLTDTGDYRPLLKQIKNSAEAHIVLDCSTDKIYEVLKQAQQIGMMSDYHSYLITS
LDLHTINLEEFRYGGTNITGFRLINEKIVGDVVRQWSIEDKGVLRSANLTTVKSETALMY
DAVHLFAKALHDLDTSQQIDIHPINCEGQSTWQHGFSLINYMKIVEMKGLTNVIKFDHQG
FRTDFVLDIVELSQSGIRKIGNWNSTSPEGINFTRTFSQKQQEIEANLKNKTLIVTTILS
NPYCMRKESVVSLSGNDQFEGYAVDLIHEISKSLGFHYKIQLVPDGSYGSYNKLNGEWNG
MIRELLEQRADLAIADLTITFEREQAVDFTTPFMNLGVSILYRKPVKQPPNLFSFLSPLS
LDVWIYMATAYLGVSVLLFILAKFTPYEWPAYTDPNGEKVENQFTLLNCMWFAIGSLMQQ
GCDFLPKALSTRMVAGMWWFFTLIMISSYTANLAAFLTVERMDSPIEGAEDLAKQTRIKY
GALKGGSTAAFFRDSKISTYQRMWSFMESARPSVFTSSNGEGVDRVAKGKGNYAFLMEST
SIEYVTERNCELTQVGGMLDTKSYGIATPPNSPYRTAINSVILKLQEEGKLHILKTKWWK
EKRGGGKCRVETSKSSSAANELGLANVGGVFVVLMGGMGVACVIAVCEFVWKSRKVAVEE
PVIVDQSYMQKTGENILAHFQKILNDAVHENLKYGGITFKYFTWTGIRLKKDTLAAITIM
DCDNTWKFFEDMKTPNVLVVAITNAECPRLPINQALM
>GmmmGluR-c GMOY010637 (TMP012848) scf-652153 433529:437851 MW: 115191.375
MATAFIRQQLVYKQQQPQTITNTKNSASLAQYTDNDKVSNSETTSVSANAYTGYRNVIDR
SADLQQGKQKVLMSSELNPYYGIDDSVEIFKYHKSPDNNNLSTTIGSTVMHRNSNAPENL
NSLKSHETMQKLETNDEKFNWQMNNSAMVTDNALPEPVKATEMKTNVTGIHVEASSQHAD
EIPWLSKKSQNPNINGRDLKPSLASMTMMMIKATKTESVNMQNTADYTMTERNFFTNGIK
NNNLHATTEHQDTLMEKVYSTDVSILESEKQNERNLSKTNNTSHVSTTVHAITKLFQGKM
GNAGELKASTEESGRNDKEKINRVVNKNERETNLTKSINIREDSVQHLCKEVSMQPLMVN
KSKSDGTVKPLLAIGENFKKTQKAEMRNIKGEPQIAFIKTSEQDTRNSSEGEKKRYHKVR
EVSSKGADILFDLESLIPVNGFEEPNRNCSENLFRQSDKNSHQITKATESHLRDRLQVNI
KDKRNFSEVNERKLLRKWRESPAEISKTFEMPLTTKRIAITKDGSTDEKQLQIDQIMTAQ
PTSTQRFSYNVEEQSKADTLQLLKNETLSTVRKIPKVFQQSRSNTTNSAERIVNVSEKLR
SDFTNRENLEIFLNSKFNSNNSRDRKLKISQQMDNYSKRGDCDLKQKLSKICLNDSLNEL
KSRNTVNSYSKTNKNDVKINIENYKNIPSLLLNKFLRPKRKKLIEAKLINFENIASTSII
TSLPYIQHSNITQMSSKYETNIENYVENIERNMTEYSINGTNTTETEEITEFVTNDIMIT
ESALKATNEISYNDLCCDGAKNHNNSLILEADGVAENDDVLTTTRTTQSVAKTQTPLIPE
TNINNAINNSRTSGAFHDMITITKAFATIPTLVSTDIISVLSENNNNNNNFQLLAETTKP
SITNLLTTMGTAITASIIATTTAATSLWPVKHAAVVEGEVILGGLMMVHSREDTITCGPI
MPQGGIQALEAMLYTLDQGNKNQLLPNVTLGAHILDDCDKDTYGLEMAVDFIKDFHAAAA
VEFE
>GmmmGluR-d GMOY010638 (TMP012849) scf-652153 547067:569030 MW: 60099.797
MSFGTQSLMNAYACLLLMTTFAAHGFAIAISDHQFIPTKIAVVGDIKTANKEETPNFDTE
MYSIANFSSESEEMAQQLSQQNLQETIKTKQTCHQYREQIELTHLHEQHECHRQQRRYLQ
EQTKPNKHRKWPHRHHHPNPHSGHHHHHHHHRNKLKLPSPELMNYQNDEQLMLKNSHLSK
KYQNLKQTTDYDFTDNFNSLSKPHNHNHHQQHHSTLMKTAKNNYFALTSRSGRHRGRVSH
THSNKSNKIFRSLPDNSEHSFEQQHFNQKHLMSSTIVGQRHWPVKREAVVEGDVILGGLM
MVHSREDTIMCGPIMPQGGIQALEAMLYTIDCINKVQLLPNITLGAHILDDCDKDSYGLE
MAVDFIKGSISNIDDAEYHCNKTQVRKVISGVVGAASSVTSIQVANLLRLFRIPQVSFFS
TSPELSNKQRFEYFSRTIPSDHYQVKAMVEIVKRLGWSYVSIIYEESNYGIKAFEELEEL
LARHSVCIAVKEKLVKDSGVAEEIAYDNIVQKLLTKPRARGK
>GmmmGluR-e GMOY010639 (TMP012850) scf-652153 582077:604521 MW: 93170.75
MRAVRRNNATGSFSWIGSDGWSARNLVSDGNEPEVEGTLSVQPQANPVRGFEEYFLNLTV
ENNQRNPWFVEFWEDRFQCRYPGSSSTPYNNYNRTCTTEERLSRENTDFEDQLQFVSDAV
MAFAYALRDMHRDLCGGRPSLCEAMKPTKGGDLLKYLRKVQFEGLSGDHFRFDNNGDGPA
RYNIIHFKQSIEGQYHWVKVGEYYEGELRLNMSAVQFKLLHPKPPESVCSLPCERGQAKK
YVEGESCCWHCFNCSTYQIRHPLDETQCLTCQLGTLPDIAKQKCQTIPEIYLRPESAWAI
GAMAFSSTGILITLFVIGVFVRHNDTPIVRASGRELSYILLAGILKCYAVTFALVLRPTN
VVCAIQRFGVGFCFTVVYAALLTKTNRIARIFKAGKQSAKRPSFISPKSQLILINGVWMV
IAPSHAMYHHPTREDNLLVCDSYIDASYMIAFFYPIVLIVICTVYAVLTRKIPEAFNESK
HIGFTMYTTCVIWLAFVPLYFGTANHVPLRITSMSVTISLSASVTIACLFSPKPYLSYLF
LPLYIILIRPDRNVRQSMMPPRYSNVQRTGGTGQTSIMAPAVVTAATCAQNDNIQKHITP
GQTEHSIKAKKLCEMATQTITIGSSILTNLDSQTYQQIPYADDPYVPNNNRVDEKFKQQT
DKVSDSNGLYSESNAGIAAVASGCIATSQTMLENNNKINQQTQDQQSNTVTNNGINVVMT
TPSGNYLSVDILQFSLLSTSAPPLSTERKSNASSCINQDHLVNNKQPALVEQNTTTINNN

NHHNNSFCDNNVNDSSNSSSILTNETIAKNSLSRESNPVKKSQRCAIPVALDII
>GmmmGluR-b GMOY005828 scf-648767 91607:102427 MW: 144849.88
MRGLLRYTALHLLSSLLLPVKTFVLIAFLIIIFTYFTSLTQAEIFTLGYLTGSQRSPGNL
DYQKPGITISGAISLAVNEINRGPLKDLGHSLDFLVAETYGDEINSIRQTASLWTEHAVA
YIGPQETCVHEGRMAAAFNLPMISYYCTHRDTSNKRDFPTFARTRPPDTQISKSVVSLLL
AFNWTQVQNNICDNQSLPFSYITFLYLNHADSLYQPVADTILTILNSAGIVIRDVKTWQT
IYHHGFKVNPFDELVEYTYTNTRIYLILGHYYEHVGLMVSLQRKGLLARGDYFVVGIDIE
QYDPSSPDKYLRGLLLERSDKLAEIAFQSYLGIYPSAPVSFAQFAKEVNKYMEQPPFNFP
NPLGFFGGTKKISAEAAYLYDAVHLYADALIKVLKAGGEARNGTAIIEAIKGVKYLSAMG
YHVYIDENGDAAGNYTVLARGLARNRRNKTVPGLVPVGTFRQTCSDKLPDLKLYGKIAWA
GAGRPEAEPKCGFTGEKCISIAGGALLILGIVSLVLYRNWRYEQELDSLLWKIDFREVQM
HENEKEKEANNAQKQTRSTHPLIRTSQVSLSSNPDADFRYTTIFTPIGLYKGQLYAIKKV
RKKCVDITREMKKELKLLRDTRHDNICAFIGACTEPPNICIITEYCTRGSLKDILENDDV
KLDNMFIASMVADIIRGVLYLHESPIRCHGSLCTSNCLVDSRWVVKLSDFGLFDFKKGIE
DNSTDLQSIAVKSIKLLYRAPELLRQGPASLVMGTQKGDSYSFGIVLYEIHVRHGPFGET
GLTPMKCLEKVLHVQDNLYPYRPSLQPLETSFDCVREILKECWSERPEERPDFKIIRAKL
RPLRKGMKPNIFDNMMAMMEKYANNLEALVDERTDQLQEEKKKTEALLLEMLPRPVAEQL
KKGHKVDPESYEQVSIYFSDIVGFTAMSAESTPLQVVDFLNDLYTCFDSIIGHYDVYKVE
TIGDAYMVVSGLPIRNGDLHAAEIASMSLHLLNAVSEFKIRHRPGNKLLLRIGIHTGPVC
AGVVGLKMPRYCLFGDTVNTASRMESNGVPLKIHCSRQCKELLEKLGGYYYQERGVIPIK
GKGEQRTYWLLGEDAEARKKRSYERSQRRGSKALNKYVQGTIKQEKEKQHQQQQQENHQG
NAKREREKSQKILGDEQLAQQHLSVANHLGIRSSLKNKNLPRNSMTRSSSLESPKKLRFA
AGNLLEHHRYHSDEALLEVITDTYKNAMRRSSASSTYSRYEESNLSCHSWEDFNNCKRQR
RPSSYPTANTPLLLNLNDT
>GmmmGluRA GMOY000333 scf-639555 91658:96144 MW: 147458.62
MAFFKYQKALRLNDLYILQKVFIILLIFFACTVNMENVSSKYSTGKWWHHPHSLALTNKK
IISVKSVIFGIVESAFAKDLSAAMLLVDARRPVLEIKFGESSNHREESVVTPKSVITSSA
LAQISQITVSPLNESGLCDEKFLKVRTTRDAKARKNDNQDINVETTTEQVVSEYHPTAYP
GNTWFALENQTKLNTITERFPGKFIPHQRSSPNKIKNKAKAADSLSIYLPGDIILGGLFP
VHEKGENSPCGEKVYNRGVQRLEAMLYAIDRINHDQNILPNITLGVHILDTCSRETYALN
QSLQFVRASLNQNLDTSAFECPNSEKPRLRKDVSAGPVFGVIGGSYSSVSLQVANLLRLF
HIPQISPASTAKTLSDKTRFDLFARTVPPDTFQSETLVDIVKALNWTYVSTIHSEGSYGE
YGIEAFHREALDRQVCIATAAKVPSNADEKTFDSIIQKLLSKSTARGVILFTRAEDARRI
LLAAKRAKLSQPFHWVASDGWGKQTRLLDGLEDIAEGSITVELQSEYIGDFDRYMKELTP
ETNTRNPWFNEYWQATFNCVLAEDLSNQNGMKRCDKRYRLSEKVGYQQESKIQFVIDAAY
AFAYALDNLRKDVCASTSYQNVYIAEHLQHGSSSSGWLRKTINSDSFSCHAMEVYDGKDF
YNNYLLNTSFIDLAGSQVKFDRRGDGLARYDVLNYQRLAPNSSSYYYKIVGRWFNSLELN
LNQVVWNGNVKQPASACSLPCGVGMIKKQQGDTCCWVCDKCESYEYVHDEFTCLDCGAGY
WPYRNKTSCYALPIEYLRWSSLFGLVPMIIAIFGICMTLTVITLFIQNNDTPLVRASSRE
LSYMLLAGILVCYSNTFVLIAKPTIYSCVLQRFGVGVGFSIIYSALLTKTNRIARIFHSA
SKSAIRLRFISPKTQVVITTSLIGVQILITLIWMMVEAPGTRFYYLNRTMVILKCKMQDN
SFLISQIYNMILITVCTLYAIKTRKIPENFNESKFIGFTMYTTCIIWLAFVPIYFGTGNS
YEIQTTTLSVAISLSASVALVCLYTPKVYILVFHPDKNVRKLTMNNHTYNRAPVANARPS
AESACWKVSSVITEYQTGTLKNTSINLLSTANDEESGRAPTDITTAKTTYVNECITSANK
YDVNRDKNNSQTSVTFNFNIVGTTYSLKASGTAWSCNDLDQRSSSQGEWNSLNAQKFNNV
GKNYIPDQTTLYDPTGKMNSYNLPAAAPIASTLTSRFSQVERDHKLFQKAEAGNKQNHLS
RYNHEFCLKFPSGSKNTNENCCEPTCSCHIPLQSSKIVAFSDKSETESEPDS
>GmmmGluR-a GMOY003230 scf-645661 95252:108946 MW: 116263.375
MKSSVNQQVLLYVNKDQQSTLHHQRRNCCVHPSHCRHHHHHQYRSTYQNHYYSGHHYYLN
MLPHLILAAIITLLVENATAYLPTEIASSATTVTTATAGSISSSSITSLNSIVKIGDGSM
ITREAIEQSLVTIHDIATENLGTLCISTLYRPLQVPINSERYESSRQKADLAASILQEVG
IVRHGGLSDALAKGLLSDEYTTGARILALNLTNGSVQSYVWWIKKNHEKMGEMLRFEEDG
LQIGKKPSPTYPWFADESTSPTLRSPKFAPSPPNIYYKGWWTFPYFSCSLSKWILSYSIA
IPPNGRHHGLRGFISIDIDVTGLRVNQCEAPVYRFNYQQMQSRRLHLADAQAIDAINDIQ
AFHKSHKCHRQTMMCDYRLPTAESPTITTSKILTTPYTWSRGAYQCLCKRGYYSIRHPDG
FNGTIMEIAWKEHQDNISNYYADVFTCLKCAPGCDACSGPEPCLADYNWPFRISLLTISI
GCATGTILLAGYLFHHRKVKVFKVASPIFLLITLIGCAIMYLEMVAIFPFLDTSWCIVTK
WTRHMGFCITYTSLLMKTWRVSLTYRVKSAHKVKLTDQQLLQWMVPILLVMLIYLGTWTI
SDTPFAEFIHDQNGLKFRQCSYNWWDHSLAIGEVFFLAWGIRVCYNVRNAESLYNEARLI
SYAIYNIAIVNITMAIIHLFIFPEAGPDIKYTLGFVRTQLSTTTTIALVFGPKISRVFKG
QGDKWDQKAKVRSITASFSLNGVGLVPEESPDLYQENEELKEQIQKLAHQIEFMKTVHMQ
INNRHLKPKPGGYFTITSTSFQAPFSKSNMSSTQTQTSKTEETVSGSASKGDNNVFDDNG
QELFKQFKEIFIPEDDINSWYYALNASSDDDDSDDHSSIAHIRLEQLMAELHDVPNSSQS
STKTVTTALTGSENSDILTIYEDITTNSPTSSSSTYAIQTPSPSLEFLLKRSSTQFQFQT
QTISDDDAMTLTEIVEIHRAPSIALIPFNNDHLLAIEMDMESQLNNNTSCSLSSTLTDSK
TLDIRSPIVV

