PUTATIVE FUNCTIONS OF NOVEL ANTENNAE CHEMOSENSORY GENE REPERTOIRE OF MALE TSETSE FLY, *Glossina morsitans morsitans*

BILLIAH KEMUNTO BWANA

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN GENETICS OF THE UNIVERSITY OF EMBU

NOVEMBER, 2022

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature …….…….. Date ……

Ms. Billiah Bwana
Department of Biological Sciences
B520/1221/2018

This thesis has been submitted for examination with our approval as University Supervisors.

Signature …….…….. Date ……

Dr. Julius N. Mugweru
Department of Biological Sciences
University of Embu

Signature …….…….. Date ……

Dr. Franklin N. Nyabuga
Department of Biological Sciences
University of Embu

Signature …….…….. Date ……

Dr. Paul O. Mireji
Kenya Biotechnology Research Institute
Kenya Agricultural and Livestock Research Organization

DEDICATION

I dedicate this work to my family and colleagues for their prayers, love and support never falling short.

ACKNOWLEDGEMENT

First and foremost, I want to thank the Almighty God for the grace he has granted me this far. In a special way, I register my gratitude to the University of Embu for giving me a special opportunity to pursue my degree, I feel forever indebted. I also register my sincere gratitude to my supervisors Dr. Julius Mugweru, Dr. Franklin N. Nyabuga and Dr. Paul Mireji for their unmeasurable support in my research journey. May the good Lord bless you abundantly for I am a better person today because of you and this inspired journey with you beside me has been inked into memory forever. I appreciate my family for sincerely praying, encouraging and supporting me even in the hardest times. I am grateful to my colleagues for the constant and resourceful advice, sharing of ideas and knowledge throughout this journey. Finally, I appreciate personnel from the Department of Molecular Biology, Biotechnology Research Institute, Muguga-Nairobi and Gloria-Isabel from VectorBase, University of Notre Dame-USA for making this research a success.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

AAT	Animal african trypanosomiasis
CDC	Center for disease control and prevention
CPs	Chemosensory proteins
FAO	Food and agricultural organization
GRs	Gustatory receptors
HAT	Human african trypanosomiasis
IRs	Ionotropic receptors
mRNA	Messenger ribonucleic acid
OBPs	Odorant binding proteins
ORs	Odorant receptors
PCR	Polymerase chain reaction
RNA-seq	Ribonucleic acid sequences
SNMPs	Sensory neuron membrane receptors

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Tsetse fly is a primary vector of Human African Trypanosomiasis and Animal African Trypanosomiasis. Tsetse fly exploits chemical cues from the environment to distinguish a non- from suitable hosts. Genes in tsetse fly antennae code for proteins and receptors that directly or indirectly mediate chemoreception. While chemoreception-associated genes have been annotated, antennal genes in the *Glossina. m. morsitans* genome with important functions have not been characterized. Antennae-specific raw reads from adult flies exposed to four treatments, namely, fed, unfed-exposed to ε-nonalactone attractant, unfed-exposed to δ-nonalactone repellant and unfed-exposed to paraffin diluent (control) were mapped onto *G. m. morsitans* gene-set. Reads that did not map were isolated and *de novo* assembled into transcripts. Protein-coding gene regions associated with these transcripts were predicted, annotated and curated as partial/complete genes. Annotated putative orthologs/homologs for these genes in *Drosophila melanogaster* (*Dm*), *Musca domestica* (*Md*) or *Anopheles gambiae* (*Ag*) genomes were identified. Finally, differential expression of the novel or existing genes in relation to odor exposures relative to no-odor control (unfed flies) were assessed and expression of existing genes quantified through qPCR. Results showed that 45.21% of the sequenced reads did not map to the gene set. These reads assembled into 72,428 unique transcripts that yielded 592 genes among which 202 were novel and 390 were improvements of existing genes in the *G. m. morsitans* genome. Among the novel genes, 94 had orthologs in *Dm*, *Md* or *Ag* and 88 had homologs in UniProt databases. These orthologs were putatively associated with non-canonical olfactory roles, thus providing insight into their specific roles in antennal physiological processes. A novel gene (GMOY014237.R1396) and 15 existing genes were differentially expressed in response to the attractant or repellent. Differential expression through qPCR analysis unveiled three antennal transcripts, i.e., the coat protein epsilon, cyclin-dependent kinase and odorant receptor 45, all three up-regulated in response to the attractant. Novel genes sequences were adopted by VectorBase, updating the existing *G. m. morsitans* annotations. This study identified 108 potentially tsetse fly-specific antennal genes. The novel antennal genes could be used as baseline data in studies of other tsetse fly species, and with an orientation towards attraction and or repellency in their control.

CHAPTER ONE

INTRODUCTION

1.1 Background

Tsetse flies (Diptera*; Glossinidae*) are primary insect vectors of Human African Trypanosomiasis (HAT, commonly known as Sleeping sickness) and African Animal Trypanosomiasis (AAT, commonly known as Nagana) in livestock (Leak, 1999). Different species of tsetse fly exist and in East Africa *Glossina morsitans morsitans* and *Glossina pallidipes pallidipes* are the major vectors of trypanosomes (Franco *et al*., 2014). Adult tsetse flies solely feed on vertebrate blood and are therefore efficient vectors of the trypanosomes (Mullen and Durden, 2019).

Trypanosomiasis is widely spread in sub-Saharan Africa where the risk of infection is high and untreated human infections are fatal (Büscher *et al*., 2017). Human African trypanosomiasis is a threat to millions from 36 Sub-Saharan African countries while AAT is distributed in 38 countries (Aksoy *et al*., 2017; WHO, 2022). Outbreaks of trypanosomiasis in sub-Saharan Africa hinder agricultural production due to anemia and animal abortions (Holt *et al*., 2016; FAO, 2018). The AAT severely obstructs agricultural development, restricting nutritional sources and economic success in areas invaded with tsetse flies hence resulting in economic losses of about 4.5 billion US dollars annually (Shaw *et al*., 2014).

Trypanosomes causing trypanosomiasis present a complex life cycle revolving between the insect vector and the mammalian host. Bloodstream trypomastigotes in an infected mammalian host are ingested by tsetse fly when obtaining a blood meal. The trypomastigotes establish in the midgut as procyclic forms that exponentially proliferate and mature into epimastigotes. The epimastigotes then migrate to the salivary glands or the proboscis where they transform into meta-cyclic forms ready for transmission to an uninfected mammalian host (Gibson and Bailey, 2003; Dunn and Adigun, 2018).

Despite significant efforts, tsetse flies and trypanosomiasis remain major public health threats in sub-Saharan Africa (CDC, 2012). Vaccine development and trypanocidal drug development strategies for HAT and AAT have failed due to the emergence of resistance and adverse side effects (Mullen and Durden, 2019). *Trypanosoma brucei rhodesiense* is resistant to independent treatment of suramin and melarsoprol that on the other hand causes induced

ancephalopathic syndrome in patients and contributes to about 50% of the deaths (Dunn and Adigun, 2018).

The ability of tsetse fly to detect and respond to volatile and non-volatile odor cues in their environment helps them identify and distinguish suitable from non-suitable hosts for a blood meal. Tsetse fly species, sex and ecological location largely affect this behavior (Gikonyo *et al*., 2000; Gikonyo *et al*., 2002; Gikonyo *et al*., 2003). Odor cues can be derived from host breath for instance acetone or from host urine as microbial by-products such as 4-cresol. Acetone, mostly used as a standalone attractant is also a component of the well established *G. m. morsitans* attractant POCA (3-*n*-Propylphenol, 1-Octen-3-ol, *p*-Cresol and Acetone) widely used in its control and of other savanna tsetse fly species (Willemse *et al*., 1991).

In light of the knowledge that tsetse fly use olfactory cues in their antennae to identify their suitable host, traps saturated with host odors have been employed in tsetse fly control (Chahda *et al*., 2019). Traps laced with different host derived odors show differences in catches for both sexes and species because each *Glossina* species respond differently to odors due to their gene guided host species-specific preferences (Omolo *et al*., 2009; Wachira *et al*., 2016).

Tsetse fly antennae sensilla is the major appendage involved in chemoreception to perceive odors (Liu *et al*., 2010). The *Glossina* antennae harbor genes expressed in odorant binding receptors corresponding to chemical odors (Chahda *et al*., 2019). These genes encode for odorant binding proteins (OBPs), chemosensory proteins (CSPs), gustatory receptors (GRs), odorant receptors (ORs), sensory neuron membrane proteins (SNMPs) and ionotropic receptors (IRs) that are essential for chemoreception (Liu *et al*., 2012; Masiga *et al*., 2014). Expression of the genes encoding for these proteins and receptors is vital for host finding and obtaining of a blood meal by tsetse flies (Obiero *et al*., 2014; Nyanjom *et al*., 2018).

Macharia *et al*. (2016) annotated 30 OBPs, 5 CSPs, 2 SNMPs, 14 GRs, 30 IRs and 14 ORs encoding genes from *G. m. morsitans* and postulated that the chemosensory genes mediate response to different odors. The team identified 127 chemosensory genes from whole female RNA sequencing. Caers *et al*. (2015) found 39 neuropeptide precursor encoding genes and 43 neuropeptide receptor genes in *G. m. morsitans* that are postulated to take part in the response of this species to their host odors.

The current annotations of six tsetse fly genomes for *G. m. morsitans, G. pallidipes, G. austeni, G. brevipalpis, G. palpalis* and *G. fuscipes* are incomplete as the annotations were heavily

dependent on existing transcriptomes from tsetse fly of specific sex, treatments and different organs for example midgut but not inclusive of all body organs, geared towards particular research interests and needs at the time (International Glossina Genome Initiative, 2014: Attardo *et al*., 2019). As a consequence, annotation of antennal gene repertoire for tsetse fly species from antennae-specific divergent transcripts is investigated in this current study.

1.2 Statement of the problem

Management strategies towards the tsetse flies and the trypanosomes they transmit to humans and animals have been unsuccessful due to changes in fly behavior, and evolution of the trypanosome parasite though constant antigenic variation. Effect on non-target population and the dynamic tsetse fly host range has rendered tsetse fly control methods such as aerial insecticide spraying and insecticide treated animals inefficient (Kuzoe *et al*., 2005; Percoma *et al*., 2018). Tsetse fly response to olfactory cues has guided in the design of repellent or attractant odor baited controls, a more effective strategy in tsetse fly management. This has been achieved through antennal gene expression elucidation and functional characterization. However, only a small fraction of the antennal genes has been identified. From contemporary molecular techniques and computational analysis, only about 50% of the RNA reads from the *G. m. morsitans* antennae map absolutely to the known gene sets in the reference genome. This therefore, leaves about 50% of the gene set unidentified.

1.3 Justification

Host finding and selection in tsetse flies is enhanced by olfactory signals and their response is mediated by chemosensory proteins encoded by highly expressed olfactory genes at the antennae (Zhang *et al*., 2017). Knowledge on the expression of an olfactory gene repertoire in respect to antennae perceived host odors has facilitated the development of odor based technologies for tsetse fly control. As a result of limited knowledge of genes, a few studies confined to a few genes are available. This study contributes to the available knowledge on the identity and function of *G. m. morsitans* antennal genes by utilizing the fast developing technologies in genome sequencing and bioinformatics. This will give a better understanding of the molecular basis for response coordination in the insect that can be exploited in developing olfactory based management strategies towards the insect vectors.

1.4 Hypothesis

Glossina morsitans morsitans antennae harbours novel chemosensory genes with unknown biochemical and olfactory functions.

1.5 Objectives

1.5.1 General objective

To identify novel *G. m. morsitans* antennae expressed chemosensory genes and their putative functional roles.

1.5.2 Specific objectives

1. To annotate novel antennae expressed chemosensory genes in male *G. m. morsitans* antennae.
2. To determine putative functional roles of the identified *G. m. morsitans* novel chemosensory genes.
3. To quantify the expression levels of the identified *G. m. morsitans* novel chemosensory genes.