Appendix 3

Dataset S3: *Glossina morsitans morsitans* chemosensory-related proteins – OBPs (32 peptides), CSPs (5 peptides) and CD36-like (15 peptides)

>GmmOBP1
MKTTAVILLALFALVSADYKLRNQEDLNKARKECMEAKKVTPELVEKYKKFDFPDDEITR
CYIECIFDKFQLFDSQTGFKNDNLIAQLGQSKDNKDEVKADIEKCADKNTEKSDSCTWAF
RGFKCFISKNLPLVMESLKKN
>GmmOBP2A
MFSKHVTLLMLLMSSSYSLNENWKRPTPQTVVQVAQKCLRLQTGLNIETLHDDPKQVRCF
FENLSLWDKYNGFKAERLAHVFNKRQMMNEILVAVNYCNDKARQDDANKWAFEAYSCFAM
GPLGNWTNLFIKNAYKKVLKEKGL
>GmmOBP2B
MKTIIVIVFLVTLATVWGHHHHEHHDDDDYVVKTREDLFKYRDECSNKLNVPADLLEKYK
KWQYPDDEVTKCYMKCMFEHFGFFNEKQGFDVHKIHKQLMGAHGTVDHSDETHEKIAKCA
DKKPEDTDPCAWAYRGGVCFINSNLQLVKSSVN
>GmmOBP3
MMEITFNTVSREHCQGKETINKNFTTKMNKFFLFAIILHIYFIEKQQVLSISLEEAHNSE
ILRKCFEEIDQSQYNSSEVLLEKFKNYAYWSHEEIPCFARCIASEKGWFDIDLSRWNKQR
LVDELGANMYNYCRFELNRAFKNVCSFAFKGLKCLKQAEMNVIITHNNLLECVKEKSISM
DQLLEYYHFPQLEHIPCLFKCFADKSHLYTVNYEWNVLNWLKAFGPIRNENADISICRVN
ANEREKMDICAIMYEEYNCWERLNYNTDGISVTYKKALKKIFNF
>GmmOBP4 scf-650660 238203: 238721 MW: 20000.766
MKSKIIFDTRVRRKVMFRVTLILLAIVTPALFSENRYMEFLADFKHCKRERGVGRFELDR
LRVGNLAYPSYEAKCFLGCLYERTGILKNGVLQNDVLKKNVGYIANRVLLDEVLPPCYAV
SGTNKCDIAFELKKCFKNVGFDKVWITVPWEDNTDPQYIAAMKLIDDLANVKY
>GmmOBP5A
MRFHIILKLMSWMCLMRTIESKSVIDLLEGKIYAPAQYQLKPADNFASSPVNKRQMPTSE
IPKNMQQFQDTLNEAKFKCARAMRLDSNKLLMYEDQPSLREKCLMACILKRMKLMDSDYK
LSVPTISHIAGMISDENPLLISVAAATASNCNNAINAREPCEAANQINKCIANELKAHKL
NLIY
>GmmOBP5B (=Dmel obp19b)
MMKYFEIFVVFALFSIMLVVTNAEDDDENEIGMTLDELADALESFAEDCEPKPERDHIKQ
LLTNDENPHENSKCFRRCLMEQFELIDEGQSQMNKDKVVDMMSMMYADNKETLEEIVDHC
NTKNGGTTEKCENAHQHGMCILNQLKEKGFKVPEVKE
>GmmOBP6
MFKLLLVTVLMLGILSVEAEIDVQEEIAKFILLANECREEVGAKEADIQDLIHKHPSAGQ
EGKCLRACLMKKYEVLDANGKLVKSVALEHAKKFTNSDENKLKIAGTIIDMCSAMDTVGD
TCEAAEQYSECFKKQADTYGITLEI
>GmmOBP7
MKLITVIVFSIDFLLFIDASPSGVQEGIVLHQCLAPFGGYTLENDQRLQRFKQWSDTYEE
FPCFTNCYLNNMFNIYNETQGFNEENVIKRFGRSVYNACKEKLIQGNNSCEIAYNGFHCL
INREDDPFILIDNIEDISMEAKRAMKECLHKFNTDEWQYLSDYVRFPVQEPIPCYTRCFV
YKMQLYNHRLRSWNIAAMQRLLGVPAEHANIENCLSLSKRRNNNMCAWIYKEMTCFSLSQ
>GmmOBP8A
MKKYHIYIVTFAITLLMSFGLNNAQKPRRDENYPPPDFLKSFKIIHDVCVEKTGATEEAI
KEFSDGEIHEDPALKCYMNCLFHEVNVVDDAGELHFEKLVRMIPEPFLEMVKHIIDACES
HIPKGETQCDRAWSWHVCFKQTDPVLYFLP
>GmmOBP8B - 94730: 98758 MW: 30064.756
MLLKRDWFLLIFLTKLSLIRIHSAPLRRDSNYPPKELLTGAKPLHEKCVKETGVTEEAIR
EFSDGEVHEDEALKCYMNCFFHELGAVDDKGDVHLETLNLIMPGSFVEAILKPAQHCIHP
EGDTLCHKAWWFHQCWKKADPEYPPKELITMAKPFHEACVRHTGVTEEAIKEFSEGNIHE
DEALKCYMNCFFHELGLVDDKGDVHLETLHQSMPGSFVDLILKPAQHCVHPEGDTLCHKA
WWFHQCWKKADPVVSNLMEET
>GmmOBP9
MTLSGKYRFLTVYTMLIVLLSAWTRAQQPRRDDEYPPPAILKLAKPFHDICVEQTGVKEE
AIKEFSDGEIHEDEALKCYMNCLFHEFDVVDDNGDVHLEKLFSKIPAALRDLLMEASKGC
VHPEGDTLCHKAWWFHQCWKKADPVHYFLV
>GmmOBP10
MNCFFHELGLVDDKGDVHLETLHQSMPGSFVDLILKPAQHCVHPEGDTLCHKAWWFHQCW
KKADPVVKLFLFFSIIAALLSVIKILKFCCCSFLYDISFTMNAKFLHARPRLIFIREKLQ
LCFSFVFLYTSLEGEPICQAEKWSKVNAQVAA
>GmmOBP11

MKLFLIILSTTVALVFAKFDIRTKDDALKAHEECHEEFQVPDDIYEQYLDYQFPEHKLTN
CYVKCWVEKMGIFTENRGFNEKNIVAQYTYENFKNLESVRHGLEKCIDHNEWETDVCTWA
NRVFSCWLKVNRHVVRKMFT
>GmmOBP13
MKFICIVCLSLACYVALVFSSTKDDFEKILQSCREDMQINENDLRTLSASPNDVSEGVKC
YMKCVMEKQGHFKNGALLEEAVIKSLESSPADHNDQNQMSAIVKECKKEIGSNECETAFK
VSMCLREHKVDFEI
>GmmOBP14
MFGKNSITLIIVCLVFVCLLKSSNVYGGATEEQMRSAANLMRDVCLPKFPKVSKETADGI
RNGNLPDNKDAKCYINCVMEMMQTMKKGKFLYEGALKQVDLLMPDSYKEEYRPGLAKCKD
SANGIKNNCDAAYAVLSCLRAEITQFVFP
>GmmOBP15
MQVLQNNYICKTLQNYVKTCVIEENISTKDLKLFMAWNFSNISNEGKCFFSCFHEKIGLT
INGVLQKKIAFGHLKRIFDRETAEFVLGECVNLVGKDKCETAYQFEKCLFNIEYNRLAK
>GmmOBP16
MPLLDLKETESFWELCQQSHNITDEEFENFNAFQSIDMEPDRKFKCYAHCLLSNLKYLNT
FSGKFDIEDFKQQDGIEDEDVAVIAKCKKLYDNINDPCEYGFNILQCILMFEPTE
>GmmOBP17 GMOY007314
MKSWILVLLTVGAVTIIDGSNDKAAADNKVILLYHKRACLEMEGLSEEVFPGDDVHEIFA
TMFQLESEVVPYETKCFLRCWLKRIQVMGDHLTMLKKKMNPDGTCERAARAASRGDECEF
AFLYQKCDHLLDVNEFDY
>GmmOBP18
MTRIFKGIFVLLGSLLLVAAQNKQNDDKMKKQALRLHNYCTKKVDSNTEYLLAALYNKTE
HTESFKCYLQCIFDSLGLVDSNNQVNLEKLINFAPTEIHEHILELHRACDTQPGKDSCDI
VYTTSQCYYELKPASREYIEYMMH
>GmmOBP19
MLERYNAEIDEETQALERECLKEENIPGRFNLPNYSLTDHFLGQIEYAEIAPKNAKCFLR
CWYKKMGILKENLVTSAGPIPELRQHMRECNEVATEWAQNQSNGDECEFAWSFYTCMHES
LVKCLT
>GmmOBP20
MLFTLFLIVFIFSSREASALNETSRFVLKEPNVRFAQMRCAEKYPDARPFPNYPDTPANH
CYVYCLFYKLGLIDLRSRDLDVQKKLQDVCEDFGFEIVKSLPKSLSGRCQDYYKILVECK
HKYRDLFEFIFNERSPKTTNEIGESATEICANGLYAADNVDFLVKPVLVEQLKLVCIFAN
FHYLDAYQRVDVEEIMISYDEAQALNQHTREIIEDCAHRANALYRINDYGDMALTLNTCL
RRESSDYSKVFALRDKNSRKY
>GmmOBP21 - 14269: 14825 MW: 17497.0
MKGINFDFLINDRDL+DEDWQPKTVADIKSIRNECLKEHPLSNEQITKMKNFEFPDEEEV
RQYLLCTALKMEVFCAHQGYHPNRIAKQFKMDMNEEEVLEIAEKCHDSNPDNSSVDVWAF
RGHKCMMSSAIGDKVKAYIKKRQEENAAKNA
>GmmOBP22
MRKNPKKFSLSNQSKISDDFFQMSERCMRLEKVPDRYKAQFTEFQFPNDPIVHKYILCVN
RELQIWDNNQGFDIEKIYQQYKGRANEEVVLPIISQCNQDAKQRNYELWCYKAFLCILDT
QVGEWFKEDVRRQQTRTLTNGHQ
>GmmOBP23
MAIFIFVTMKAADTVDFDDVIEECNSSFSIPTDYLTSFNSTGSLPDVTDKTGMCFLRCFY
EKSGFIKNWKLIDAKIRKYMWPATGDSIEICEQEKSKEPNSCVRLYAIIKCLMLRAIVDA
RNKPV
>GmmOBP24
MKFLINLCLLITVIELSIVRSEELTKENALAAANDCKDETGATDDDIAAILEHKPADSTE
GKCLRSCVMKKFGIMNDDGKLVKEKALEVAQIMITDDDKKELAIEVVEACENLSVNEDHC
EAAIEYGACLKEHAEEQGLLPPDF
>GmmOBP25
MKVFIIVMVLMVIGSTSAFRIIHKTFTKLFECQKREQVPIEAWFAANLPDSDSTKLDPNH
QCGVYCQNEAFGLTTNGILNPEAILRFLPELEKTYNVEEMVKHCRHAGASNNCEGALKLS
ICFENYRLPETSLQ
>GmmOBP26 LUSH homolog
MMRNGCAPKFKLTTEQIDGLRVGNFDENNKDLKACFIIVYITQLAGTLTKKGELSAQKAL
AQIPMILPVEMQKVALASLEHCKDIQKNYKDPCDRLFFTTKCVYEYAPDDFTFP
>GmmOBP27
MTALLNILLTSVSVALRCRTDDGPSEAELKRVTRNCMRKLSENYMQTTLNGHQQNQHAQQ
HNKEDENANSRTNADNGGDYLNYNQRNQFQSHENYNNNNNNNSTRYNNRYNNKSKNSNNN
NSTDSFCLAHCFFEQMQMLNNRRYPDQHKVLYVLTKDIRDRELRNFYTDTIHQCFHYLES
QHRRDKCQFSRDLINCMTEYAKGNCDDWNDI
>GmmOBP28 GMOY012237 scf 650245 25814..26230
MKYWVLTLIAVELIITFSRKNMRAFAEDWQPSLLALRTKCIEKESLSTHLYTGDNLGEVF
DTMLHQKSDQVSRHSKCFLRCWLKETRSILDNFSINAERYDAVDRYCERDAKVQANSDEC
EFAFLLLKCERAPYVETP

>GmmOBP29
MEKPLHRSVQGVQYVEIKGIKAFKILKNEVAQAQAPMNDHDDDDDDDGGLVYMETLDSVM
KNDDKFWPHDMNEEYVEGGGGEDVQDNFGYISYQDAPHRRMARLIQRIGRVDDISNGIYH
PTLVPFEDKRIAGCLLHCVYAKNNAIDKMGWPTLDGLVDFYSEGVNEHGFFMATLRSVNL
CLRAVTNKYHVDRHKLPEKGESCDLAFDVFDCISDQITGYCMDHYKP

>GmmSNMP1 - 103427: 115278 MW: 61002.797
MYISVTHSNMTKNIYLWIQRSVVIAFGSIFISAGIYLYLNWIDIFTRARGKQMFLGPESP
AFSGWKTPPFSLNFDVYLFNWTNPEDFHRNSNKKPHFVELGPYRFVEKLEKVDIVWHTNN
HSVSYRKKSLYHFDPENSKGSLSDKVTSVNVVAHSIALKFKDDSNFQKMVIARTLKMYNA
GVSITKTADEWLFTGYVDPFLSLGNLLSKFNKDMKIPYDRVGYLYTRNNSATYDGHFNVF
TGADDIRKMGQIHTWNYKKHSETFGGECGQVTGSMGEFFPPNLTPQDTLWLFVPNICRTV
SFHYADSVKIHNVNAYKYSAGERLLDNGTLFPSNKCFCIGGKCERSGVFNIGPCAYNASM
YISLPHFYKADPYYLDAIEGLRPVKEKHEFFMTVQPNLAVPMDVGGGLQGNYLLEPIEHL
PPFDRIQRAFMPLMWAEERVRVPPEIAESISLVPLIILIGHIFTGMLLALGVILVCWYPA
KLITSNYLCSKQKVGFLKRFTSNTNMKTTSLPRKSSSDLENTDLLQKNKITIIGS*
>GmmSNMP2 -scf-648879 - 215595: 251498 MW: 60102.133
MQIFPFHFQNVIITDGSEQYKRFVQLPQPLSFKVYVFNVTNSHKIQLGRNECHLKNILFE
FLRQFRTKRVQHFSRDGSKITYVQDQLYIFDEEASAPLRESDNIVVLNMHMNAFLQVFEK
EITDILQGFANRINHRLNRTPGVRVLKRLMDQPFKWVHNIFLSVYDNVKSVLEISENDPS
LAILLVHLNANLKGIFNDPKSMFVSTTVKNYLFDGVRFCVNPQGLAKAICNQIKESSSKT
LRELKDGSLAFSFFHHKNGSGQELFEVHTGKGDAMKVMQIQKLDDSHNLQIWLNASENNE
ASMCNQINGTDASMFPPFRQPGDNMYIFSTDICRSVQLFNQHPIEYKGIPGYRYSIGENF
VNDIGPEHDNDCFCVDKLTNVIKRKNGCLYAGALDLTTCLEAPVILTLPHMLGASNEYTS
TVRGLKPDAKKHQTYVDVQQVKEQTEKKFLLTGTPLQGGKRVQFNMFLKTINRITITENL
TTVLMPAIWIDEVKSNCICLFKICIIFFIICLLNIIEYFQFVV
>GmmCD36-3 (Croquermort-like acting on neural tissue on most rhodopsin in Dmel) scf-648975 14277:26128 reverse
MYISVTHSNMTKNIYLWIQRSVVIAFGSIFISAGIYLYLNWIDIFTRARGKQMFLGPESP
AFSGWKTPPFSLNFDVYLFNWTNPEDFHRNSNKKPHFVELGPYRFVEKLEKVDIVWHTNN
HSVSYRKKSLYHFDPENSKGSLSDKVTSVNVVAHSIALKFKDDSNFQKMVIARTLKMYNA
GVSITKTADEWLFTGYVDPFLSLGNLLSKFNKDMKIPYDRVGYLYTRNNSATYDGHFNVF
TGADDIRKMGQIHTWNYKKHSETFGGECGQVTGSMGEFFPPNLTPQDTLWLFVPNICRTV
SFHYADSVKIHNVNAYKYSAGERLLDNGTLFPSNKCFCIGGKCERSGVFNIGPCAYNASM
YISLPHFYKADPYYLDAIEGLRPVKEKHEFFMTVQPNLAVPMDVGGGLQGNYLLEPIEHL
PPFDRIQRAFMPLMWAEERVRVPPEIAESISLVPLIILIGHIFTGMLLALGVILVCWYPA
KLITSNYLCSKQKVGFLKRFTSNTNMKTTSLPRKSSSDLENTDLLQKNKITIIGS*
>GmmCD36-4 (Croquermort-like, scavenger receptor class B) scf-648975 38840:47668 reverse
MRVRQAYLKPGTQLYDIWKQLPIPVTLDVYLFNWTNPEEFHRGSKGKPRFVELGPYRFIE
KPDKVDIVWHTSNHSVSFRKKSIFHFDHENSKGNLSDRITSVNTVALTIALKFKDDNNFQ
KMLAARALKMYNAGISITKTADQWLFTGYTDPFLTIGSLLSKFNKYIKIPYDRIGYLYGR
NNSAAYEGYFNMFTGADDIRKMGQLHSWNYKEHNAVFEGECGQVKGSAGDFYPPNLTPQD
TLWAYIPNLCQAIPLDYTESVQIHDLTGYKYSGGEKLLDNGTLFPSNKCFCVGGKCERSG
VFNVGPCVYNASMYISFPHFYKADPYYLDAIEGLKPEREKHEFFLTVEPSSGIPLEVGGG
FQINYLLEPIKFLPPFDHIPRAFIPTIWIEKRLKLTQEFTALISPVPLITWIGHLTSGVL
SILGVTILCWYPVRYFIVSYMRPKQKIHFLSRASSKTHLTKNSLHMESLKSEEIMDMLQQ
KPPNSELIS*
>GmmCD36-5 (lysosome membrane protein-like,scavenger receptor class B) scf-648975 53123:60891
reverse
MQGKEKEIMAEENSSSLLRNLKTFLKYLNILLSISLAGVGTILTIFSSTLFDKYLENELQ
LKPNNSVTKTWIKPDVNISLDVYLFNWTNSHEFLDPRIKPQFQQVGPYGYDEVPFKSILK
WHSDDNTLEYHKLHTFYFNETRTKGSLQDRITSINALLVRNGTSVLSPTYRISTGADGVN
SYGQLKFVNGRNHTKHVPSGCSQVRGSSGELQRVNLVKYEPIEYYVADFCRRFWLEYDDE
VMVDGVLGYRYKMGNMILDNGTMYSENKCFCNGKCLPAGVVNITSCAWDMPFFISLPHFL
NADDYFVSRVEGLQAKRELHEPFIILEPRTGLLMEFRGRFQLNIYLEPSSHLNCFPNKRE
LLFPIVWFDAYMHMSVKTAFLLRFIQNFNFYSQLLGMFLITVTMTTLLWRPVKKCCFIHY
PQHLEISAMQDDDDDDENIKKNYTEDLANKEALLGESSNFIGLFGESKKIYS*
>GmmCD36-6 (Croquermort-like, Ceratitis capitata) scf-648975 62531:67482 reverse
MPQKSTNMKIWTKSRKRTLVSALGFLLSIFAILCAMFWERVFDSIMAKEMVLRPNSQVFQ
KWKNPPLSLNLDIYLFNWTNPADFRNLSTKPILEQCGPYRFVEKPDKVDIRWHPENSSLT
YRRKSFFYFDVNGSNGSLSDEIITVNPVALSAAGKGKQWDPVRRKMVDVGLNLYQQKMSV
KRTVDELLFTGYSDDMLDMARAMPLFGKDVEVPFDRFGWFYTRNGSADLTGVFNVYTGQQ
DIKKLGQMFSWNYKRHLGFFESYCGFANGSAGEFQPPNLTPKSIVKLFTPDMCRTIPLDY
KETQIAEGIKGYKYAGGHRSIDNGSLYPENKCFCGGNCVPSGVMNISSCRFGSPVFMSYP
HFYQADDFYLNQVEGLQPEADKHEFYMVLEPKTGISLEVAARFQVNMLVEPIRGISLYEN
IPRVFFPMIWFEQKVRITPDLAKNLKTLPYVLLGGQIIAGLLFLIGLILLCWYPIKYVWS
THKIQEIKVYPAENGKAINSKQSELKILETEKKTPDSSPLLEKNSQLKPAITPKSPKRDT
DANDVNHK*