CHAPTER TWO

LITERATURE REVIEW

2.1 Tsetse flies and Trypanosomes

Tsetse flies of genus *Glossina* that transmit African trypanosomiasis, are geographically distributed depending on their ecological and feeding preferences as well as oviposition sites (Bogitsh *et al*., 2018). This genus is composed of 31 tsetse fly species and all adults are exclusively hematophagous presenting medical and veterinary importance as vectors of trypanosomes in wild and domesticated animals as well as humans (Bourn *et al*., 2001). *Glossina* species in Kenya include; *G. pallidipes, G. longipennis, G. austeni, G. brevipalpis, G. swynnertoni* and *G. fuscipes* dispersed in the western, southern and coastal regions and geographically coexist (Ngari *et al*., 2020).

Tsetse fly exhibits a specialized feeding method where the female ingests blood as a sole source of energy and nutrient. These nutrients are later availed to the uterine developing larvae in a highly modified accessory 'milk' gland (Attardo *et al*., 2006). Tsetse fly harbor about 250 proteins that assist in blood feeding and digestion (Alves-Silva *et al*., 2010). High temperatures above 34°C impact blood feeding and digestion by increasing their rates respectively hence enhancing the transmission of trypanosomiasis (Terblanche *et al*., 2008)

Trypanosoma parasites causing nagana in animals and sleeping sickness in humans have a complex life cycle rotating between their insect vector and a mammalian host. *Trypanosoma brucei rhodesiense* and *T.b. gambiense* cause HAT and *T.b brucei* cause AAT (Simarro *et al*., 2011; Savage *et al*., 2016). Infected tsetse fly transmits the parasite to a healthy host during a blood meal. In the mammalian hosts including humans' bloodstream, the parasite is presented as a slender extracellular proliferative form which then differentiates into G_1 stumpy form that is ready to be taken up by the tsetse fly. The trypanosomes express variable surface glycoproteins to evade the host immune response as they differentiate and increase in number. Upon uptake of a blood meal short stumpy bloodstream-form trypomastigotes in the insect midgut transform to procyclic forms that infect the fly (Matthews, 2005; Chou *et al*., 2010; Ooi and Bastin, 2013).

Procyclins differentiate by changing their variable surface glycoproteins and migrate to the salivary glands where they exist as epimastigotes. Later epimastigotes proliferate and transform

to metacyclic forms ready to infect a new mammalian host (Matthews, 2005). The disease condition in humans is characterized by infection of the central nervous system and haemo-lymphatic trypanosome proliferation (Thuita *et al*., 2008).

2.2 Olfaction in tsetse flies and its exploitation in their control

Insects depend on olfaction for feeding, host identification, mating and oviposition (Andersson *et al*., 2015). Olfaction is mediated by the sensilla that are located on the insect antennae (Hu *et al*., 2016). *Glossina* also use olfaction signals in locating their host for food, mates as well as larviposition (Masiga *et al*., 2014). This process is based on the activity of olfactory proteins that include: odorant binding proteins, chemosensory proteins, odorant degrading enzymes and chemoreceptors that include: gustatory receptors, ionotropic receptors and odorant receptors (Liu *et al*., 2010; Obiero *et al*., 2014; Nyanjom *et al*., 2018).

The sensilla on the antennae house dendrites of olfactory receptor neurons in which each responds to a specific cue which can be pheromones or conventional odors (Carey and Carlson, 2011). Each odorant receptor neuron gene expresses olfactory receptor proteins which are encoded by odorant, gustatory and ionotropic receptor genes. Insects physically respond to the stimuli based on olfactory receptor neurons activated. Hence, identification and establishment of gene networks as a result of receptor encoding gene activation enlighten on tsetse fly response by different stimuli (Carey and Carlson, 2011; Clark and Ray, 2016).

Odor molecules are transported to the odorant receptor neuron membrane by the OBPs where they coordinate with the receptors to propel an action potential (Gadenne *et al*., 2016). Genes encoding for these proteins are highly expressed in respective olfactory tissues. Characterization of the entire gene collection gives a better understanding of their role in the chemosensory pathway in regard to different odors (Zwiebel and Takken, 2004). Odorant receptors, therefore, are at the forefront in chemical stimuli detection and transforming them into electrical signals (Bohbot and Pitts, 2015) and impact the behavioral response which can be exploited in control. Even though different odorants can interact with the olfactory receptors, the most prominent mechanism that has been exploited in preventing hosts from vectors is the activation of olfactory receptors resolved for aversion of particular species repellents (Clark and Ray, 2016).

The use of repellents has been one of the strategies used in preventing animals and humans from tsetse fly bites (Liu *et al*., 2012). Saini *et al*. (2017) revealed that the savanna tsetse flies

avoid feeding on the waterbuck and postulated that it might have an odor that repels the flies. In addition to acids, ketones, and phenols, the major repellent identified previously is the compound δ-octalactone (Gikonyo *et al*., 2002). The use of traps enriched with host odors has been shown to be more effective. The flies respond to attractants by flying upwind towards the odor (Wachira *et al*., 2016) and it's presumed that the OBPs or CPs single out the vaporous odors of attractants as they go through the antennal pores and are conveyed through the sensilla to ORs (Nyanjom *et al*., 2018).

2.3 Antennal chemosensory genes in tsetse flies and their role in olfaction

Liu *et al*. (2010) established that some genes in the antennal library encode for proteins that are involved in semiochemical transport and metabolism, neuron signaling and basic metabolism as well as the building of cell constituents. Other than encoding for proteins that are responsible for chemoreception, the antennal genes also encode for enzymes that break down the same proteins when feeding ceases (Hu *et al*., 2016).

Studies show that approximately six gene families in *Glossina* are involved in olfaction and they encode for proteins and receptors that include OBPs, CPs and SNMPs, and the ORs, IRs and GRs receptors (Zhang *et al*., 2017). Both OBPs and CPs transport hydrophobic odorants and activate the receptors, but the CPs are also vital in sex identification. The SNMPs are essential for detecting pheromones, the ORs recognize brain signals, and IRs are involved in odor detection (Tian *et al*., 2018). Other GRs are exhibited in taste organs and used in contact chemoreception. Gustatory receptors genes also encode for carbon dioxide receptors at the antennae (Yuvaraj *et al*., 2018).

In *Glossina*, the chemosensory genes are organized as scaffolds that occur away from each other unlike other insects such as *D. melanogaster* where they occur in clusters (Moindi *et al*., 2018). Tsetse fly chemosensory genes are distributed across the genome and considering the 127 chemosensory genes and eight pseudogenes observed on *Glossina* it is evident that the tsetse fly have a narrow range of chemosensory genes compared to other dipterans such as *D. melanogaster* (Macharia *et al*., 2016). Distinct clusters of odorant and gustatory receptors and the corresponding gene clusters in *G.m. morsitans* have a similar function as their homologs in *Drosophila* chemosensory genes (Masiga *et al*., 2014; Obiero *et al*., 2014).

The clustering of ORs and GRs encoding genes signifies that a common regulatory mechanism exists in response to familiar stimuli due to joint collective gene expression (Masiga *et al*.,

2014). However, there are only a limited number of ORs and GRs since tsetse fly specializes on expression of a few chemosensory genes that influence adaptive behavior. Due to a restricted diet, *G. m. morsitans* lack GR genes which are associated with sweet tastes in other dipterans. Similarly, *Glossina obp*56i and *obp*19 genes have sequence deletions between C3 and C4 cysteine residues that are conserved in other insects (Macharia *et al*., 2016).

Most of the OBPs are less conserved in *Glossina* but highly conserved in other insects such as *D. melanogaster* throughout their genomes and hence have a preserved function (Macharia *et al*., 2016). The CPs and OBPs are soluble proteins highly concentrated in the sensilla of the tsetsefly antennae. These two classes of proteins are characterized by a four-cysteine and six-cysteine signature, respectively. The CPs mediate interactivity between the odorants/pheromones and odorant receptors while OBPs are associated with odor recognition (Pelosi *et al*., 2006). The CPs in the *G. m. morsitans* have been associated to host finding in females and preventing mating between females (Moindi *et al*., 2018).

The use of RNA with conventional quantitative PCR has shown that two *G. m. morsitans* CPs genes, *GmmCSP1* and *GmmCSP2*, are highly induced in the antennae irrespective of the age of tsetse flies and hence projected to be essential in olfaction (Liu *et al*., 2012). However, *GmmCSP1* and *GmmCSP3* show similar expression levels before and after feeding, an indication that nutritional status does not affect their expression. *GmmCSP2* is highly induced in newly emerged females and starved for 24 hours or 48 hours and hence strongly tied to influencing 'host seeking' physiological state (Liu *et al*., 2012).

Liu *et al*. (2010) identified three OBPs and 30 putative genes from *G. m. morsitans* antennae and from the body without heads, respectively. Macharia *et al*. (2016) on the other hand, identified 30 OBPs genes and three pseudogenes using a whole female RNA-seq. Evidence from RNA-seq in other dipterans such as the *D. melanogaster* shows that the odorant binding proteins genes are highly expressed at the antennae and are associated with feeding. This approach has also contributed to other novel genes identification such as the ammonium transporter gene (Menuz *et al*., 2014).

2.4 Odorant degrading enzymes expressed by tsetsefly antennal genes

Identification of odors by insects in the olfactory sensilla entail three steps; transport of odor conciliated by OBPs along the sensilla lymph, their interaction of ORs on the odorant receptor neurons membrane and finally inactivation of the odors (Steiner *et al*., 2017). Numerous studies

show that some insect pheromones are broken down by respective odorant degrading enzymes belonging to esterases, cytochromes, aldehyde oxidases and s-transfarases P450s enzyme families (Chertemps *et al*., 2012; Chertemps *et al*., 2015). For example, a protein encoded by the juvenile hormone esterase gene *(jhdup)* in *D. melanogaster* is able to degrade different food odors (Steiner *et al*., 2017). Esterase 6 expressed in the male antennae of *D. melanogaster* degenerate sex and aggregate pheromone cis-vaccenyl acetate, indirectly stimulating egg-laying but it's also useful in olfaction (Chertemps *et al*., 2012; Younus *et al*., 2017).

2.5 Tsetse fly antennal gene expression with respect to environmental cues

In other dipterans other than *G. m. morsitans* such as *Anopheles gambiae,* OBPs, ORs and IRs are down-regulated after a blood meal during the resting period to prevent the activity of the chemoreception system, and the OBPs are up-regulated about 24 hours after a blood meal (Taparia *et al*., 2017). Moindi *et al*. (2018) established that *GmmOR*33 and *GmmOR*45 odorant receptor genes in *Glossina* are vastly expressed in female and male antennae, respectively. Other gustatory receptors genes are highly expressed in female flies in response to certain cues for instance genes for Gr21a are activated several hours after feeding and *obp83a* gene is activated when hungry suggesting that they are essential for host finding (Macharia *et al*., 2016).

Obstruction of OBPs expression in *G. fuscipes* through double-strand RNA interference blocked attraction of this tsetse fly to 1-octen-3-ol in behavioral assays. For this instance, *GffObp83a1, GffObp83a2* and *GffObp83a4* were silenced, a demonstration of their significance and probable expression in response to attractive odors in nature (Diallo *et al*., 2021). Further, these OBPs also show an affinity to δ-octalactone, geranyl acetone and guaiacol constituents of Waterbuck Repellent Blend (WRB) that elicit repulsive behavior in *G. fuscipes* (Diallo *et al*., 2021).

2.6 Characterization and functional analysis of essential antennal genes in tsetse fly

Whole-genome sequencing, comparative modeling coupled with quantitative polymerase chain reaction (qPCR) assays have been widely employed in gene identification and characterization (Duan *et al*., 2015; Macharia *et al*., 2016; Tian *et al*., 2018). *Glossina* antennal genes, features and functional roles have been identified by comparative homology but this method is unreliable due to the lack of homologs in close relatives which is attributed to olfactory genes

undergoing positive selection pressure to facilitate adaptation to different odors (Macharia *et al*., 2016; Robertson, 2018).

Glossina genes resemble those of other dipterans in structure and sequence length. Exceptions include genes such as *Or67c, Or67d* and *Or43a* which are expanded in *Glossina* than in other dipterans. *Glossina m. morsitans* has six copies of Or67d paralogs, which code for cis-vaccenyl acetate receptor rather than five paralogs as in other dipterans (Macharia *et al*., 2016).

Organ-specific transcriptome and genomic sequencing approaches promote discovery of novel genes (Montagné, *et al*., 2015). Novel gene characterization and their putative functions in numerous insects' antennal transcriptome have been identified with similar methods. The use of molecular techniques and computational analysis in regard to the tsetse fly antennae only, ascertaining essential genes in tsetse fly and their functions illuminates on its biology as well as loss and gain of gene function.