>GmmCD36-7 (Croquermort-like, Ceratitis capitata) scf-648975 62531:63990 reverse
MYLFLLFSLLIYYGLKSAAGKGKQWDPVRRKMVDVGLNLYQQKMSVKRTVDELLFTGYSD
DMLDMARAMPLFGKDVEVPFDRFGWFYTRNGSADLTGVFNVYTGQQDIKKLGQMFSWNYK
RHLGFFESYCGFANGSAGEFQPPNLTPKSIVKLFTPDMCRTIPLDYKETQIAEGIKGYKY
AGGHRSIDNGSLYPENKCFCGGNCVPSGVMNISSCRFGSPVFMSYPHFYQADDFYLNQVE
GLQPEADKHEFYMVLEPKTGISLEVAARFQVNMLVEPIRGISLYENIPRVFFPMIWFEQK
VRITPDLAKNLKTLPYVLLGGQIIAGLLFLIGLILLCWYPIKYVWSTHKIQEIKVYPAEN
GKAINSKQSELKILETEKKTPDSSPLLEKNSQLKPAITPKSPKRDTDANDVNHK*
>GmmCD36-8 (Croquermort-like - NINAD, Dmel) TMP007264-RA:CDS scf648454 19001:20785 reverse
MCCECCGVTQRKAWVFSSAVILAIFGILLVVMWPEWSVSLVHNNLLIKVGTDNYNSWVKA
PIPIYLVFYLFNWTNPEDITKANIKPNFVEMGPYVFLEKHSKENLSFYSNDTVSYYQRRT
WFFVPEKSNGTLEDMITTAHPITATVADQMRYKNKIIKKVLNFMLNHEGGNLYVTRPVKE
WIFDGFQDELIDFLSLFNTSKINIPYKRFGWMVERNASLDYDGLFTIHTGVDNIHNLGQL
MHWNGKNTSDFYSPPCNTIKGTPGDLFPPELDSQEPITVFVTDVCRYLNLKPNGSTELYG
LQAITWEGTNATLDSGAYYPEQECFCDASIDECPRGGVADCKKCLHNAPIYASFPHFYLA
DEYYANAVTGMKPDPEKHKFTLAIEPHTGIPVEVKARIQINMMITPDDTFDVYRGVQHFL
MPMFWFEEIAILDEKLANKAKLALNLDSYGVIGGIILICLAVTLSIIGVVLTIMKKWKHI
PDDDETILTDNAENANNAENTQ*
>GmmCD36-10 (Scavenger receptor class B member 1-like) TMP013099-RA scf-652157 2092908:2101875 forward
MFRKYLLRSPRILIRKICDILRVENLFLNSEATAEPIRVTTFENGFTIAVSVLLLILGIL
VVCYFFTLINAVVDYQVALRPGGQTYGWWAKPPVEPKISVYVYNVTNANEFLSNGSKPIV
NEVGPYVYTESWEKVNIVKNENGTLSYNVRKIYVFREELSAGSDDDVVIVPNIPIFLRLA
MASIMDILKIKPFVQVSVGQLLWGYEDPLLKLAKDVVPKEQKLPYEEFGLMYGKNGTSPD
RVTIFTGVEDVTKFGIIDKYNGKSHLPHWLSEECNTLNGTDGSIFPPHIDEQRILYIYDK
DLCRLMPLHFEKEVETKGGVKGFRFSPPTNVFADVERNPDNMCFCPAGQPSCAPNGLFNV
SLCQYGKYKLMKKLAQAEITLVSFALRFADSPVMLSFPHFYLADDSLRTAIEGISPPEKE
KHQLFIDVQPVMGTTLRARARVQINLAVSQVFDIKQVANFPDIIFPILWFEEGIDYLPDE
ITDLMNFATTVPPKIRVILTIIFFAVGALLLLISVFCLIRNSHRQSTLHLEGNNYLATAQ
IDMQKKMAKERY*
>GmmCD36-11 (Scavenger receptor, Gmm) TMP013105 scf-652157 2268547:2272905 forward
MSLPSDKVTTEQQLTLTCDDCIDDDVAVQQSAEKQHQRQSQRQHYRKHQHQHQHQQQQQQ
QQHQQCMFNSNMNADKRPQKTENDLSLKIKYKRSESRNKKTQSLFELFFEMVGEKGQRTR
EMGTLILLGTMFLFFIISLTGFFVMWFTEYYNNIFLSNLVLSRNSETAEKWMNPNSKYDT
FLKVHIFNYTNIKDYLEGKAEKIEIKDLGPLTYKEHTTKVNVVFNDNYTVTFRDHRNYEF
LPDKSSYGEHEKIFVPNVPLLAADFLIDQMRGLKKMTASVAIKAIGGNAFKTLTPSQYLW
GYRDKISSLNFASGKSHFGLLMNRNGTSLDSLQINTGEDDLRKFGLVTQFNGMPLLDFWS
EEQCNRIDGSDPSMFPPHLIENRSTLNVFLQVLCRKIPLKFEKQVTIFNNIEALRYRTPM
NVFSHPSENSENECYCRNTQKCLPSGIINATKCYDNIPIYPSSPHFFAADPDIYKHLDGI
EPRQELHQTFADIHPRFGFPINGASRIQINIAVHKGSIVEQQLRRLRRDTILPLIWIEIT
TGDFTEDVIDTLYASTYGLNLIQCSLKYGTLLMCLIFFTLIVASFYYLAKKREIQLEKGE
KILKAELKALNRIHLSSASLAQMQS*
>GmmCD36-12 (lysosome membrane protein like) TMP004071-RA scf-642607 82697:90874 forward
MKINIFARSTSPSPNIKILLAGFLGIILMLLATTIFIIDPVKTITKSQMSFRKGSTLFNV
WQKPPLEVFITVYMFNVTNYDRFSSGVDDKMKLAEVGPYVYQEWLENHNATFHANNTVTF
VPKRTVKFVRERSVGDPKEDKIIVPNIPYLGATSAASSFSMITAMALRALTSKLQSEAML
DITVHDYLWGYEDRLVYLASKLIPQIINFEKFGLLERMFNEGHNVVNMHLPGSKISKDRQ
PKYYREFSINTWNNEKSIHYWTNSRKQNLTNCNRVYGTFDGSLFPHDIEKGEILRIYRKT
FCRTLPIVFTHPSTINGIEAYNFKLHEKSFDSDLSDGDSSCFCKNRKCLKKGLGNITPCY
YNIPLAISFPHFFNADPSLLKPFEGLEPNESKHGTEIALQPQLGIPMRVKSRFQVNLAMS
DVSFNGEVQRFRNMVLPIFWLQIGLDDLTPYLKTLLILTFKVGPIAQGSLVIILIATGLS
LIAYAIWKSLPETLTRYGTIKTIKFPLDRRVSVRTPDASIQPLLQDEKTIELQNWIDEDE
KGSVFTFTPVEDSGDTF*
>GmmCD36-13 (Scavenger receptor) TMP005916-RA scf-647496 34564:69209 forward
MGLQKQYLRYGRTARDRLLGWLGISTRTENRVDNLAQVNTSTPIIVAGRESVNASLNTTP
TTTTNNTNQPRRSNQTPHRNRTTPLSMLISQGAKLSNNRLAVIIIGIVTVILGILLTTIP
WLDYFILQNLRLWNNTLSYHYWQRPGVIRLTKVYIFNVTNPDGFLNGEKPRLQEVGPFVY
REDMQKVNVKFHDNFTVSYQHKKILEFVPELSIDKNTPIVTPNIPLLTLTSLSPKLGYVL
SKTISVILTAAKFKPFINVTADQLVFGYDDPLVALAHRFYPKHIRPMERMGLLLARNGTL
TEVSTIKTGHNGMNEFGYIERLNGLDYLPHWKQRPCTSITGSEGSFFPPRDITKSDIVFI
YDKDLCRVIPLKYSHGIKKDGIHADLYYLPEDSYGDSQHNLENRCFDARDYKPIKGLQNI
SPCQYGAPKLGVPLEGQVRIQLNLRVTQAKDVFAVRNFRTFNFPIMWLEEGVSELTPVIR
RWIYLATVFAPTIIPITSYLMILGGAFAIMYTFVRVYQNYVFARDPTLEILEMGRRSLRR
GSGFIAQHQHKLIHRDSYVLLKTGSNVLINDDVEQDDSS*
>GmmCD36-14 (Croquermort-like - NINAD, Dmel) TMP009115 scf-649782 11756:13500 reverse
MCCRCCGVLQKKIWLFGSATLFLILGIVLTVWGPGLADNFIDNMVVLKEGSTTFEKWTNI
PVPVYMQMYMFNWTNGKEVQESGVKPNFERVGPYVFRETDIKTNITWHENKTISFKPQRT
WHFEPKMSSGNLDDLVTAPHIPSLAASNFMRKKSSIMKRFFSSILNKNGGALYQTHTVNE

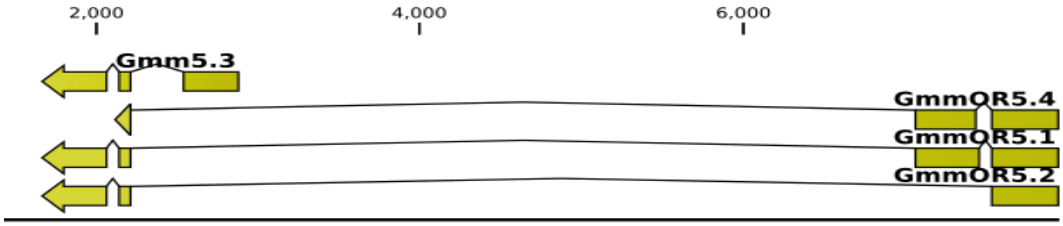
WLFDGFYDEFLAFAKKQNNSLIPPIPSDHFGFLLNRNGSSDYEGTFTIHTGQGNVREMGE
IKLWNGQNHTGYFSGECGRINGSTGELFAPKRDPNEYVTVFSRDTCRIINLMPIGTDTFR
GIEAIHYETQAETFDNGALNPDMKCYCQDPDNCHKTGASDISTCAEGVPMYISHVEFRDA
DPSYANSTTGHKPIDESDRFFIIMEPRLGIPLKMNVAIQVSLHVQPDKDITILQNINEFY
APLFVGKSSGEVDAKLAKKIKLLLNARPIAFYSGVASLVLSIILLLIGIYLSLTNRW*
>GmmCD36-15 (lysosome membrane protein-like,scavenger receptor class B) TMP008437 scf-648956 20927:40721 reverse
MISGHLSDMKRIIMGDEAWICAYYPETTDLSREYRAKCDRKNRVNGKKLIMRENNEMFDL
WAKPPVDLYIKIYLFNITNAEAFLAGREKMNVQQLGPYVYRELFTHENVTFNANDTMSTL
PRHPLVWQEHLSEGNKEDDPVVMLNIAMLAISHLTADRNIIVRLSLNSLFSTLQSEPIVR
MTAKEFMFGYNTKLTSLGNTFLPNWIYFDKVGIIDRMYDFDSDYETFYTGRSDPSLSGLY
ATYRGNTDLPNWPEKHCSNIETASDGSKFRSYIKPNDTLKFFRKSMCRPIHLVRADTDIV
TKCGLKGYRYRFEDNAFDNGRYNIQNKCFCRKGYCQPIGLLDVTDCYYGFPISLSLPHFM
DSDPGLRMNITGLQPDKEKHSSEFIIQPQSGLPLSLTAKVQINLHFKNMRAFRQLQAFSY
QTIPTLWFDITMPQLPDHMNILFSMYLNVLPYVEPIVFWSCSIIGFTLVFYAITRATLRM
SNLGHSTHISDGNRYGKANLLNSQNGVYKSCEMKQIEGKEKSRLLPDETYGHSNVNDEYC
EKEESNRTRSYILDLEPTLSASDCGSNTGSSSNDEGNDSDTATMTISRNSSSSSCMGMQV
SSKSNDHNNDDVEKQTIISFTSSSGYDTSITES*
>GmmCD36-9 (lysosome membrane protein-like,scavenger receptor class B) TMP013098-RA: scf-652157 2035294:2042370 reverse
MHTYAFAFFGKKHTLNPTWQEFQFNPFVLYSCLFCIIFPYFTPIAGKLISLTITAVVCSV
LFLASLHINYQWEFIKEHVRFRRNSPQQNEWTHSPHGMLRVYMFNVTNAESFLNGTDLRL
KIQQIGPIAYHVTGLNEILSQTSDSVTFRRNPHNIFEFDPSASSSPDILNQTIIMPNIIL
LSSAAKLHDWVFFVRHAFNAITINESAFLKETINYFLWDFTIPTLSLLAHYVPNIVSNCG
LLYNAIRPKELIYNVKIGVDNGIENFFRVNTFNNKTYFPQQRAFVKRAKKSDEYCPVILD
NSFDNSFFPPLLTRETELNIIATESCRTLKLNYDRDVVWQGFKGEVPTAFSAPHFMDTAY
NFTKHFEGLSPDKEKHEAEVILEPTMGIPLEEKYRFQVNIPLPDMKGFNKDLQRFSHMVI
PSFWYEYDLDDMSTLTTILMHISVHIVPNIQAIFMVIFLVLIVYSCLRIYLLLTNKTLRE
LLCATYKKK*

>GmmCSP1 - 209793: 211764 MW: 22410.896
MGANNQDYRRFFAYEGLILRCRSERYHITPLSSLDHASRWLMQYDLEKRRKQNCSAQVLL
QQTAVNMKYLTIVAVIATLSAVVVMGAEEKYTTKYDDVDVDEVLKSDRLFKNYYNCLIDQ
GKCTPDARELKKSLPDALQTECSKCSEKQKKTSEKVIKHLMDHKPEEWKVLQTKYDPEGI
YYSKYKARDAKA
>GmmCSP2a variant 1
MLISFVRIPQIDKKFLIINNMLRFFGICIVTAIIAWDSVRSLPHPPATTAAPFKQSYDNK
FDNVDLDEILGQERLLKNYVKCLEGTGPCTPDGKMLKETIPDAMATDCAKCTPKQKYGSE
KVTHFLIDNRPEDWERLEKIYDPAGTYRTAYLMGKGEKKKTNLAITTTERNNDSIPNA
>GmmCSP2b variant 2 - 59985: 60641 MW: 19846.854
MLISFVRIPQIDKKFLIINNMLRFFGICIVTAIIAWDSVRSLPHPPATTAAPFKQSYDNK
FDNVDLDEILGQERLLKNYVKCLEGTGPCTPDGKMLKGYANFRDNTGRYGDRLRQMYPET
EIRFGEGDAFLNRQSSRGLGAFGENLRSRRDISNRVLDGERGKEEDKLSNHYY
>GmmCSP3 - 200468: 200854 MW: 14628.069
MKFLTIVVAVVLMTVAVIAEEQYTTKFDNIDVDEILASDRLFDNYFKCLVDEGKCTPEGR
ELKKTLPDALETACAKCNDKQKATVDKVIRFLTEKKPDQWKALQAKYDPAGEYLKKYRSE
AEKRGIKV
>GmmCSP4 - 2032955: 2034988 MW: 14004.424
MKSTFCCLALLVVYLSVIVTAQKSYTNKFDGVDVDSVLSNERILTNYIKCLMEKGPCTPE
GRELKKLLPDALKSDCTKCTDVQKKNSQKVINYLRANRPGEWKLLLNKYDPSGDYRAKYE
KQA
>GmmCSP5
MKRIMISRVCVTCVIYLLMAAVVVECDEKNINKLLNNQVIVSRQIMCVLEKSPCDQLGRQ
LKAALPEVIVRNCRNCSPQQAQNAQKLTTFLQTKYPDVWAMLLRKYKT

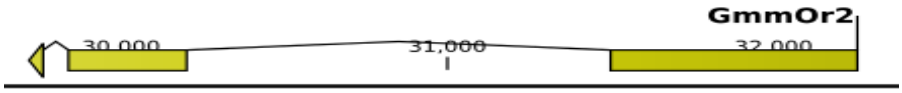
Appendix 4

Genomic structures of ORs and GRs in *G. m. morsitans*

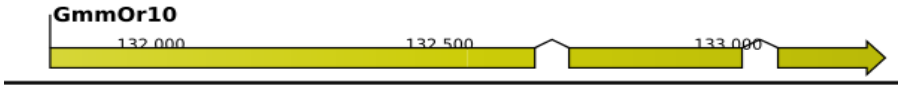
Scf-639717



scf-648756

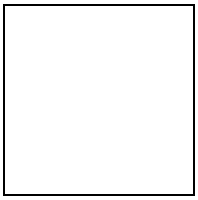


Scf-644232

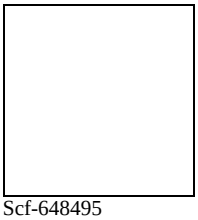


Scf-651846

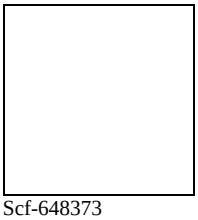
Scf-648410



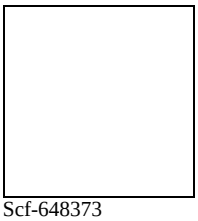
Scf-645812



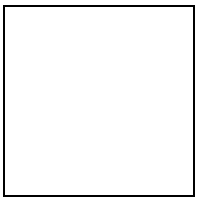
Scf-648495



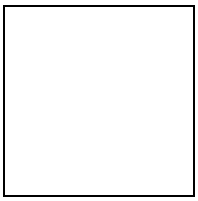
Scf-648373



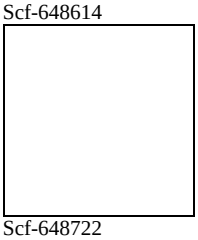
Scf-648373



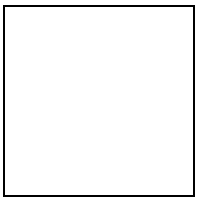
Scf-648080



Scf-648614

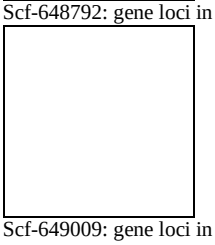


Scf-648722

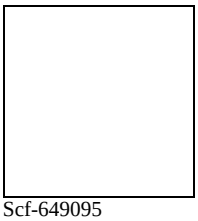


Scf-648722

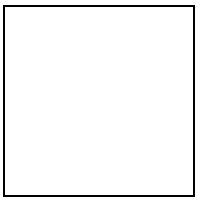
Scf-648792: gene loci in tandem



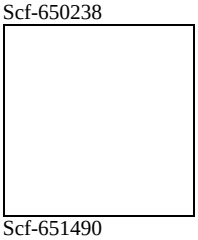
Scf-649009: gene loci in tandem



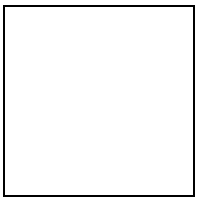
Scf-649095



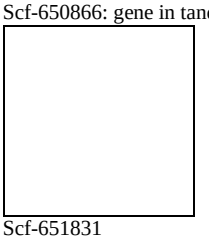
Scf-650238



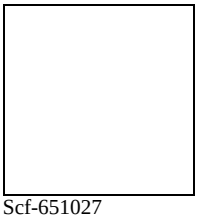
Scf-651490



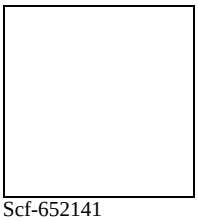
Scf-650866: gene in tandem; Or28 pseudo gene



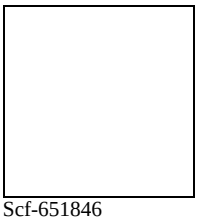
Scf-651831



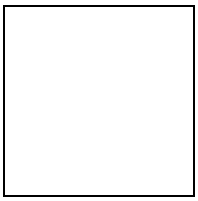
Scf-651027



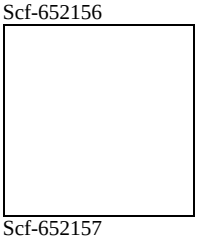
Scf-652141



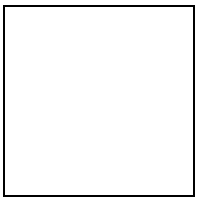
Scf-651846



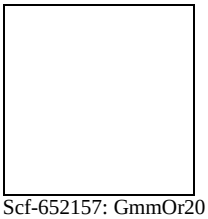
Scf-652156



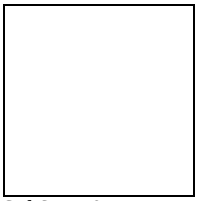
Scf-652157



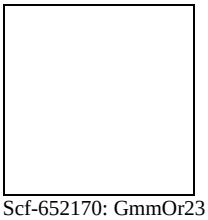
Scf-652170



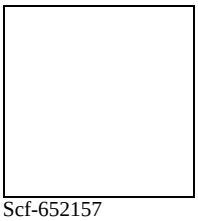
Scf-652157: GmmOr20 is a pseudo gene



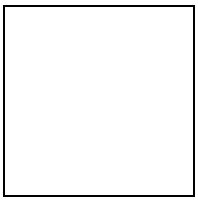
Scf-652170



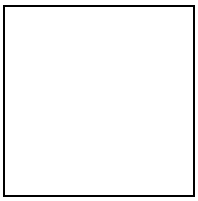
Scf-652170: GmmOr23 (likely pseudogenized because it has only 2 THMM)



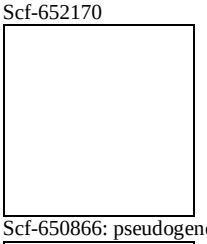
Scf-652157



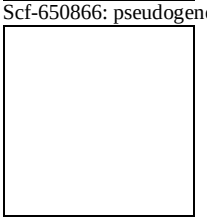
Scf-649048: genes in tandem



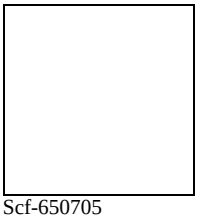
Scf-652170



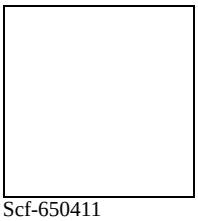
Scf-650866: pseudogene



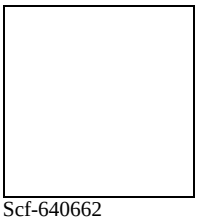
Scf-648928

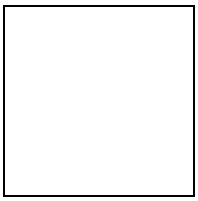


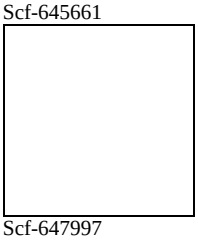
Scf-650705

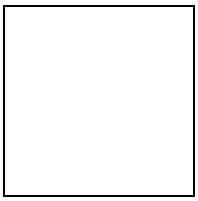


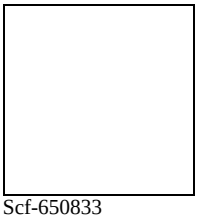
Scf-650411

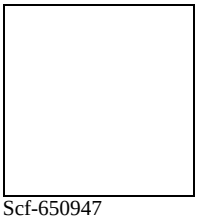

Scf-640662

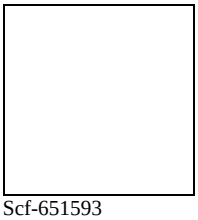

Scf-645661

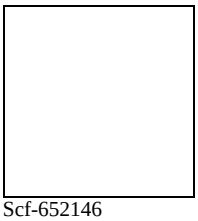

Scf-647997


Scf-648889

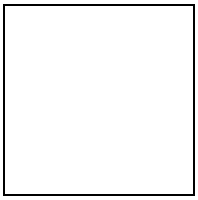

Scf-650833


Scf-650947

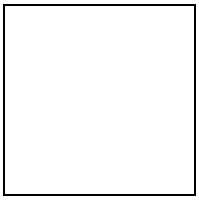

Scf-651593



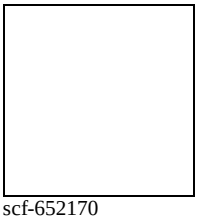
Scf-652146



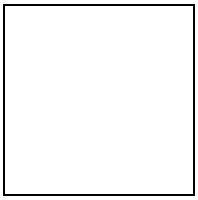
//



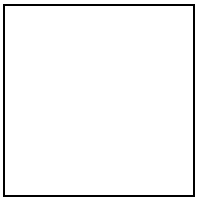
Scf-652170



scf-652170

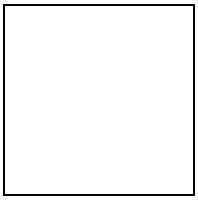


scf-652170



Appendix 5

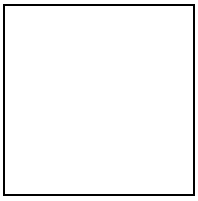
Conserved motif blocks in *G. m. morsitans* ORs



Appendix 6

Phylome database tree of *Glossina morsitans morsitans* OR41-46.

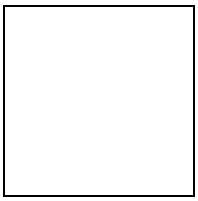
The expansion is probably by five duplication events – shown by red boxed internal nodes on a highlighted grey background. The blue nodes represent speciation events. The tree was searched using TMP010054 as seed sequence. All the five Glossina genes are homologs to single copy genes in each of the drosophilid species (*D. yakuba*, GE21738; *D. melanogaster*, OR67d; *D. pseudoobscura*, Q2LZK3; *D. mojavensis*, GI11463), and distantly related to highly expanded clusters in *T. castaneum* (all collapsed leaves) and mosquitoes. The tree was generated using Whelan and Goldman (WAG) amino acid model.