2.7 Transcriptome based annotations in tsetse fly

The International Glossina Genome Initiative (2014) annotated *de novo* a 366 Mb *G. m. morsitans* genome using whole female transcriptomes from multiple sequencing methods and estimated the genome to comprise of 12, 308 protein-coding genes. Consequently, further annotations of organ and process specific associated genes such as chemosensory genes in this species has been effectuated (Liu *et al*., 2010; Obiero *et al*., 2014; Macharia *et al*., 2016). Four genomes including those of *G. pallidipes, G. austeni, G. brevipalpis* and *G. fuscipes* have also been annotated but there is a decrement in the number of chemosensory genes as compared to *G. m. morsitans* genome. This decrease, however, constitutes of a small amount of predicted genes (Macharia *et al*., 2016: Attardo *et al*., 2019). Transcriptomic quantification, orthology, enrichment and expansion of these chemosensory genes in *G. m. morsitans* has been extensively explored (Kabaka *et al*., 2020; Gakii *et al*., 2021).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study RNA Libraries

To annotate, characterize and establish functions of novel chemosensory antennal genes, *G. m. morsitans* antennal transcriptome (RNA-seq) libraries previously prepared and deposited at the Sequence Read Archive (SRA) (ncbi.nlm.nih.gov/sra) under study accession numbers PRJNA344035 were used. The RNA-seq libraries data was generated from laboratory reared colonies of *G. m. morsitans* maintained at Yale University, New Haven, USA, insectary.

3.1.1 Study insects

Prior to RNA extraction, antennae were collected from male *G. m. morsitans* colony flies maintained at Yale University insectary. The flies originated from a small population of flies originally collected from Zimbabwe. The flies were maintained at 24°C, 50 – 60% relative humidity (RH) and received defibrinated bovine blood (commercially supplied by Hemostat Laboratories, Dixon, CA, USA), via an artificial feeding system every 48h (Moloo, 1971).

3.1.2 Odorants

The δ-nonalactone repellent (98–99% pure) was sourced as racemic mixture from Sigma-Aldrich (Taufkirchen, Germany). The ε-nonalactone attractant racemic blend was not commercially available and was synthesized in the laboratory using Scheme 1 method as previously described by Gikonyo *et al.* (2002). Briefly, 1-Pyrrolidino-1-cyclopentene was allowed to react with propargylbromide in acetonitrile to create 2-propylcyclopentanone which was then converted to δ-octalactone using *m*-chloroperbenzoic acid. Then ε-nonalactone a structural variant of δ-octalactone was made by increasing the six-member carbon ring of δ-octalactone to seven. The structure of ε-nonalactone was confirmed by high-resolution mass spectrometry (HR-MS), carbon 13 nuclear magnetic resonance (^{13}C NMR), hydrogen nuclear magnetic resonance (^{1}H-NMR), and infrared (IR) spectrophotometry as previously described by Wachira *et al*. (2016). These odorants were tested using a two-choice wind tunnel at 10^{-3} dilutions in paraffin oil (1% vol/vol), a dilution previously adapted for assessment of laboratory responses of *G. m. morsitans* to odors (Chahda *et al*., 2019).

3.1.3 Assessment of male *G. m. morsitans* responses to odorants

Colony male *G. m. morsitans* teneral flies (1 - 3 days old) maintained at Biotechnology research institute insectary, Muguga, were collected for odorant response assessment. This was to affirm that the flies respond to these odorants and that the odorants elicit receptor responses in the antennae of the flies. The flies were maintained at optimum insectary conditions. Briefly, 1 - 3 days old male *G. m. morsitans* teneral flies were fed with blood meal post-eclosion and then starved for 72h to 'induce' hunger. The flies were sorted into groups of 30 flies each.

The flies of each group were released sequentially at the midpoint of the two-choice wind tunnel with the odorants suspended in cotton at one arm of the tunnel. Pure air was released uniformly from an air cylinder at 12.3 l/min to both arms of the tunnel (with odorant or without) and observations were made after three minutes. The flies were handled in the three independent replicates for each odorant and dilution (10^{-1}, 10^{-2} and 10^{-3}) or paraffin oil control (each odorant and dilution run independently) and the wind tunnel was cleaned before and after each experiment by releasing clean air through it for ten minutes. Data collected from the observations was analyzed using a Two-Way Analysis of Variance in Graphpad prism version 8.0.0 (GraphPad Software, San Diego, California USA).

3.1.4 Treatment of tsetse flies

Blood-fed 1 - 3 days old teneral male *G. m. morsitans* were collected from the colony and starved for 72hrs to 'induce' hunger and to potentially prime them into 'host seeking' physiological state. Both male and female tsetse flies require a blood meal for energy and it is expected that they use similar cues to find the mammalian host. Also, to enable us standardize the experiment, we used only male insects which were abundant in our rearing colonies. A reserved group of flies was continually fed to assess the effects of feeding on host-seeking. The starved flies were separately placed into three independent replicates each comprising of 50 flies in one-liter transparent glass jars. The two odor treatments i.e., an attractant and a repellent, and a control made of diluent paraffin oil were delivered using a strip of Whatman filter paper. Briefly, three replicates of 50 flies each for each treatment (attractant, repellent) and diluent paraffin oil and two replicates for the fed flies were placed independently in the glass jars, adding up to eleven sample jars. Then, 100µl of the treatments or the diluent paraffin oil were pipetted unto the filter paper and immediately suspended on the respective glass jars'

tops. The glass jar tops were closed with screw caps and moved into table tops under insectary conditions for five hours.

Since *G. m. morsitans* have their peak activity of blood-meal seeking in early mornings and late afternoons (Pilson and Pilson, 1967), the exposures were performed from 7:00 am for 5 hours to coincide with the morning peak activities. Antennae were extracted from each fly in each treatment. The flies were snap frozen at the end of the exposure by placing the eleven jars containing the flies in -80^{o}C freezers. In each treatment and replicate, the pairs of antenna from the tsetse fly heads were carefully hand-dissected and pooled into 1.5ml Eppendorf tubes under liquid nitrogen as described by Menuz *et al.* (2014).

3.1.5 Isolation and RNA Sequencing

Total RNA was then extracted by mechanically crushing the antennae with disposable RNAseq-free plastic pestles in TRIzol reagent (Invitrogen, Carlsbad, USA) following the manufacturer's protocol. Briefly, 1 µL of Trizol reagent and isopropanol were added in each Eppendorf tube to lysis the antennae tissue cells and then incubated for ten minutes at 4^{o}C. The antennae samples were transferred into centrifuge tubes and centrifuged at 12000rpm for ten minutes. The samples were then washed with 1 µL of 75% ethanol and vortexed before the extracted total RNA was suspended in 60 µL of RNase-free water and mixed thoroughly.

Potential genomic DNA contaminants were eliminated from the total RNA samples by digestion using TURBO DNase (Ambion life technologies, TX, USA). The extracted total RNA was mixed gently with 1µL of TURBO DNase reagent and buffer and then incubated for 30 minutes at 37^{o}C. The total RNA samples were then incubated at room temperature for 5 minutes after adding 1µL of the DNase inactivation reagent and centrifuged at 10000rpm for 1.5 minutes. A qualitative assessment of total RNA purity was done through Polymerase chain reaction method using tsetse fly specific *beta-tubulin* gene primers as documented in Bateta *et al*. (2017).

The quality and integrity of total RNA samples was verified using Agilent Bioanalyzer 2100 (Agilent, Palo Alto, CA, USA) following the manufacturer's instructions. Briefly, 1 µL of the total RNA suspended in RNase-free water was pipetted and introduced to the Agilent Bioanalyzer calibrated using RNase-free water. Only total RNA samples that had absorbance ratio (A_{260nm}/A_{280nm}) of between 1.8 – 2.0 were used in the complementary DNA synthesis.

Complementary DNA (cDNA) was then generated from the remainder (60 µL) of the pure total RNA using *Illumina* TruSeq RNA *Sample Preparation Kit (*Illumina, Hayward, CA, USA) and the cDNA (101bp paired-end read) sequenced on Illumina HiSeq 2500 at Yale University Center of Genome Analysis (YCGA), New Haven, CT, USA. A paired-end read sequencing platform was performed to facilitate subsequent accurate alignment of the reads onto reference transcripts (Nakazato *et al.,* 2013). Thus, eleven libraries consisting of three replicates each of antennae from attractant, repellent or paraffin oil control exposed flies, and two from fed flies were sequenced and all raw reads sequences deposited at the Sequence Read Archive (SRA) under study accession number PRJNA344035.

3.2 Computational analysis

3.2.1 Mapping of male *G. m. morsitans* antennae specific reads to the genome or gene-set

The quality of the reads in each transcriptome library was established using FastQC version 0.11.0.9 (Babraham Bioinformatics) software package (Andrews, 2010). The FastQC results were then used to clean (trim) and remove low quality reads from respective transcriptomes using trimmomatic software version 0.38 (Bolger *et al.*, 2014) that implemented *1)* -phred33 scale of quality scores commensurate with the RNA-Seq data quality and format and *2)* settings that permitted sequential cleaning of leading or trailing three nucleotides within 4:15 sliding window leaving at least 36 nucleotides long reads. This cleaning process generated *1)* paired reads of forward and their counterpart reverse reads surviving, *2)* unpaired (orphaned) reads where the forward or reverse reads did not survive and *3)* none, where neither forward nor reverse reads did not survive the cleaning process.

The clean surviving paired reads category from different treatments and replicates were then separately mapped onto *G. m. morsitans* transcripts version 1.9 or genome version 1.0 from VectorBase (Giraldo-Calderón *et al.,* 2015). The mapping was performed using STAR RNA-seq aligning software version 2.7.3a (Dobin *et al*., 2013) through default settings with Binary Alignment Map (BAM) output format.

3.2.2 Assessment of attractant or repellent responsive transcripts in male *G. m. morsitans* antennae

The quantification of the number of reads aligning onto each transcript in the respective BAM files was performed using Salmon software version 1.2.1 (Patro *et al*., 2017). Subsequently, differential expression of the transcripts was established by comparing the relative abundance

of the aligned reads from libraries derived from tsetse fly that were fed, exposed to attractant or repellent odor relative to control using DESeq2 software (Love *et al*., 2014).

Transcripts were considered differentially expressed if the test statistics P-value (adjusted for false detection rate; FDR) was less than 0.05 with at least a two-fold change in the difference between the fed, attractant or repellent and control treatments in either direction (up-regulated or down-regulated). The putative functional roles of the differentially expressed transcripts were derived from their annotations and their associated *D. melanogaster* orthologs in VectorBase (Giraldo-Calderón *et al*., 2015). Since the antennae are functionally specialized for olfaction, and potentially enriched with associated canonical chemosensory gene transcripts, expression profiles of these transcripts were examined separately and transcripts with at least a two-fold change in difference between the fed, attractant or repellent and control treatments were assorted.

3.2.3 Assessment of gaps in annotation of antennae expressed genes in male *G. m. morsitans*

Taking into consideration that a substantial amount of reads did not correspond to any transcripts in already annotated/existing transcripts from the mapping above, this denoted a gap in the current annotations of the *G. m. morsitans* genome. Furthermore, it supported the hypothesis that the unannotated genomic regions might be accounted for by the unmapped reads as exhibited by transcript and genome mapping differences.

Forward or reverse components of the clean surviving paired reads category from different treatments and replicates as explained above (section 3.2.2) were pooled separately. They were both mapped (paired reads) onto *G. m. morsitans* transcripts version 1.9 or genome version 1.0 from VectorBase (Giraldo-Calderón *et al.,* 2015) using Bowtie2 ultrafast short sequence reads aligning software version 2.3.5.0 (Langmead and Salzberg, 2012) with settings that also isolated unmapped reads from each mapping procedure.

Unmapped paired reads from the transcript mapping procedure (associated with the potential gap in annotation) were collected, *de novo* assembled (unmapped reads) into transcripts and the quality of the assembled transcripts assessed using the short read Trinity *de novo* assembly software 2.10.0 (Grabherr *et al*., 2011). The unmapped reads were then mapped back onto the *de novo* assembled transcripts using Bowtie2 version 2.3.5.0 to establish the proportion of reads that were revealed to be incorporated/employed in the *de novo* assembly.