Appendix 7

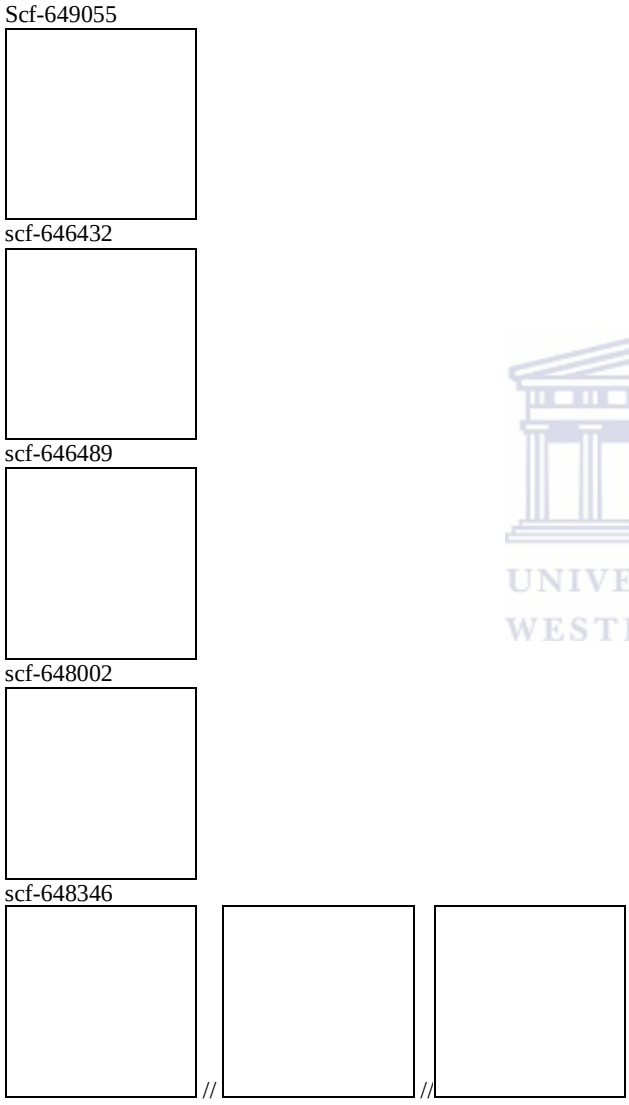
Conserved motifs in *G. m. morsitans* GRs

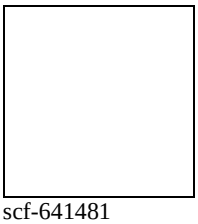
Motif 1 [RK][AQ][LK][WY]LD[VLM][KSD]E[LY][LT][QK]Q[LF][GN];
Motif 2 – QFY[RE]AL[KQ]PLLI[LI][LSF]xI[LY]G[VLC][TM]PIx[RIL][SQ]xPK;
Motif 3 – [NT][LA][AK]G[FYL]FN[IV][ND]REL[YL]F[GLT]

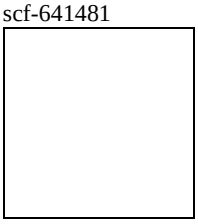


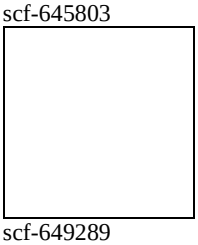
Appendix 8

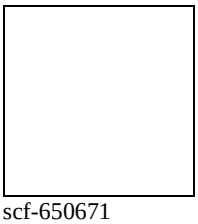
Gene and genome structures of *G. m. morsitans* glutamate-gated receptors
GmmGluRIA.a and GluRIIE.a are encoded in tandem on scaffold 7180000651742; Clumsy, GluRIIA and GluRIIB on scffold 7180000649055; and IR75a, 75b and 75c on scaffold 718000065067. Gene loci GmmKaiRIA had putative three splice variants, while GluRIIA, NMDAR2, KaiR2-lkd and IR8a were each annotated two splice variants. The clumsy gene was identified on the complementary strand. The horizontal bold type line represent the genome scaffold sequence with numerical sequence indices; yellow bars are the gene coding exons for each gene locus, with the terminal arrows indicating direction of coding strand; gene names annotated are indicated at the 5' ends of the genes. The annotations were supported by RNA-sequence data. The image was generated from CLC-Genomics workbench suite.

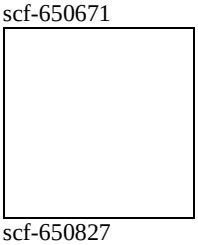


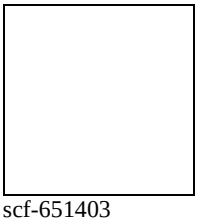

scf-641481

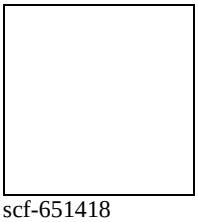

scf-645803

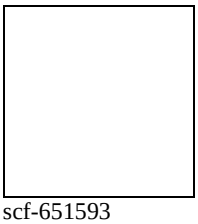

scf-649289


scf-650671

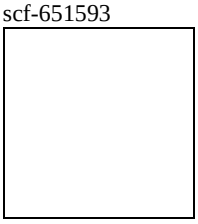

scf-650827


scf-651403

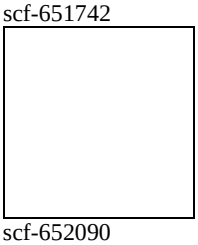

scf-651418



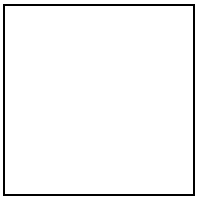
scf-651593



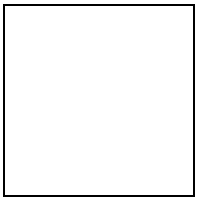
scf-651742



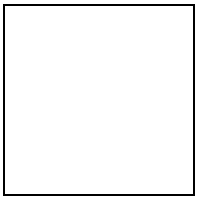
scf-652090



//



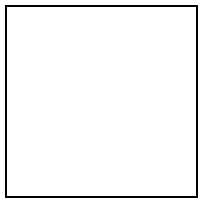
//



Appendix 9

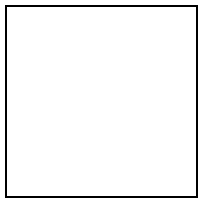
Conserved Glossina glutamate-gated ion receptors ligand-binding domain (LBD) sites.

Upper panel show LBD S1 ligand interacting residue R (column highlighted and marked with *); lower panel show LBD S2 ligand interacting residue T and D/E (columns marked with *). The green horizontal bars mark length of predicted beta sheet domain for the S1 and S2. Sequences were aligned using MUSCLE tool (Edgar, 2004), viewed using Jalview (Waterhouse *et al*., 2009), and secondary structure domain predictions done using Jpred program (Cole *et al*., 2008)



Appendix 10

Conserved Glossina glutamate-gated ion receptors trans-membrane domains (TMs).
A – trans-membrane domains TM1; B – TM4; and C - TM2 which is a re-entrant loop that form the heteromeric ion pocket, and TM3. Column marked '*' is a Q/R RNA editing site common in some receptors. The red horizontal bars mark length of predicted alpha domain for TM1-4. Sequences were aligned using MUSCLE tool (Edgar, 2004), viewed using Jalview (Waterhouse *et al.*, 2009), and secondary structure domain predictions done using Jpred program (Cole *et al.*, 2008).

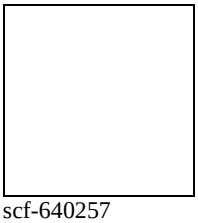


Appendix 11

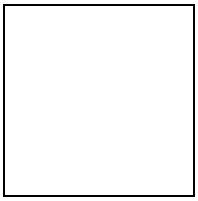
Gene and genome structures of *G. m. morsitans* OBPs, CSPs and CD36-like genes

The yellow bars show the gene exons linked, with arrow indicating direction of coding strand. The gene names are labeled from the 5' end. Thick black horizontal line represents the coding scaffold with numerical nucleotide indices shown with numbers. GmmOBP2A|2B and OBP5A|5B are encoded in tandem on their respective scaffolds. GmmCSP2 had detectable splice variants CSP2.1 and CSP2.2 in which similar internal exons were encoded on different open reading frames of the the same locus. GmmCD36-3, -4, -5 and -6 are arranged in tandem on scaffold 7180000648975, CD36-6 encodes putative splice variants. The GmmCD36-9, -10, and -11 are also in tandem on scaffold 7180000652157 and are flanked on both ends by chemosensory proteins (GmmCSP4 and CSP1).

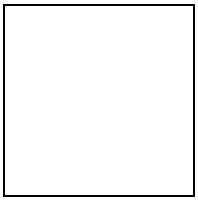
scf-639213



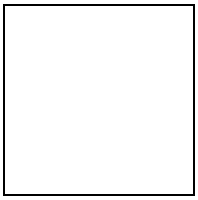
scf-640257



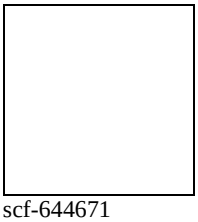
scf-640662



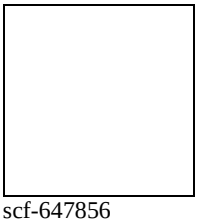
scf-641423



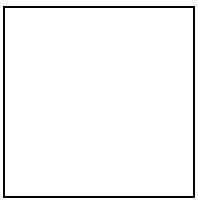
scf-644614



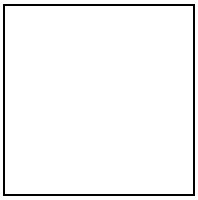
scf-644671



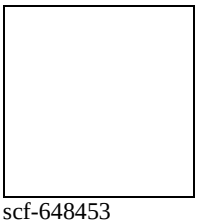
scf-647856



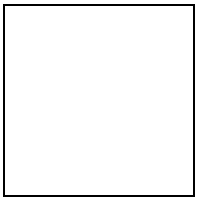
scf-648041



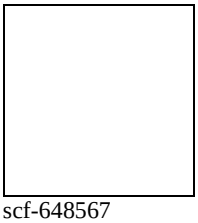
scf-648041



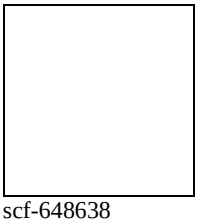
scf-648453



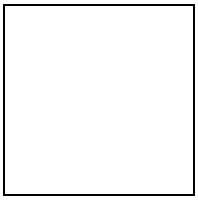
scf-648462



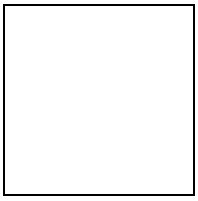
scf-648567



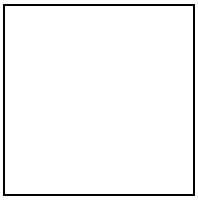
scf-648638



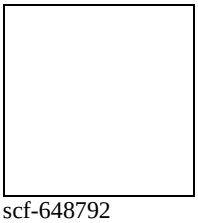
scf-648638



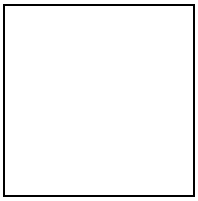
scf-648638



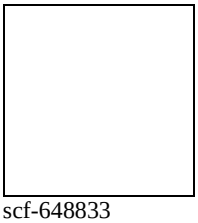
scf-648778



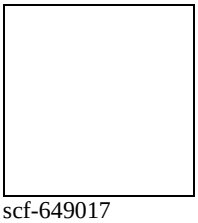
scf-648792



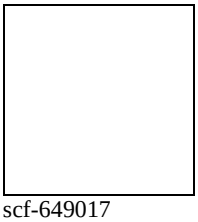
scf-648825



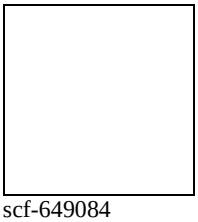
scf-648833



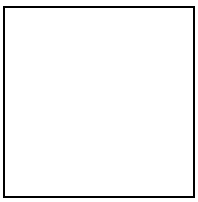
scf-649017



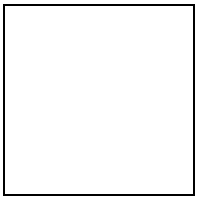
scf-649017



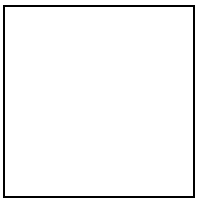
scf-649084



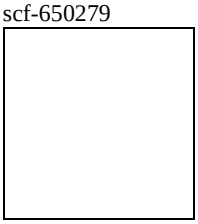
scf-649084



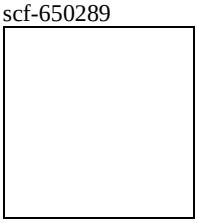
scf-650245



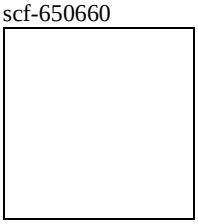
scf-650279



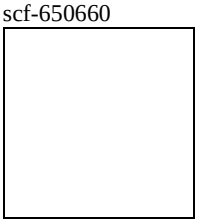
scf-650289



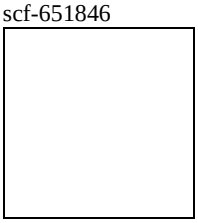
scf-650660



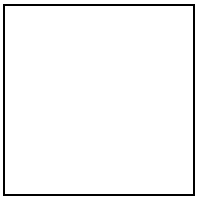
scf-650660



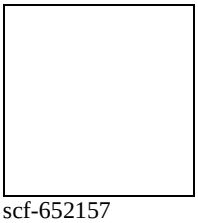
scf-651846



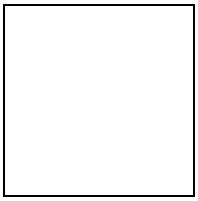
scf-651861



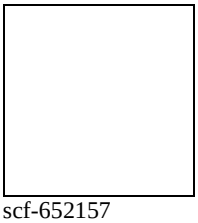
scf-652014



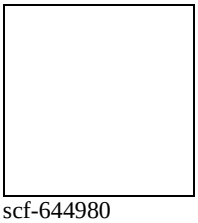
scf-652157



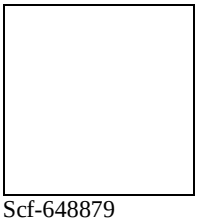
scf-652157



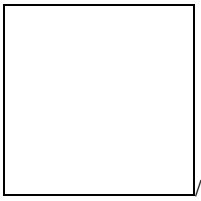
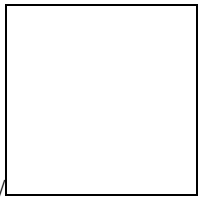
scf-652157