The longest transcripts with open reading frames (most representatives of the respective genes) that could putatively yield peptides at least 100 amino acids long were isolated using TransDecoder software (Haas, 2018). These transcripts were queried for their putative functions/homologs in protein database UniProt release-2020-04 (The UniProt Consortium, 2019) or corresponding transcript in *G. m. morsitans* transcripts version 1.9 from VectorBase using Basic Alignment Search Tool (BLAST) analysis for protein (tBlastx) and nucleotide (Blastn) sequences, respectively. A transcript was considered *1)* a homolog of a UniProt database gene if it had an e-value < 0.001, at least 95% query coverage and 100 amino acids, and *2)* corresponding transcript of *G. m. morsitans* transcript if it had an e-value < 0.001, at least 95% query coverage and identity and length of 300 nucleotides.

Transcripts with neither homologs nor corresponding transcripts in either database were considered as novel transcripts. The longest transcripts with open reading frames were then independently used to predict novel protein-coding genes in the *G. m. morsitans* genome version 1.0 from VectorBase using MAKER computational pipeline (Campbell *et al*., 2014). This pipeline employed *ab initio* gene predictions, transcript evidence, and homologous protein evidence from UniProt/Swiss-Prot protein database (The UniProt Consortium, 2019) publicly available at https://www.uniprot.org/ (accessed on 10 June 2020).

Finally, an assessment for proportion of the longest transcripts with open reading frames that MAKER used in the prediction of protein-coding genes was performed by searching in the corresponding genes for the transcripts using BLASTn (Altschul *et al*., 1990). A *de novo* transcript with an e-value of < 0.001, and at least 95% query coverage and identity were considered correspondent to the predicted genes.

3.2.4 Annotation of the novel antennae expressed genes in male *G. m. morsitans*

The final gene models generated by MAKER software were manually curated by inspecting and refining the precise gene structure and putative function in graphical browser-based curation Apollo software platform in community VectorBase (Dunn *et al*., 2019). The major steps in the manual curation included, *1)* investigating exon/intron structure integrity and setting start and/or stop codons based on the concatenated RNA-seq evidence track as well as existing tracks in VectorBase, *2)* verifying consistency and accuracy of the curated gene models by querying them against known homologs in *D. melanogaster* within VectorBase, and *3)* internal validation and provision of stable sequence identities and adoption by VectorBase.

The improvement in the annotation depth of the genome was assessed by mapping the original concatenated reads onto a combined transcript dataset consisting of *G. m. morsitans* gene-set (version 1.9) and the newly annotated gene transcripts using Bowtie2 software version 2.3.5.0. An assessment on the proportion of the *de novo* assembled transcripts utilized and accepted by the MAKER computational pipeline in processing of the gene predictions was performed through a nucleotide search using Basic Alignment Search Tool (BLASTn) of the *de novo* assembled transcripts as query against the newly curated genes as the subject. Novel genes, i.e., the predicted protein-coding genes, were isolated from among the newly annotated genes by performing a nucleotide search using BLASTn of the annotated genes against *G. m. morsitans* gene-set (Version 1.9).

Identification of the putative functions of these novel genes, i.e., those without corresponding gene in the *G. m. morsitans* gene-set (version 1.9) was done by, *1)* identifying homologs in Uniport/Swiss-prot protein database (The UniProt Consortium, 2019) using BLASTp (Altschul *et al*., 1990) search against UniProt protein database (accessed on 10 June 2020), accepting hits with e-value < 0.001 as significantly homologous, *2)* identifying protein domains and GO terms associated with the predicted protein-coding genes using standalone InterproScan software version 5.52-86.0 (Jones *et al*., 2014), and 3) identifying orthologs in *Musca domestica*, *Anopheles gambiae* mosquito genomes obtained from VectorBase release 53 and *D. melanogaster* genome from FlyBase (Larkin *et al*., 2021) using default settings in OrthoFinder software version 2.5.4 (Emms and Kelly, 2019).

3.2.5 Assessment of differential responses of the novel antennae expressed genes to *G. m. morsitans* attractant or repellent odor cue

Assessment of the differentially expressed novel genes in response to attractant (ε-nonalactone), repellent (δ-nonalactone), or feeding was performed using RNA-Seq by Expectation Maximization (RSEM) – EBSeq pipeline (Li and Dewey, 2014). Briefly, RSEM transcript references for annotated novel gene transcripts were built and then separately mapped the clean paired reads and replicates from each of the treatment libraries (fed, unfed, exposed to repellent or attractant) individually onto the novel transcripts using Bowtie2 version 2.3.5.0.

Estimated read counts for respective transcripts or their isoforms from each library and replicate were then extracted. Subsequently, a count matrix was generated from comparisons

of the read counts from attractant, repellent, or fed treatment libraries to the unfed library (control) using the RSEM–EBSeq pipeline. This analysis thus generated a list of relative expression levels of each transcript/isoform in the treatments relative to the control. The transcripts/isoform were considered as differentially expressed if there was a two times post-fold change and false discovery rate (FDR) correction of <0.05. The annotation procedure is summarized in Figure 3.1.

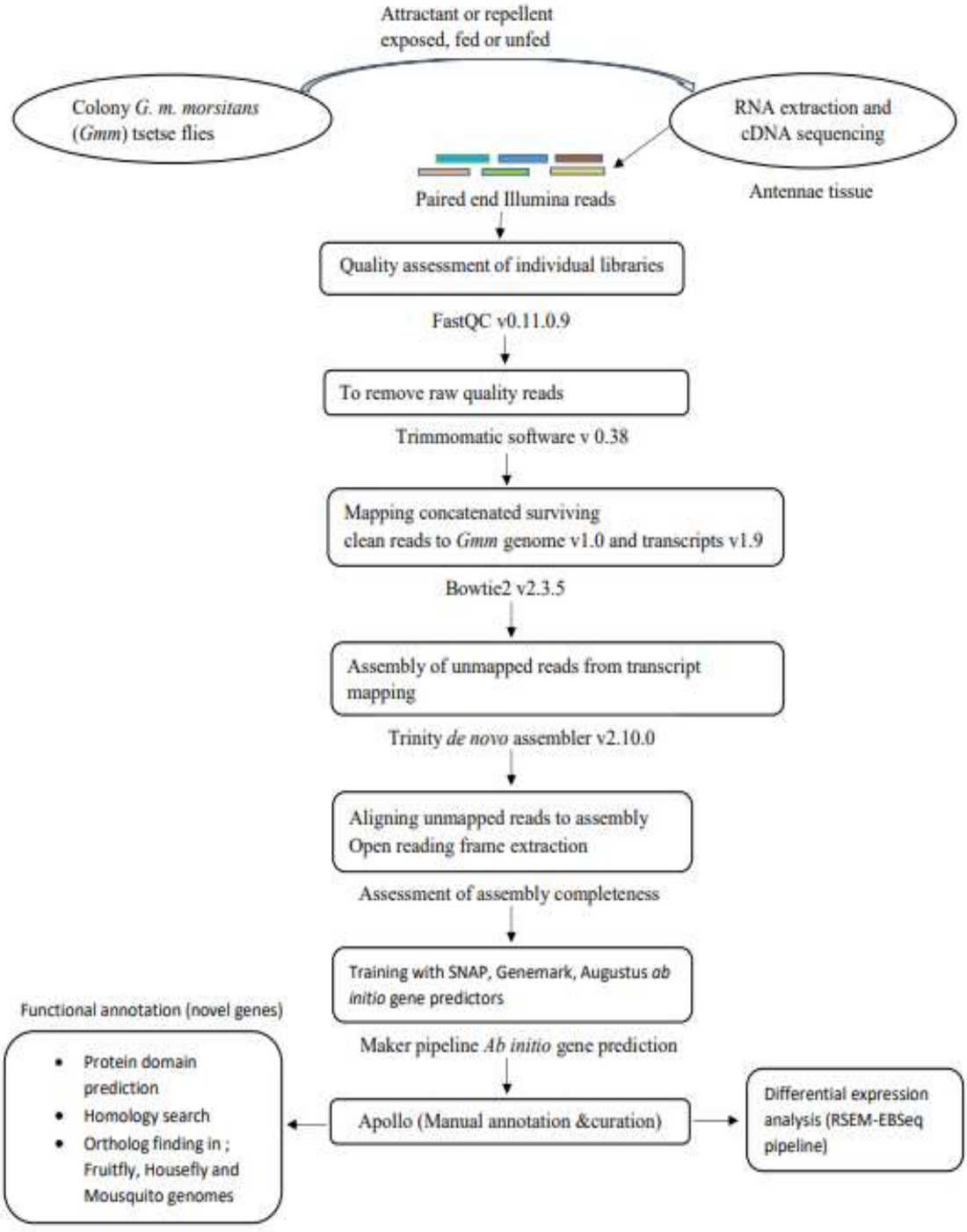

Figure 3.1. Summary of the steps followed sequentially in the annotation of the novel *G. m. morsitans* antennal genes through *de novo* assembly and manual curation.

3.3 Validation of the global differentially expressed antennal transcripts

To validate the computational analysis, quantitative Real-Time PCR of the global differentially expressed transcripts (section 3.2.3) was performed using gene specific primers.

3.3.1 RNA extraction

Total RNA was extracted from antennae of 1 - 3 days old teneral male *G. m. morsitans* that were reared at Biotechnology research institute and exposed to attractant or repellent of 10^{-3} dilution, or diluent paraffin oil as previously described (section 3.1.3). The fed group was not included due to absence of chemosensory genes that were induced or suppressed in response to feeding as compared to the control.

Briefly, the flies were snap frozen at the end of exposure by placing the jars containing the flies in -80^{o}C freezers. The pairs of antennae from the tsetse fly heads were carefully hand-dissected in each treatment and replicate (five replicates were used) using liquid nitrogen as described by Menuz *et al.*, (2014). Total RNA was extracted from the antennae samples using the Isolate II RNA extraction kit (Bioline, UK) following the manufacturer's protocol. Briefly, the antennae samples were homogenized and mixed with 350 µL of the lysis buffer and then centrifuged at 12000rpm for 1 minute. 350 µL of 70% was then added to the samples followed by an equal amount of membrane desalting buffer. Contaminant DNA was digested with 950 µL of DNase. The samples were sequentially washed with 200 µL, 600 µL and 250 µL of RW1, RW2 and wash buffer respectively. The extracted RNA was then eluted in 60 µL of RNase-free water.

In Each treatment, the biological replicates consisted of tubes that each had 50 antennae (from 25 flies) pooled together. The quality and integrity of the total RNA was verified using a Nanodrop 2000/2000c Spectrophotometer (Thermo Fischer Scientific, Wilmington, USA). Only total RNA samples that had an absorbance ratio (A_{260nm}/A_{280nm}) of between 1.8 – 2.0 were used in the downstream processes. For long term storage, the total RNA was stored at - 80°C.

3.3.2 Complementary DNA synthesis

Pure extracted RNA samples were used for cDNA synthesis. Prior to cDNA synthesis, Dnase treatment was done on the extracted total RNA to remove potential genomic DNA carry-over and contaminants. This was followed by a two-step reverse transcription PCR using the iScript^{TM} cDNA synthesis kit (Bio-Rad Laboratories, Inc, Hercules, CA, USA) following the manufacturer's instructions. A 20 µL reaction mix consisting of 19 ng/µL RNA, 1 µL iScript reverse transcriptase, 4 µL 5X iScript reaction mix (pre-blended with oligo (dT) and random hexamer primers) and nuclease free water was prepared. The mixture was down spun then RT-PCR set up in the Mastercycler gradient Nexus thermal cycler (Eppendorf, Hamburg, Germany). The following cycling conditions were used: priming for 5 min at 25°C followed by reverse transcription for 20 min at 46 °C then RT inactivation for 1 min at 95 °C. The cDNA generated was stored at -20^{o}C till used.

3.3.3 Quantitative real-time polymerase chain reaction

Nine genes transcripts from among the global differentially expressed transcripts were randomly selected for validation and specific primers (Table 3.1) were designed using primer3plus protocol to probe them (Steve and Helen, 2000). These markers were used to assess the expression levels of the selected genes and this was determined individually through quantitative RT-PCR. The reaction mix consisted of uniform 2 µg cDNA template, separately amplified in three independent replicates with 5 µL of iTaq Universal SYBR Green super mix (Bio-Rad Laboratories, Inc, Hercules, CA, USA) in presence of 0.5 picomoles specific primers for the various genes (Table 3.1). The reactions were performed in Strategene MX3005P, real time qPCR machine (Agilent Technologies, California, USA). The *G. m. morsitans* housekeeping genes *beta-tubulin* or GAPDH were used as reference genes. The crossing threshold values were recorded for all the sample reactions and subsequently used to quantify products of amplification using comparative Ct ($2^{-\Delta\Delta Ct}$) method (Livak and Schmittgen, 2001).

3.4 Expression pattern analysis of the selected differentially expressed transcripts

The Ct values for each selected gene in each treatment and replicate obtained from the real-time qPCR run were analyzed with the comparative quantification delta Ct method to quantify the expression levels of each gene transcript based on the number of amplified transcripts.

Table 3.1 Primers used for quantitative PCR validation (RT-qPCR) of global differentially expressed antennal transcripts in male *G. m. morsitans*

Gene/Transcript	Primer sequence 5' - 3'	Annealing temp (°C)
Tsetse fly β -tubulin	CCATTCCCACGTCTTCACTT	60
Tsetse fly β -tubulin-rev	GACCATGACGTGGATCACAG	
GAPDH	CTGATTTCGTTGGTGATACT	55
GAPDH -rev	CCAAATTCGTTGTCGTACCA	
GMOY005226-RA	CAGCAATGGCCGAAAAGGAT	58
GMOY005226-RA-rev	ACGCAAGTCATACGACAGCA	
GMOY008016-RA	AGGAATTTGTCGTTGGCACA	60
GMOY008016-RA-rev	TATCAGATCGGTGCAGCAGG	
GMOY007896-RA	TCGGCTCAATGCGAATACCC	58
GMOY007896-RA-rev	AAGGACGTATGTGCCAGCAA	
GMOY009893-RA	GGGCTAAACGTACCCCGAAA	53
GMOY009893-RA-rev	GTGTAGACGGCGCTATCAGT	
GMOY003789-RA	AGAGACTGCGTGGAAGGTTG	51
GMOY003789-RA-rev	CCGCCTTAAAAGTCATGCCG	
GMOY001391-RA	AGGCATTCCCAGCTAACACC	57
GMOY001391-RA-rev	ACAGTTCAAAAAGCGTCGGC	
GMOY003590-RA	AGGCGGAACCGATGGTAATC	57
GMOY003590-RA-rev	TGTGGCCCAGAAAAACCCTT	
GMOY006073-RA	ATGATGACACCACACGGTCC	57
GMOY006073-RA-rev	TGTCACGGCCATGCTAAGAG	
GMOY002035-RA	GCGCATCTACCGCAAAACAT	60
GMOY002035-RA-rev	CATGAGGAATCGCCGTCACT	

Rev- denotes the reverse counterpart of the primers for the housekeeping genes or test gene transcripts

CHAPTER FOUR

RESULTS

4.1 Odorants induce distinct physical responses in *G. m. morsitans* flies

The assessment of *G. m. morsitans* response to odorant treatments (attractant, repellent, and negative control) at different levels of concentration (10^{-1}, 10^{-2}, and 10^{-3}) using a two-choice wind tunnel showed no interaction effect ($F_{4, 18} = 0.1132$, $p = 0.9762$). The data however revealed that the odorant effects were statistically different ($F_{2, 18} = 11.41$, $p < 0.001$). The attractant ε-nonalactone had high number of flies compared to the repellent δ-nonalactone, and surprisingly the number of insects in the negative control (paraffin oil) was almost similar to those on the attractant (Figure 4.1). Odorant concentrations did not show differences ($F_{2, 18} = 0.3396$, $p = 0.7165$) in inducing tsetse fly responses (Figure 4.1), conveying that diluting the odorants did not determine the observed physical changes in tsetse fly responses.

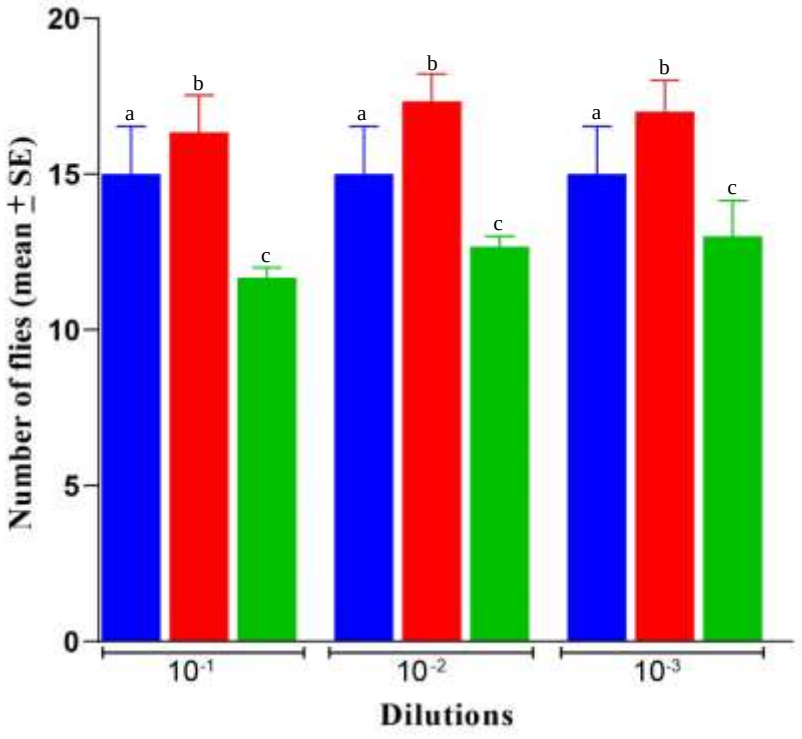

Figure 4.1. Assessment of *G. m. morsitans* response to attractant ε-nonalactone (red) or repellent δ-nonalactone (green) as compared to diluent paraffin oil control (blue) using flies

means from a two-choice wind tunnel. Bars marked with the same letter are not statistically significantly different.

4.2 Antennae-specific transcriptomes and annotation gaps in the *G. m. morsitans* genome

Approximately 588 million reads were obtained as a result of sequencing all the eleven RNA libraries from the antennae of adult male *G. m. morsitans*. The libraries constituted between 23 and 74 million reads as shown in the second and third replicates of the fed and control, respectively (Table 4.1).

Table 4.1 Quality statistics for read libraries from male *G. m. morsitans* under different treatments

Treatment	Rep*	Input§	Paired	Forward only	Reverse only
ε-nonalactone	1	31,613,354	27,542,555	3,191,259	226,061
	2	54,614,728	53,014,279	1,277,959	294,485
	3	69,839,051	69,838,974	77	0
δ-nonalactone	1	39,779,002	35,024,794	3,075,953,	525,061
	2	46,005,014	41,819,337	3,363,603	356,507
	3	73,706,273	73,706,252	21	0
Control	1	54,871,262	49,243,856	4,772,241	364,879
	2	68,515,983	64,428,721	3,360,752	416,877
	3	74,587, 703	74,587,696	7	0
Fed	1	51,370,991	48,023,641	2,575,519	461,537
	2	23,086,651	20,978,730	1,599,080	196,793
Total		587,990,012	558,208,835	20,140,518	2,842,200

§ - Total amount on reads obtained from sequencing antennae RNA of teneral male *G. m. morsitans* exposed to an attractant (ε-nonalactone), repellent (δ-nonalactone), diluent paraffin oil control or fed; * - Replicate for each treatment. Reads of good quality based on trimmomatic software were as paired (forward and reverse counterparts survived), forward only (reverse counterpart dropped) and reverse only (forward counterpart dropped). The attractant, repellent or paraffin control treatments were performed in three replicates while the fed had only two replicates.

More than 97% (least computed percentage from the first replicate of the repellent δ-nonalactone) of the reads in these libraries passed the quality control test as clean paired

(87.12% - 100%) or unpaired (0% -10.91%) reads (Figure 4.2). These results showed that the obtained sequences were of very good quality.

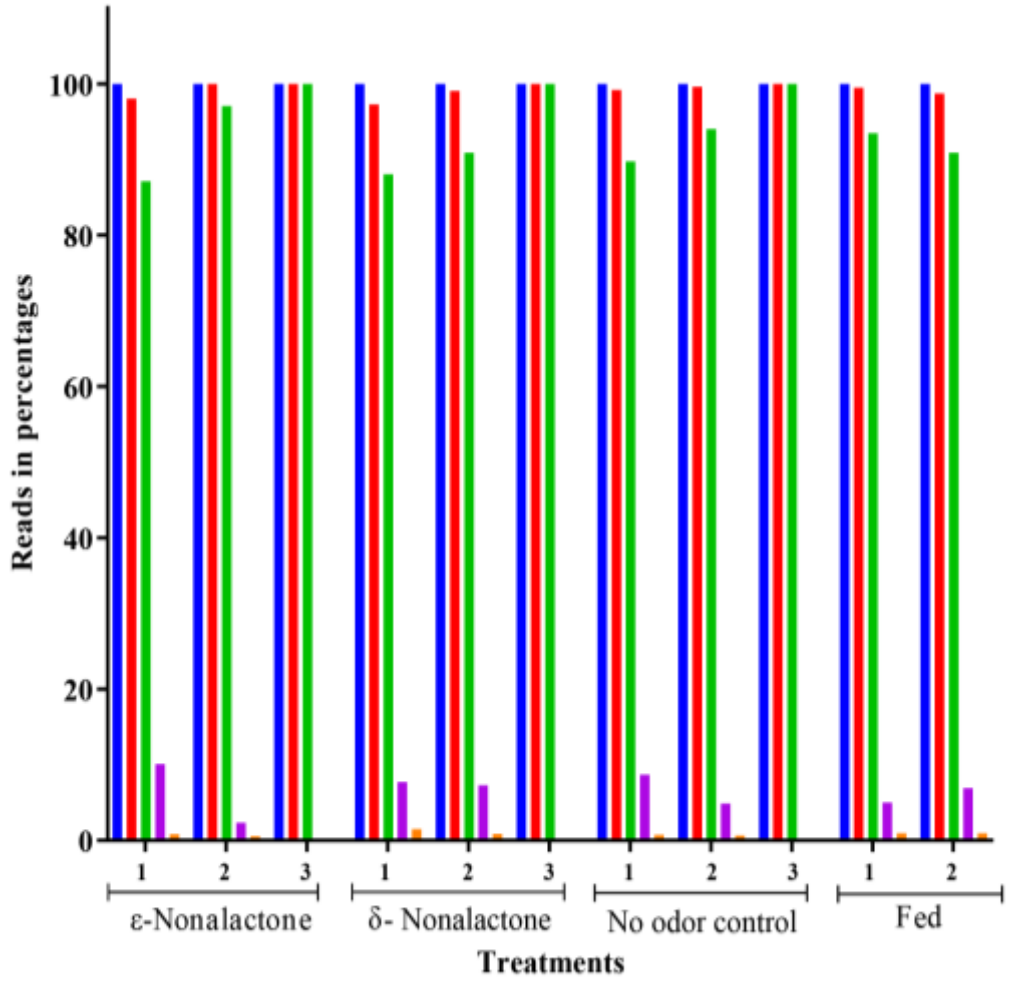

Figure 4.2. Quality assessment statistics of individual male *G. m. morsitans* RNA-Seq antennal libraries obtained from fed/unfed teneral male *G. m. morsitans* exposed/unexposed to attractant (ε- nonalactone) or repellent (δ-nonalactone). The total percentage of the clean reads (presented in red) comprised of both paired and unpaired reads. Paired reads (Both surviving, presented in green) had forward and reverse counterparts pass quality check while unpaired had either the forward (purple) or reverse (orange) reads only pass quality check. All percentages were computed against the total input read percentages (blue). The attractant, repellent, unfed (No odor control) treatments were performed in three replicates marked 1,2 and 3 while the fed had only two replicates marked 1 and 2.

More than 81% (presented by the first replicate of ε- nonalactone) of the clean paired reads from each library mapped onto the genome, of which between 78.12% (from first replicate of ε- nonalactone) and 91.58% (from second replicate of the fed) mapped uniquely and 3.29% (first replicate of control) and 5.01% (second replicate of ε- nonalactone) mapped to multiple loci (Table 4.2).