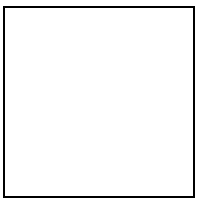
scf-644980



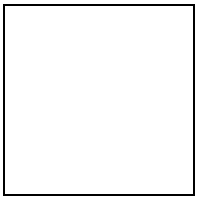
Scf-648879

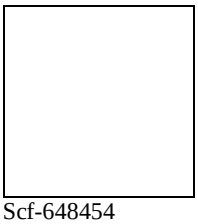
Scf-648975



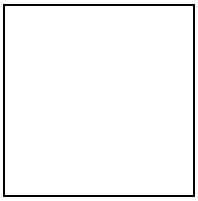
scf-648975



Scf-648975



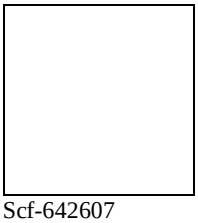
Scf-648454



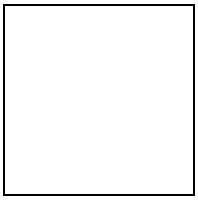
Scf-652157

Scf-652157

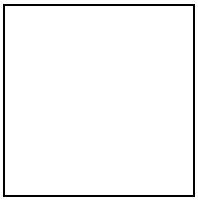
Scf-652157



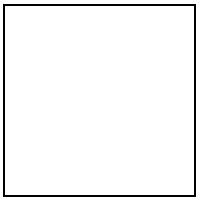
Scf-642607



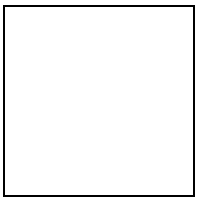
Scf-647496



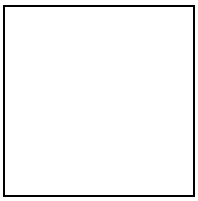
//



Scf-649782

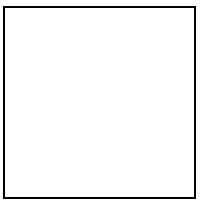


Scf-648956



Appendix 12

Functional classification of OBPs, CSPs, and CD36-like genes



Appendix 13

Predicted transmembrane domains of *G. m. morsitans* SNMP1 and SNMP2

The predicted transmembrane domains of SNMP1 are located at the sequence terminal ends, while those of SNMP2 are both on the C-terminal end

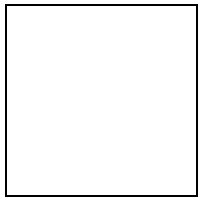


Table S1. The genomic annotation of *Glossina morsitans morsitans* OR and GR genes

G. m. morsitans	**Scaffold**	**Gene region**	**Reciprocal BlastP search**		
		(coding strand)	**NCBI acc. no.**	**E-value**	**Description**
GmmOR1	scf7180000648683	14760..27560 (-)	AFH96943.1	0.00E+000	odorant receptor co-receptor [Chrysomya rufifacies]
GmmOR2	scf7180000648756	29714..32260 (-)	NP_525046.1	5.00E-126	odorant receptor 2a [Drosophila melanogaster] >sp|O46077.2|
GmmOR3	scf7180000648228	109437..110684 (-)	XP_002100282.1	1.80E-088	GE16253 [Drosophila yakuba] >gb|EDX01390.1|
GmmOR4	scf7180000642438	111181..112437 (+)	XP_002088410.1	7.20E-080	GE18551 [Drosophila yakuba] >gb|EDW88122.1|
GmmOR5a	scf7180000639717	1658..7954 (-)	XP_004530390.1	5.30E-070	PREDICTED: odorant receptor 33b-like [Ceratitis capitata]
GmmOR6	scf7180000651846	662439..668878 (-)	XP_002080251.1	5.00E-073	GD10347 [Drosophila simulans] >gb|EDX05836.1|
GmmOR7	scf7180000651846	655700..658243 (-)	XP_002080251.1	2.40E-078	GD10347 [Drosophila simulans] >gb|EDX05836.1|
GmmOR8	scf7180000651846	659294..662305 (-)	XP_001961367.1	3.50E-076	GF13833 [Drosophila ananassae] >gb|EDV38189.1|
GmmOR9	scf7180000652157	136730..138889 (-)	XP_002080251.1	1.90E-096	GD10347 [Drosophila simulans] >gb|EDX05836.1|
GmmOR10	scf7180000644232	131824..133273 (+)	XP_004526336.1	6.60E-104	PREDICTED: odorant receptor 46a, A-like [Ceratitis capitata]
GmmOR11	scf7180000652156	1267206..1269387 (-)	XP_004526336.1	2.20E-074	PREDICTED: odorant receptor 46a, A-like [Ceratitis capitata]
GmmOR12	scf7180000651831	454103..455258 (-)	XP_002098364.1	9.00E-087	GE10342 [Drosophila yakuba] >gb|EDW98076.1|
GmmOR13	scf7180000645812	24048..26362 (-)	XP_002019305.1	8.00E-117	GL12310 [Drosophila persimilis] >gb|EDW37939.1|
GmmOR14	scf7180000641298	67504..69789 (+)	XP_004534467.1	3.00E-043	PREDICTED: putative odorant receptor 45a-like [Ceratitis capitata]
GmmOR15	scf7180000651027	85495..88783 (+)	XP_001361066.3	1.30E-040	Or45a [Drosophila p. pseudoobscura] >gb|EAL25642.3|
GmmOR16	scf7180000649095	5203..6553 (-)	XP_004534466.1	2.20E-054	PREDICTED: putative odorant receptor 45a-like [Ceratitis capitata]
GmmOR17	scf7180000648564	107291..115929 (+)	XP_002135335.1	2.10E-044	odorant receptor N [Drosophila p. pseudoobscura] >gb|EDY73962.1|
GmmOR18	scf7180000648792	266134..274630 (+)	ACO83222.1	1.00E-106	putative odorant receptor [Stomoxys calcitrans]
GmmOR19	scf7180000648792	277481..283655 (+)	ACO83222.1	0.00E+000	putative odorant receptor [Stomoxys calcitrans]
GmmOR20	scf7180000652157	706731..709253 (-)	XP_004518350.1	9.90E-008	PREDICTED: putative odorant receptor 85d-like [Ceratitis capitata]
GmmOR21	scf7180000652170	3479799..3481562 (-)	XP_001994082.1	6.90E-037	GH22929 [Drosophila grimshawi] >gb|EDV94818.1|
GmmOR22	scf7180000641538	58384..59452 (+)	AAD26356.1	4.80E-012	odorant receptor DOR62 [Drosophila melanogaster]
GmmOR23	scf7180000652170	25006662..25007967 (+)	XP_002036307.1	5.00E-067	GM17404 [Drosophila sechellia] >gb|EDW52230.1|

GmmOR24	scf7180000652157	888049..890329 (-)	XP_001359521.1	3.10E-040	Or85b [Drosophila p. pseudoobscura] >gb|EAL28667.1|
GmmOR25	scf7180000649009	35111..37791 (+)	XP_002047978.1	2.50E-038	GJ11611 [Drosophila virilis] >gb|EDW70320.1|
GmmOR26	scf7180000652170	7565092..7567386 (+)	XP_004535567.1	1.00E-061	PREDICTED: putative odorant receptor 67c-like [Ceratitis capitata]
GmmOR27	scf7180000650866	8500..9930 (-)	XP_004535566.1	7.20E-063	PREDICTED: unknown protein LOC101461468 [Ceratitis capitata]
GmmOR28	scf7180000650866	2343..4333 (-)	XP_004535567.1	6.10E-014	PREDICTED: putative odorant receptor 67c-like [Ceratitis capitata]
GmmOR29	scf7180000652141	245471..248227 (-)	XP_004535567.1	5.70E-037	PREDICTED: putative odorant receptor 67c-like [Ceratitis capitata]
GmmOR30	scf7180000652141	215509..218352 (-)	XP_321153.1	1.40E-115	AGAP001912-PA [Anopheles gambiae str. PEST] >gb|EAA01023.2|
GmmOR31	scf7180000650238	26161..28277 (-)	NP_523470.3	1.00E-039	odorant receptor 24a [Drosophila melanogaster] >sp|P81913.4|
GmmOR32	scf7180000648410	219323..223133 (-)	XP_002075406.1	2.60E-073	GK17746 [Drosophila willistoni] >gb|EDW86392.1|
GmmOR33	scf7180000648614	40250..49045 (+)	NP_523721.1	3.00E-067	odorant receptor 49b [Drosophila melanogaster] >sp|Q9V6H2.1|
GmmOR34	scf7180000652170	23283182..23286806 (+)	CBA13932.1	5.50E-022	odorant receptor 85d [Drosophila melanogaster]
GmmOR35	scf7180000648722	141942..145983 (-)	ADK48351.1	1.30E-067	odorant receptor 43a [Drosophila melanogaster]
GmmOR36	scf7180000648722	191876..198790 (-)	XP_004534194.1	1.10E-039	PREDICTED: putative odorant receptor 30a-like [Ceratitis capitata]
GmmOR37	scf7180000648373	4213..5709 (-)	XP_004531513.1	1.90E-118	PREDICTED: putative odorant receptor 74a-like [Ceratitis capitata]
GmmOR38	scf7180000648495	88188..91340 (+)	XP_004522725.1	7.70E-097	PREDICTED: odorant receptor 47b-like [Ceratitis capitata]
GmmOR39	scf7180000648080	333109..334447 (+)	XP_004536014.1	5.10E-081	PREDICTED: putative odorant receptor 88a-like [Ceratitis capitata]
GmmOR40	scf7180000649009	40802..42730 (+)	XP_002049113.1	1.20E-075	GJ21406 [Drosophila virilis] >gb|EDW60306.1|
GmmOR41	scf7180000649048	4425..5787 (-)	XP_002094291.1	1.50E-074	GE21738 [Drosophila yakuba] >gb|EDW94003.1|
GmmOR42	scf7180000649048	15..1464 (-)	XP_002094291.1	1.30E-080	GE21738 [Drosophila yakuba] >gb|EDW94003.1|
GmmOR43	scf7180000651490	579..1910 (+)	XP_002094291.1	9.00E-058	GE21738 [Drosophila yakuba] >gb|EDW94003.1|
GmmOR44	scf7180000648928	37939..39536 (-)	XP_002069080.1	5.40E-095	GK24034 [Drosophila willistoni] >gb|EDW80066.1|
GmmOR45	scf7180000650705	207676..209075 (+)	ADG96063.1	3.00E-089	putative odorant receptor [Stomoxys calcitrans]
GmmOR46	scf7180000645804	155235..156578 (-)	XP_002047834.1	1.30E-084	GJ13657 [Drosophila virilis] >gb|EDW70176.1|
GmmGR1	scf7180000650411	214655..216066 (+)	AFH96947.1	6.00E-092	gustatory receptor 1 [Chrysomya megacephala]
GmmGR2	scf7180000652170	18996209..18997720	XP_312786.1	4.00E-039	AGAP003098-PA [Anopheles gambiae str. PEST] >gb|EAA08342.2|
GmmGR3	scf7180000652170	18992721..18996270	XP_312786.1	2.00E-059	AGAP003098-PA [Anopheles gambiae str. PEST] >gb|EAA08342.2|
GmmGR4	scf7180000650833	228463..237409 (-)	AFH96945.1	2.00E-087	gustatory receptor 2 [Chrysomya megacephala]
GmmGR5	scf7180000647997	53981..55829 (+)	XP_001354191.2	3.00E-045	Gr66a [Drosophila p. pseudoobscura] >gb|EAL31243.2|
GmmGR6	scf7180000652170	14942509..14944789	NP_995642.1	7.00E-058	gustatory receptor 28b, isoform C [Drosophila melanogaster]