Table 4.2 Genome mapping statistics of the individual read libraries from male *G. m. morsitans* antennae.

			Total mapped		Uniquely mapped		Multiple mapped	
Treatments[§]	**Rep[*]**	**Total reads**	**Counts**	**%**	**Counts**	**%**	**Counts**	**%**
ε-nonalactone	1	27542555	22483174	81.64	21514983	78.12	968191	3.52
	2	53014279	45548088	85.91	42890747	80.90	2657341	5.01
	3	69838974	62176304	89.03	59360627	85.00	2815677	4.03
δ-nonalactone	1	35024794	32139288	91.76	30867011	88.13	1272277	3.63
	2	41819337	38109390	91.13	36675914	87.70	1433476	3.43
	3	73706252	69751069	94.64	66811273	90.65	2939796	3.99
Control	1	49243856	44171626	89.70	42552779	86.41	1618847	3.29
	2	64428721	60705527	94.23	58550046	90.88	2155481	3.35
	3	74587696	70781511	94.90	67912957	91.05	2868554	3.85
Fed	1	48023641	45677563	95.11	43982451	91.58	1695112	3.53
	2	20978730	18776090	89.50	18024501	85.92	751589	3.58

§ - Teneral male *G. m. morsitans* either fed, exposed to an attractant (ε-nonalactone), repellent (δ-nonalactone) or diluent paraffin oil control and the antennae obtained for RNA extraction and sequencing; ***** - Replicates for each treatment. Reads obtained from all sequenced libraries were cleaned and mapped unto *G. m. morsitans* genome version. 1.0. Uniquely mapped reads aligned to one genomic region while the multiple mapped reads aligned to more than one genomic region of the genome. The total mapping reads comprises of all reads aligning to the genome over the total reads (clean) from trimmomatic in each respective library. The paraffin control, attractant or repellent treatments were performed in replicates while the fed had only two replicates. All percentages are calculated in reference to the total clean reads.

On the other hand, less than 57.88% (presented by first replicate of the fed) of the clean paired reads from each library mapped onto the *G. m. morsitans* gene set sequences, of which between 30.05% (from second replicate of ε-nonalactone) and 52.60% (from first replicate of the fed) mapped uniquely while 3.15% (third replicate of the ε-nonalactone) and 5.28% (first replicate

of the fed) mapped to multiple loci (Table 4.3). All the replicates from the repellent and fed libraries showed a >50% total mapping to the gene– set with fewer reads aligning to multiple sequences.

Table 4.3 Gene- set mapping statistics of the individual read libraries from male *G. m. morsitans* antennae.

			Total mapped		Uniquely mapped		Multiple mapped	
Treatments[§]	**Total reads**	**Rep[*]**	**Counts**	**%**	**Counts**	**%**	**Counts**	**%**
ε-nonalactone	27542555	1	11657423	42.33	10493751	38.10	1163672	4.22
	53014279	2	17882096	33.73	15928288	30.05	1953808	3.69
	69838974	3	24466055	35.03	22264326	31.88	2201729	3.15
δ-nonalactone	35024794	1	18846534	53.81	17069696	48.74	1776838	5.07
	41819337	2	23166503	55.40	20973148	50.15	2193355	5.24
	73706252	3	39704940	53.87	36106178	48.99	3598762	4.88
Control	49243856	1	20177516	40.97	18247922	37.06	1929594	3.92
	64428721	2	36689029	56.95	33439445	51.90	3249584	5.04
	74587696	3	39083878	52.40	35755886	47.94	3327992	4.46
Fed	48023641	1	27797532	57.88	25259895	52.60	2537637	5.28
	20978730	2	11334835	54.03	10307865	49.13	1026970	4.90

[§] - Teneral male *G. m. morsitans* either fed, exposed to an attractant (ε-nonalactone), repellent (δ-nonalactone) or diluent paraffin oil control and the antennae obtained for RNA extraction and sequencing; [*] - Replicates for each treatment. Reads obtained from all sequenced libraries were cleaned and mapped unto *G. m. morsitans* gene- set version 1.9. Uniquely mapped reads aligned to a single gene sequence while those mapped to multiple loci aligned to multiple gene sequences in the gene-set. The total mapped reads are a percentage calculated from the sum of all aligning reads over the total reads (clean) obtained from trimmomatic analysis in each respective library. The paraffin control, attractant or repellent treatments were performed in replicates while the fed had only two replicates. All percentages are calculated in reference to the total clean reads.

More than 42.12% of the clean paired reads from each library mapped onto the *G. m. morsitans* genome but not onto the gene set. However, pooling of the forward or reverse clean surviving paired reads category from different treatments and replicates into a single library of either improved the preliminary statistics. Using the enriched library, more than 96% of the clean paired reads mapped onto the *G. m. morsitans* genome but only 54.79% mapped onto the gene set sequences leaving 45.21% as unmapped (described herein as divergent) (Figure 4.3)

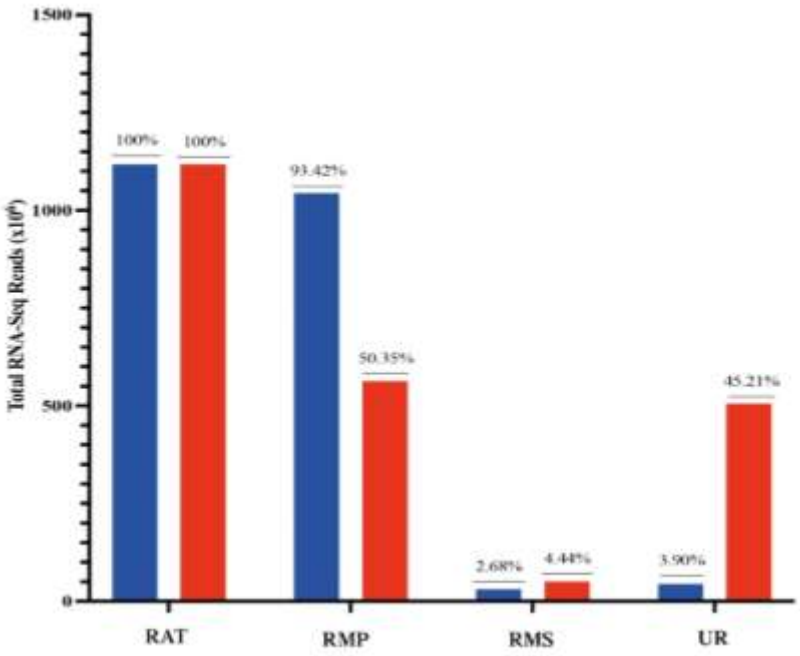

Figure 4.3. Mapping statistics of concatenated RNA-Seq library from male *G. m. morsitans* antennae. Each library (forward and reverse) was mapped to *G. m. morsitans* genome version 1.0 (Genome) or gene-set version 1.9 (Transcripts) from Vectorbase. Reads mapped in pairs (RMP) had both forward and reverse read align to a genomic region and gene sequence while reads mapping singly (RMS) had either forward or reverse read align to a genomic region and gene sequence. The unmapped reads (UR) didn't correspond to any genomic region of the genome or gene sequence in the gene- set. Statistics of the genome (blue) and gene- set (red) mapping were in reference to the total input clean reads -reads after trimming (RAT).

The mapping statistics established the unmapping reads as being of *G. m. morsitans* origin and not contaminants or of other organisms, while potentially revealing the gap in the annotations of active genes in the genome. The less than 4% of the clean paired reads that did not map onto the genome from the pooled library potentially represent reads of unknown origin such as those of symbionts or pathogens in the fly population.

4.3 The *G. m. morsitans* antennae genes are responsive to repellent or attractant odorant

Mining of differentially expressed genes in the existing *G. m. morsitans* gene set revealed 15 gene transcripts as differentially and significantly expressed in response to attractant or repellent, of which 87% (13 out of 15) were in response to the attractant (Table 4.4). The

attractant induced four gene transcripts, namely coat protein epsilon, cyclin-dependent kinase, odorant receptor 45 and a hypothetical protein, and suppressed nine gene transcripts, namely amalgam, hemolectin, regulatory particle triple-A ATPase 4-related, vesicular monoamine transporter, two scavenger receptor class A, member 5 and three hypothetical proteins. Among these, coat protein epsilon and a hypothetical protein were the most induced and suppressed respectively. The δ-nonalactone significantly induced expressions of two (homogentisate 1,2-dioxygenase and a hypothetical protein) transcripts, among which homogentisate 1,2-dioxygenase was the most expressed (Table 4.4).

Further assessment of differentially expressed (fold change > 2 or < -2) chemosensory gene transcripts revealed putative antennal induction of seven (Or67d, Clumsy, Ir60a, Gr2a, Gr28b, Obp83c-d and Obp19b) and suppression of three (Or83a, Or45b, Ir84a and Obp8a) *D. melanogaster* chemosensory transcript orthologs following male *G. m. morsitans* exposure to attractant odor. Similarly, repellent exposure putatively induced four (Or7a, Obp19b, Obp19d and Phk-3) and suppressed six (Or33a-c, Or83a, Ir84a, Clumsy, Gr66a and Obp83g) *D. melanogaster* chemosensory transcript orthologs in the G. *m. morsitans* (Table 4.5).

Table 4.4 Global differentially expressed antennal genes transcripts in male *G. m. morsitans* exposed to repellent or attractant odor

Odor	*G. m. morsitans* ID*	VectorBase Annotation*	Fold Change (log_2)[§]	P-Value	FDR P-Value
ε-nonalactone	GMOY010096-RA	Coat Protein epsilon	1.54	p< 0.001	0.0490
	GMOY005226-RA	Cyclin-dependent kinase	1.09	p< 0.001	0.0126
	GMOY008016-RA	Hypothetical protein	1.48	p< 0.001	0.0490
	GMOY007896-RA	Odorant receptor 45	1.15	p< 0.001	0.0453
	GMOY009893-RA	Amalgam	-1.55	p< 0.001	0.0490
	GMOY003789-RA	Hemolectin	-1.05	p< 0.001	0.0016
	GMOY006386-RA	Hypothetical protein	-1.45	p< 0.001	0.0453
	GMOY001677-RB	Hypothetical protein	-10.13	p< 0.001	0.0037
	GMOY000172-RB	Hypothetical protein	-22.89	p< 0.001	0.0002
	GMOY012233-RA	Regulatory particle triple-A ATPase 4-related	-10.86	p< 0.001	0.0034
	GMOY001391-RA	Scavenger receptor class A, member 5	-1.02	p< 0.001	0.0490
	GMOY003590-RA	Scavenger receptor class A, member 5	-1.26	p< 0.001	0.0453
	GMOY010623-RA	Vesicular monoamine transporter	-1.77	p< 0.001	0.0037
δ-nonalactone	GMOY006073-RA	Homogentisate 1,2-dioxygenase	1.52	p< 0.001	0.0020
	GMOY002035-RA	Hypothetical protein	1.08	p< 0.001	0.0112

* - Transcript ID in VectorBase (Giraldo-Calderón et al., 2015); FDR – False Detection Rate corrected; ε-nonalactone –Tsetse fly attractant; δ-nonalactone – Tsetse fly repellent; § - Responses in male *G. m. morsitans* to ε-nonalactone or δ-nonalactone were assessed in relation to their respective no- odor paraffin controls. The fold-changes were computed as a ratio of transcript expressed due to odor exposures relative to similar expressions in their respective controls. Consequently, positive and negative values connote respective odor induced or suppressed transcripts relative to control.

Table 4.5 Differentially expressed chemosensory gene transcripts in male *G. m. morsitans* exposed to repellent or attractant odor.

		Fold Change (log_2)[§]			
Chemosensory Gene Family	***G. m. morsitans* ID***	**ε-nonalactone**	**δ-nonalactone**	**VectorBase Annotation***	***D. melanogaster* Orthologs***
Odorant receptor	GMOY012193-RA	-0.69	1.07	Or8	Or7a
	GMOY012018-RC	-0.49	-1.55	Or5	Or33a-c
	GMOY011399-RA	-2.30	-1.65	Or21	Or83a
	GMOY009271-RA	-2.13	-0.97	Or12	Or45b
	GMOY007896-RA	1.15	0.27	Or45	Or67d
Ionotropic receptor	GMOY008188-RA	0.79	-2.29	-	Ir84a
	GMOY006490-RA	1.55	-1.13	GluR-Clumsy1	Clumsy
	GMOY002248-RA	1.39	0.47	Ir60a	Ir60a
	GMOY002585-RA	-1.16	-0.14	-	Ir84a
Gustatory receptor	GMOY011903-RA	1.28	0.69	Gr	Gr2a
	GMOY006209-RA	1.67	0.29	Gr	Gr28b
	GMOY004207-RA	0.03	-1.45	Gr	Gr66a
Odorant binding protein	GMOY005548-RA	2.03	0.17	Obp7	Obp83c-d
	GMOY001476-RA	-1.22	0.28	Obp22	Obp8a
	GMOY005550-RA	-0.86	-1.61	Obp11	Obp83g
	GMOY006522-RA	1.01	1.01	Obp19	Obp19b
	GMOY005400-RA	0.14	1.14	Obp19d	Obp19d
Chemosensory protein	GMOY010874-RA	0.96	1.81	Csp4	Phk-3

Or- Odorant receptor; Ir- Ionotropic receptor; Gr – Gustatory receptor; Obp – Odorant binding protein; Csp – Chemosensory protein; Phk - Pherokine'-' – Hypothetical protein; * - Transcript ID in VectorBase (Giraldo-Calderón et al., 2015); ε-nonalactone -Tsetse fly attractant; δ-nonalactone – Tsetse fly repellent; [§] - Responses in male *G. m. morsitans* to ε-nonalactone or δ-nonalactone were assessed in relation to their respective no- odor paraffin controls. The fold-changes were computed as a ratio of transcript expressed due to odor exposures relative to similar expressions in their respective controls. Consequently, positive and negative values connote respective odor induced or suppressed transcripts relative to control

4.4 The unmapped divergent reads are associated with antennae-specific novel genes in the male *G. m. morsitans*

The *de novo* assembly of the unmapped paired reads yielded 213,184 genes contigs comprising of 311,970 transcripts. The assembly had a GC content of 34.18% and contig N50 quality statistic of 1,445 for all genes and 903 for the longest transcripts (Table 4.6).

Table 4.6 Trinity transcriptome assembly quality assessment

	Statistics	
Attribute	**All Transcripts**	**Longest Isoform***
Genes	213184	-
Transcripts	311970	-
GC %	34.18	-
Contig N50	1445	903
Median contig length	478	367
Average contig length	857.16	626.51

*-Statistic based on the longest transcript for each single gene.

At least 96.15% (RMP + RMS) of the unmapped paired reads mapped back onto the *de novo* assembly (Figure 4.4). Using the TransDecoder software, 72,428 longest transcripts with ORFs of at least 300 nucleotides long from the 311,970 transcripts were isolated. 18,522 of the 72,428 transcripts, representing 25.57%, had unique corresponding homologs with 5,252 hits from the non-redundant UniProt database (Appendix 1). Similarly, 16,427 representing 22.68% transcripts had hits to their corresponding 4,887 counterparts in the G. *m. morsitans* gene-set query database, among which 2,374 (48.58%) were functionally characterized (Appendix 2). These searches revealed that some of the *de novo* assembled transcripts associated with the unmapped reads (gaps in genome annotation) were potentially linked to functionally annotated genes and others were novel without corresponding transcripts/genes in the databases.

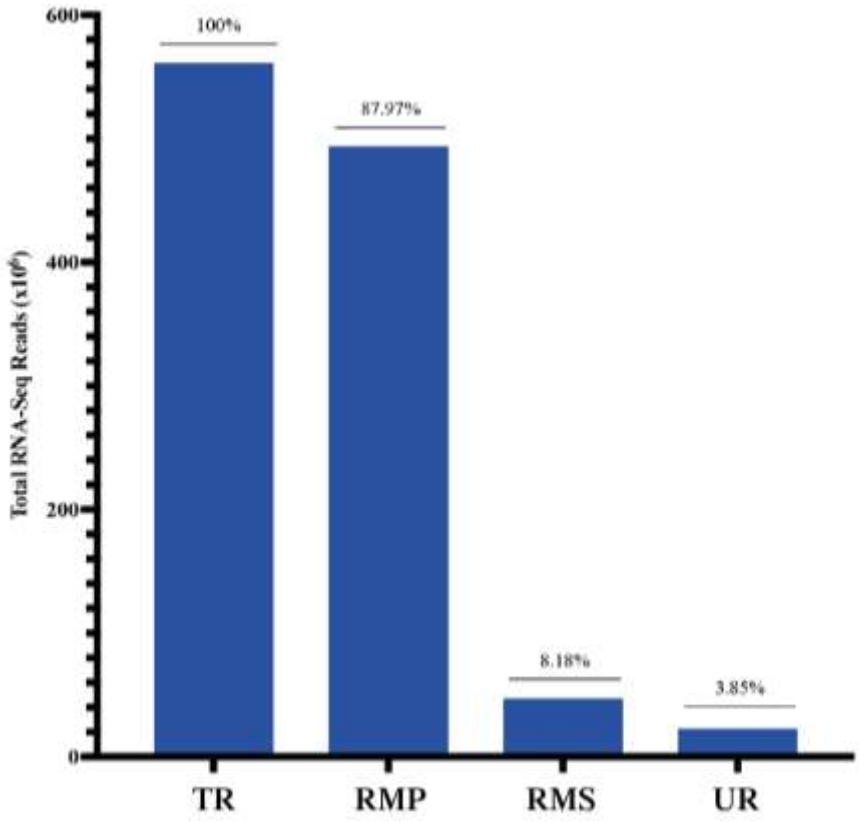

Figure 4.4. Quality assessment of read representation in *de novo* assembled antennal transcripts from male G. *m. morsitans* RNA-Seq antennal libraries. TR - Total unmapping reads from gene- set mapping that were used in *de novo* assembly. Reads mapped in pairs (RMP) had both forward and reverse reads mapped while reads mapped singly (RMS) had either forward or reverse read mapped to the *de novo* assembled transcript sequences. The unmapped reads (UR) didn't correspond to any transcript sequence.

Additionally, among the 25.57% of the longest transcripts that had unique corresponding homologs from the non-redundant UniProt database, they comprised of transcripts homologous to proteins of the chemosensory gene families. These proteins were of odorant receptor (Or30a, Or49b, Or33b, Or67d, Or74a, Or7a, Or85b, and Orco), ionotropic receptor (Ir75a and Ir93a), gustatory receptor (Gr21a and Gr63a) and odorant degrading enzymes (CyP450) gene families.

Using MAKER prediction pipeline under default settings, a total of 1,333 gene models were predicted, supported by 214,835 (68.86%) of the associated assembled transcripts - suggesting that almost 30% of the transcripts input as prediction evidence were excluded. From these

models, a total of 983 gene models were manually curated using the Apollo annotation tool as valid structural and putative functional genes. However, only 592 gene models had complete structures, of which 202 were categorized as novel genes and the remaining 390 genes were curated as modifications of already existing genes at the VectorBase. Among these 390 genes, 167 (42.82%) had their putative functions described while the remaining 223 had no known functional description at the VectorBase. Further functional characterization of the 223 genes was not done as the focus of this study was on the novel genes only.

A homology search in OrthoFinder revealed that 94 genes (46.53%) of novel genes had orthologs in at least one of the target dipteran genomes (*M. domestica*, *An. gambiae* and *D. melanogaster*) (Appendix 3), with some orthologs shared among the genomes (Figure 4.5). Five of these genes had unique orthologs specific in *M. domestica* genome but not to *An. gambiae* or *D. melanogaster* genome.

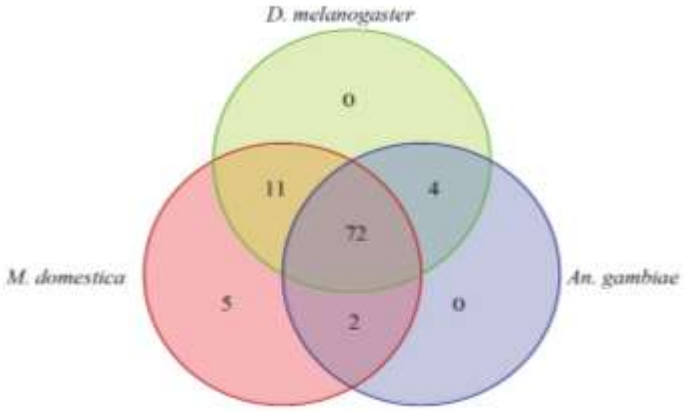

Figure 4.5. Orthologs of novel annotated male G. *m. morsitans* antennae genes in *D. melanogaster*, *M. domestica* and *An. gambiae* genomes. Majority of the genes (72) were shared among the three genomes.

In addition, the analysis against the UniProt database established that 88 (43.56%) of the novel genes had known homologs (Appendix 4). Functionally, an examination of the gene ontology characterizations of the homologs and orthologs linked to the novel genes, revealed that they are associated with molecular and biological processes such as oxidative phosphorylation, protein synthesis, transcription and translation regulation, detoxification, carbohydrates metabolism, embryogenesis, male courtship regulation, neural cell adhesion, metal ion binding,

G protein-coupled receptor signaling, memory development, immunity induction and protein degradation among others (Appendix 3 and 4).

Putative functional characterization of the curated novel genes revealed no direct association to any of the chemosensory families which might be attributed to insufficient supporting evidence for respective models hence missing the threshold. Given that a majority of the novel genes (>53%) lacked similarity to any known homolog or ortholog, it appears that these novel genes are tsetse fly-specific hence syncing with initial analyses of the *de novo* assembled transcript above.

The domain analysis of the novel gene residues in the InterproScan database revealed conserved segments related to protein kinase domain, autophagy-related protein C terminal, nitrogen permease regulator 2, beta-acetyl hexosaminidase like, RNA recognition motif, Ubiquitin family, ribosomal protein S15, cyclophilin type peptidyl-prolyl *cis-trans* isomerase and Calcium-binding domains (Appendix 5). The details of the genes whose annotations were improved are presented in Appendix 6. These genes are putatively associated with cellular or molecular functions that include protein transport, metal ion binding, neural signaling, oxidative regulation, cell degradation, gene expression regulation, response to environmental changes and olfactory roles such as odorant reception, gustatory responses and protein degradation.

4.5 The novel genes are differentially expressed in the antennae of male *G. m. morsitans* in response to attractant and repellent odors

Assessment of differentially expressed genes revealed significant up-regulation (3.5-fold change) of GMOY014237.R1396 gene (one of the annotated novel genes without a corresponding homolog or ortholog) in response to attractant exposure relative to the no-odor control group (presented with the diluent paraffin oil) (see Appendix 3). Other novel genes that were moderately up-regulated by at least 1.2-folds included 1) GMOY014112.R1263 and GMOY014071.R1219 both of which responded to the attractant and potentially associated with protein degradation and ribosomal RNA processing, and 2) GMOY014158.R1315 inducted by the repellent and is associated with putative organ development functions (Appendix 3). However, these inductions were not statistically significant (FDR was >0.05).

Other novel genes marginally up-regulated or down-regulated in response to the odorants as compared to the paraffin oil control (Figure 4.6).