GmmGR7	scf7180000648889	6876..8857 (+)	NP_995642.1	7.00E-038	gustatory receptor 28b, C-like [Drosophila melanogaster]
GmmGR8	scf7180000651593	207733..209113 (+)	XP_001968324.1	2.00E-027	GG24573 [Drosophila erecta] >gb|EDV57383.1|
GmmGR9	scf7180000652170	23328288..23329768	XP_002021945.1	5.00E-046	GL14254 [Drosophila persimilis] >gb|EDW25886.1|
GmmGR10	scf7180000645661	235649..237209 (+)	XP_001962923.1	1.00E-032	GF14188 [Drosophila ananassae] >gb|EDV32144.1|
GmmGR11	scf7180000652146	594756..597494 (-)	XP_309027.2	8.00E-024	AGAP006716-PA [Anopheles gambiae str. PEST] >gb|EAA45495.2|
GmmGR12	scf7180000650947	5268..6453 (+)	XP_002087574.1	6.00E-022	GE17747 [Drosophila yakuba] >gb|EDW87286.1|
GmmGR13	scf7180000640662	25103..26585 (+)	XP_309026.1	9.00E-024	AGAP006717-PA [Anopheles gambiae str. PEST] >gb|EAA45494.1|
GmmGR14	scf7180000652170	14938990..14942365	XP_002018685.1	3.00E-015	Gr28b [Drosophila p. pseudoobscura] >gb|EDW36881.1|

The Glossina morsitans morsitans annotated gene loci names, scaffold identity, gene location within the scaffold, in bracket (*) refers to the coding strand where (-), reverse and (+), forward strands; columns under reiterative blast searches gives best Diptera query orthologs for Drosophila, Anopheles and Aedes

Table S2. Analysis of *G. m. morsitans* OR and GR conserved domains and transmembrane helices (Tmh) locations

Name	AA	# TMh	N-in prob.	Exp. 60 aa	N-term	Tmh-1	Tmh-2	Tmh-3	Tmh-4	Tmh-5	Tmh-6	Tmh-7	Tmh-8
GmmOR1	521	6	0.87678	13.3487	In	49..71	81..100	140..162	198..220	343..365	380..402		
GmmOR2	394	7	0.8464	23.84259	In	35..57	72..91	128..150	170..192	268..290	300..322	362..384	
GmmOR3	387	3	0.31305	21.32834	Out	40..62	-	-	-	269..291	301..323		
GmmOR4	384	7	0.78613	22.51965	In	36..58	68..87	132..151	171..188	261..283	288..310	358..380	
GmmOR5a*	442	5	0.69118	22.27873	In	24..46	111..133	140..162	-	275..297	304..326		
GmmOR6	387	5	0.92706	21.53909	In	37..59	74..96	136..158	-	278..300	307..329		
GmmOR7	406	6	0.98573	20.91439	In	40..62	77..99	137..159	174..196	250..272	277..299		
GmmOR8	389	6	0.61212	5.59897	In	-	74..96	111..133	169..191	206..228	282..304	308..330	
GmmOR9	409	6	0.69855	16.69951	In	48..70	85..107	142..164	-	263..285	292..314	369..391	
GmmOR10	444	6	0.99871	18.15249	In	43..65	75..94	133..152	172..194	256..278	288..310		
GmmOR11	341	6	0.55708	29.24041	Out	15..34	41..63	73..92	131..153	183..205	-	315..337	
GmmOR12	340	3	0.42533	13.4897	In	-	88..110	130..152	208..230				
GmmOR13	391	6	0.9995	21.53028	In	39..61	66..88	130..152	184..206	263..285	295..317		
GmmOR14	341	6	0.95953	11.51876	In	50..72	82..101	144..166	194..216	275..297	307..341		
GmmOR15	446	7	0.99405	19.87949	In	41..59	74..91	128..150	178..200	263..285	295..317		
GmmOR16	387	6	0.9884	21.40992	In	39..61	67..89	136..155	185..207	262..284	294..313		
GmmOR17	541	8	0.8757	0.18817	In	155..177	192..214	252..274	306..328	387..409	419..441	489..508	518..540
GmmOR18	420	6	0.86327	0.15158	In	-	85..107	143..162	212..234	282..304	311..330	383..405	
GmmOR19	385	7	0.84056	15.3963	In	44..66	81..100	154..176	186..208	258..280	284..306	349..371	
GmmOR21	465	6	0.94926	2.52035	In	62..79	94..116	159..181	210..232	329..351	361..383		
GmmOR23	331	2	0.78159	0.52777	In	-	-	140..162	177..199				
GmmOR24	388	5	0.88063	23.0158	In	39..61	71..93	137..159	-	276..298	311..333		
GmmOR25	385	5	0.97167	21.34719	In	41..63	73..92	132..154	180..202	264..286			
GmmOR26	418	6	0.53775	22.63457	In	37..59	74..91	131..153	183..205	266..288	298..320		
GmmOR27	415	6	0.95522	21.65919	In	38..60	70..92	136..158	199..221	268..290	300..319		
GmmOR29	438	5	0.38017	30.52399	In	44..66	81..98	140..162	190..212	273..295			
GmmOR30	361	6	0.87382	19.34639	In	44..66	81..98	140..162	177..196	203..225	240..262		
GmmOR31	435	7	0.94637	20.59435	In	40..62	72..94	132..154	179..201	266..288	293..315	399..421	
GmmOR32	450	4	0.54446	11.30211	Out	51..73	-	147..169	189..211	287..309			

GmmOR33	353	6	0.35535	33.17228	Out	29..46	56..78	124..146	166..197	247..269	274..296		
GmmOR34	360	5	0.73982	14.78357	In	33..50	60..82	128..150	170..201	250..272			
GmmOR35	392	7	0.94641	21.34345	In	37..59	63..85	129..151	183..205	256..278	291..313	352..374	
GmmOR36	343	5	0.4472	22.45052	Out	26..48	-	118..140	176..198	241..263	273..292		
GmmOR37	430	4	0.26882	17.73285	Out	43..60	-	137..159	179..201	275..297			
GmmOR38	371	6	0.93987	36.80443	In	13..35	45..67	107..129	159..181	242..264	279..300		
GmmOR39	403	6	0.70258	12.75432	In	48..67	77..94	141..160	194..216	276..298	303..322		
GmmOR40	284	4	0.88639	17.14596	In	47..66	71..93	130..152	194..216				
GmmOR41	386	6	0.99681	22.52011	In	33..55	65..87	135..157	188..210	265..287	292..314		
GmmOR42	386	6	0.99529	22.3917	In	34..56	66..88	133..155	175..197	265..287	292..314		
GmmOR43	389	6	0.96647	22.61542	In	39..61	76..98	142..164	184..206	290..312	317..339		
GmmOR44	390	6	0.99827	22.19978	In	37..59	69..91	139..161	191..213	267..289	294..316		
GmmOR45	385	5	0.4536	27.4459	Out	35..57	-	129..151	174..196	260..282	286..308		
GmmOR46	348	5	0.76428	23.05999	Out	40..62	-	105..127	142..164	226..248	253..275		
GmmGR1	425	6	0.81819	0.00016	In	-	-	112..134	149..171	198..220	233..255	304..326	341..358
GmmGR2	514	6	0.72373	0.0086	In	-	-	116..138	158..180	215..237	242..264	319..341	431..453
GmmGR3	425	6	0.87702	0	In	-	-	106..125	140..162	191..213	223..245	301..323	328..350
GmmGR4	496	6	0.98521	0.00315	In	-	-	101..123	143..165	200..222	227..249	299..321	336..358
GmmGR5	467	7	0.98615	33.01159	In	13..35	50..69	82..101	145..167	317..339	354..376	434..456	
GmmGR6	443	8	0.22336	27.46018	Out	15..37	50..72	87..109	149..168	178..200	281..300	315..337	396..415
GmmGR7	402	7	0.05239	19.11982	Out	33..50	63..85	105..127	161..183	193..215	276..295	310..332	
GmmGR8	407	6	0.94288	19.62217	In	31..53	68..90	153..175	190..212	262..284	304..326		
GmmGR9	348	4	0.26473	23.23022	Out	36..58	-	-	199..221	241..263	319..341		
GmmGR10	458	7	0.04212	20.555641	Out	26..45	65..87	102..119	160..179	194..216	330..352	362..384	
GmmGR11	450	6	0.98956	0.23729	In	-	78..100	110..132	166..188	203..225	274..296	311..333	
GmmGR12	375	8	0.62907	16.91026	In	39..61	76..95	97..119	123..145	166..188	203..225	275..297	310..332
GmmGR13	457	6	0.99563	22.72625	In	36..58	73..95	138..155	170..192	265..287	302..319		
GmmGR14	309	6	0.27360	24.90738	Out	20..39	51..73	93..112	144..166	181..203	210..232		

Columns:Name - symbol of gene model assigned by community annotator; AA - number of amino acid residues in gene, the average sizes for GmmORs and GmmGRs are 387 and 427 amino acids respectively; # TMh - number of predicted 7-trans-membrane (7tm-6-olf-rcpt) domains; Exp. 60 aa - expected probability of finding a trans-membrane within the first 60 amino acids; N-term - probability of the N-terminal being located intra-cellularly or extra-cellularly; Tmh-* - sequential residue range for the predicted trans-membrane regions within the peptide sequences, dashes, "-" within the table show adjustments for possibly disabled domains. Asterix (*) - only one splice variant analyzed is shown.

Table S3. Trnas-membrane helices and loop sizes (in number amino acids)

Name	Loop 1	Tmh 1	Loop 2	Tmh 2	Loop 3	Tmh 3	Loop 4	Tmh 4	Loop 5	Tmh 5	Loop 6	Tmh 6	Loop 7	Tmh 7	Loop 8	Tmh 8
GmmOR1	48	22	10	19	40	22	38	22	132	22	15	22				
GmmOR2	34	22	15	21	137	22	38	22	76	22	10	22	40	22		
GmmOR3	39	22	107	22	10	22										
GmmOR4	35	22	10	19	45	19	20	17	73	22	5	22	48	22		
GmmOR5	23	22	65	22	7	22	113	22	7	22						
GmmOR6	36	22	15	22	40	22	120	22	7	22						
GmmOR7	39	22	15	22	38	22	15	22	54	22	5	22				
GmmOR8	73	22	15	22	36	22	15	22	54	22	4	22				
GmmOR9	47	22	15	22	35	22	99	22	7	22	55	22				
GmmOR10	42	22	10	19	39	22	20	22	62	22	10	22				
GmmOR11	14	19	7	22	10	19	39	22	30	22	110	22				
GmmOR12	87	22	20	22	56	22										
GmmOR13	38	22	5	22	32	22	32	22	57	22	10	22				
GmmOR14	49	22	10	19	33	22	28	22	59	22	10	34				
GmmOR15	40	18	15	17	37	22	28	22	63	22	10	22				
GmmOR16	38	22	6	22	47	19	30	22	55	22	10	19				
GmmOR17	154	22	15	22	38	22	32	22	59	22	10	22	48	19	10	22
GmmOR18	84	22	36	19	50	22	48	22	7	22	53	22				
GmmOR19	43	22	15	19	54	22	10	22	50	22	4	22	43	22		
GmmOR21	61	17	15	22	43	22	29	22	97	22	10	22				
GmmOR23	139	22	15	22												
GmmOR24	38	22	10	22	44	22	117	22	13	22						
GmmOR25	40	22	10	19	40	22	26	22	62	22						
GmmOR26	36	22	15	17	40	22	30	22	61	22	10	22				
GmmOR27	37	22	10	22	44	22	41	22	47	22	10	19				
GmmOR29	43	22	15	17	42	22	28	23	61	22						
GmmOR30	43	22	15	17	42	22	15	19	7	22	15	22				

GmmOR31	39	22	10	22	38	22	15	22	65	22	5	22	84	22
GmmOR32	50	22	74	22	20	22	76	22						
GmmOR33	28	17	10	22	46	22	20	31	50	22	5	22		
GmmOR34	32	17	10	22	46	22	20	31	49	22				
GmmOR35	38	22	4	22	44	22	32	22	51	22	13	22	39	22
GmmOR36	25	22	70	22	36	22	43	22	10	19				
GmmOR37	42	17	77	22	20	22	74	22						
GmmOR38	12	22	10	22	40	22	30	22	61	22	15	21		
GmmOR39	47	19	10	17	47	19	34	22	60	22	5	19		
GmmOR40	46	19	5	22	37	22	42	22						
GmmOR41	32	22	10	22	48	22	31	22	55	22	5	22		
GmmOR42	33	22	10	22	45	22	20	22	68	22	5	22		
GmmOR43	38	22	15	22	44	22	20	22	84	22	5	22		
GmmOR44	36	22	10	22	48	22	30	22	54	22	5	22		
GmmOR45	34	22	75	22	23	22	64	22	4	22				
GmmOR46	39	22	43	22	15	22	62	22	5	22				

The loops and helices are labeled sequentially from N-termini to C-termini. Bold type highlights within table indicate disproportionate expansions.