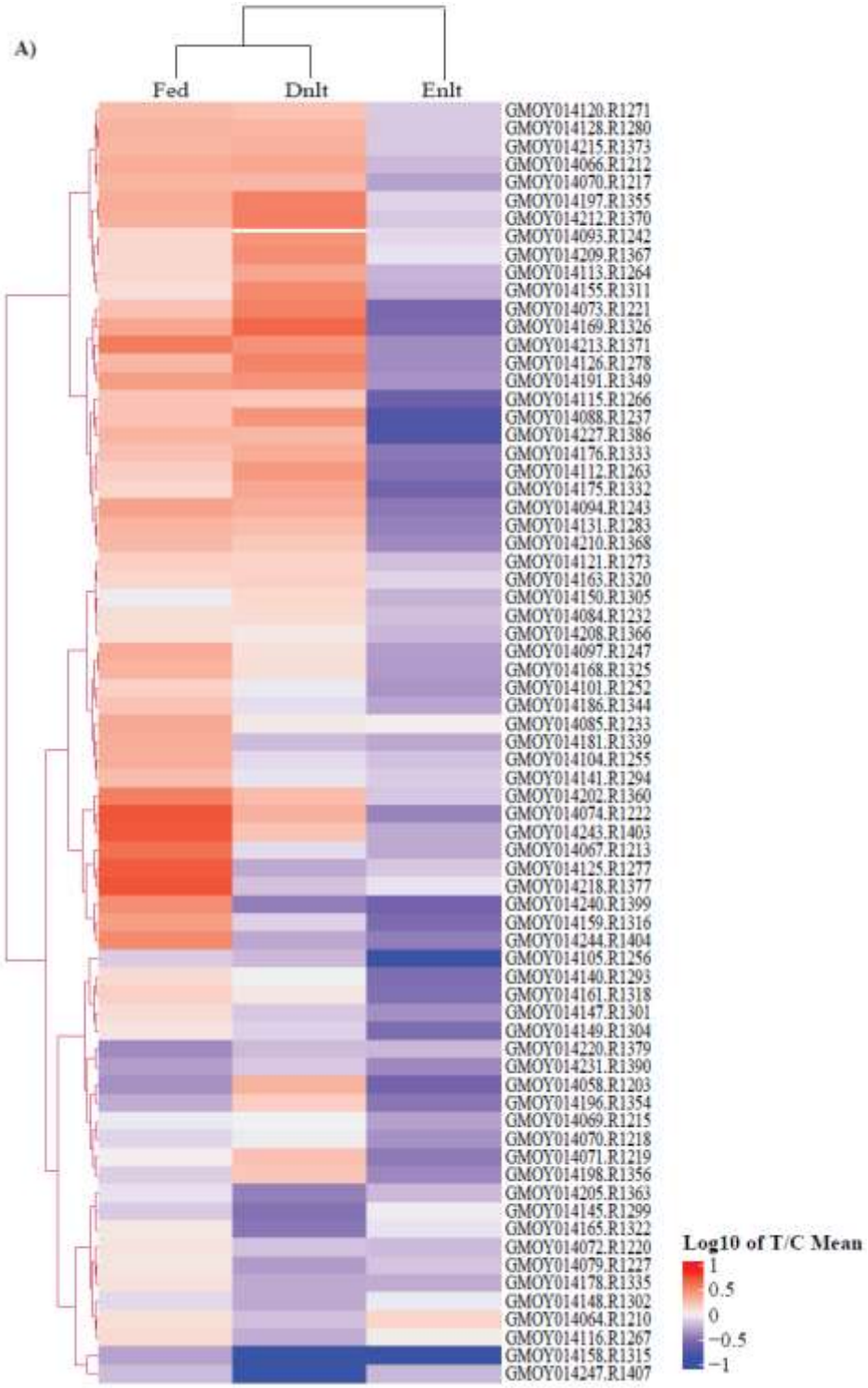

Figure 4.6 A: Heat map for 70 of 202 annotated novel genes responses to attractant (ε-nonalactone, Enlt), repellent (δ-nonalactone, Dnlt) exposure or feeding (Fed) herein referred to as treatments (T) relative to no-odor paraffin control (C). Changes were observed in the expression of the novel genes in response to the treatments. Color intensity indicate either up-regulation or suppression of the respective gene.

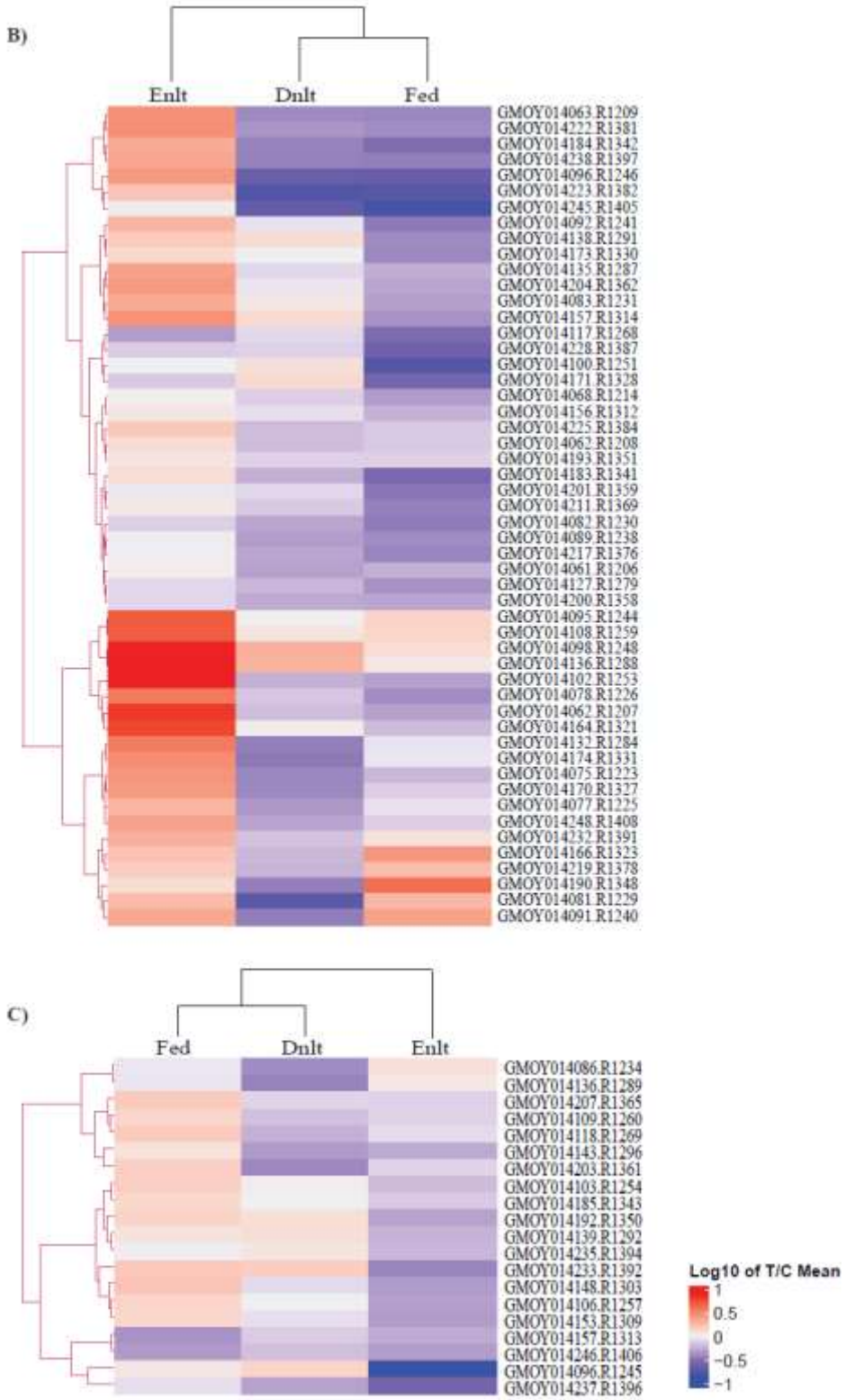

Figure 4.6 B-C: Heat map for 72 of 202 annotated novel genes responses to attractant (ε-nonalactone, Enlt), repellent (δ-nonalactone, Dnlt) exposure or feeding (Fed) herein referred to as treatments (T) relative to no-odor paraffin control (C). Color intensity indicate either up-regulation or suppression of the respective gene.

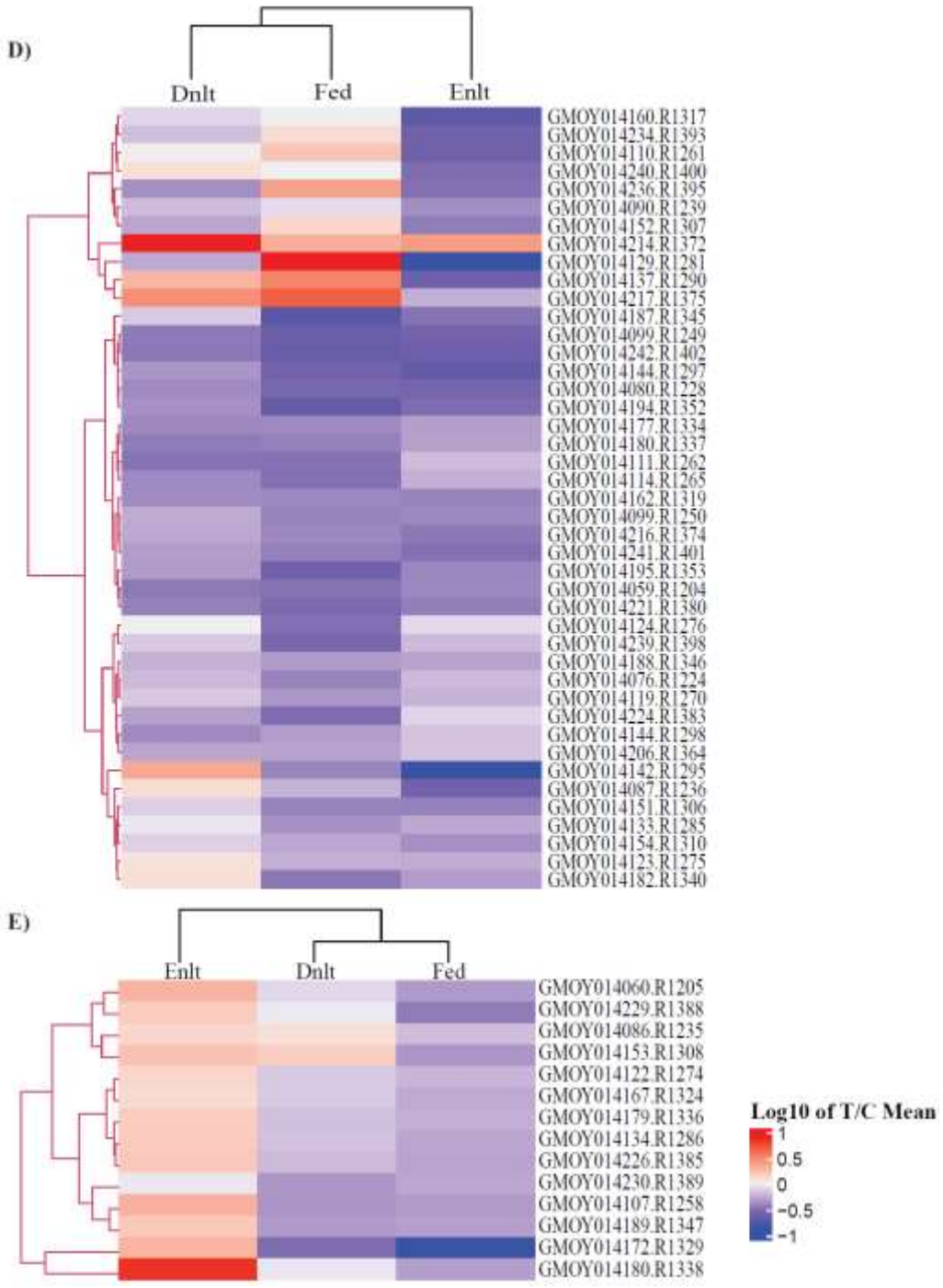

Figure 4.6 D-E: Heat map for 57 of 202 annotated novel genes responses to attractant (ε-nonalactone, Enlt), repellent (δ-nonalactone, Dnlt) exposure or feeding (Fed) herein referred to as treatments (T) relative to no-odor paraffin control (C). Color intensity indicate either up-regulation or suppression of the respective gene.

4.6 Differential expression of the antennal transcripts validated through quantitative PCR

Expression quantification of the global differentially expressed gene transcripts (Table 4.4) using quantitative real-time PCR calibrated to the *G. m. morsitans* housekeeping gene *beta-tubulin*, unveiled three transcripts; coat protein epsilon, cyclin-dependent kinase and odorant receptor 45 up-regulated in the attractant relative to the no-odor control. This outcome was consistent with the DE analysis using Deseq2 but not to expression calibrations made with the housekeeping gene GAPDH.

On the other hand, down-regulation of two (hemolectin and two scavenger receptor class A, member 5) transcripts by the attractant or up-regulation of two (homogentisate 1,2-dioxygenase and a hypothetical protein) transcripts by the repellent based on Deseq2 contrasted with their expression level calibrations made with housekeeping gene *beta-tubulin* or GAPDH where the transcripts were up-regulated or down-regulated respectively. Similarly, the down-regulation of one transcript (amalgam) by the attractant concurred with the expression levels calibrations made with GAPDH but not *beta-tubulin* (Figure 4.7).

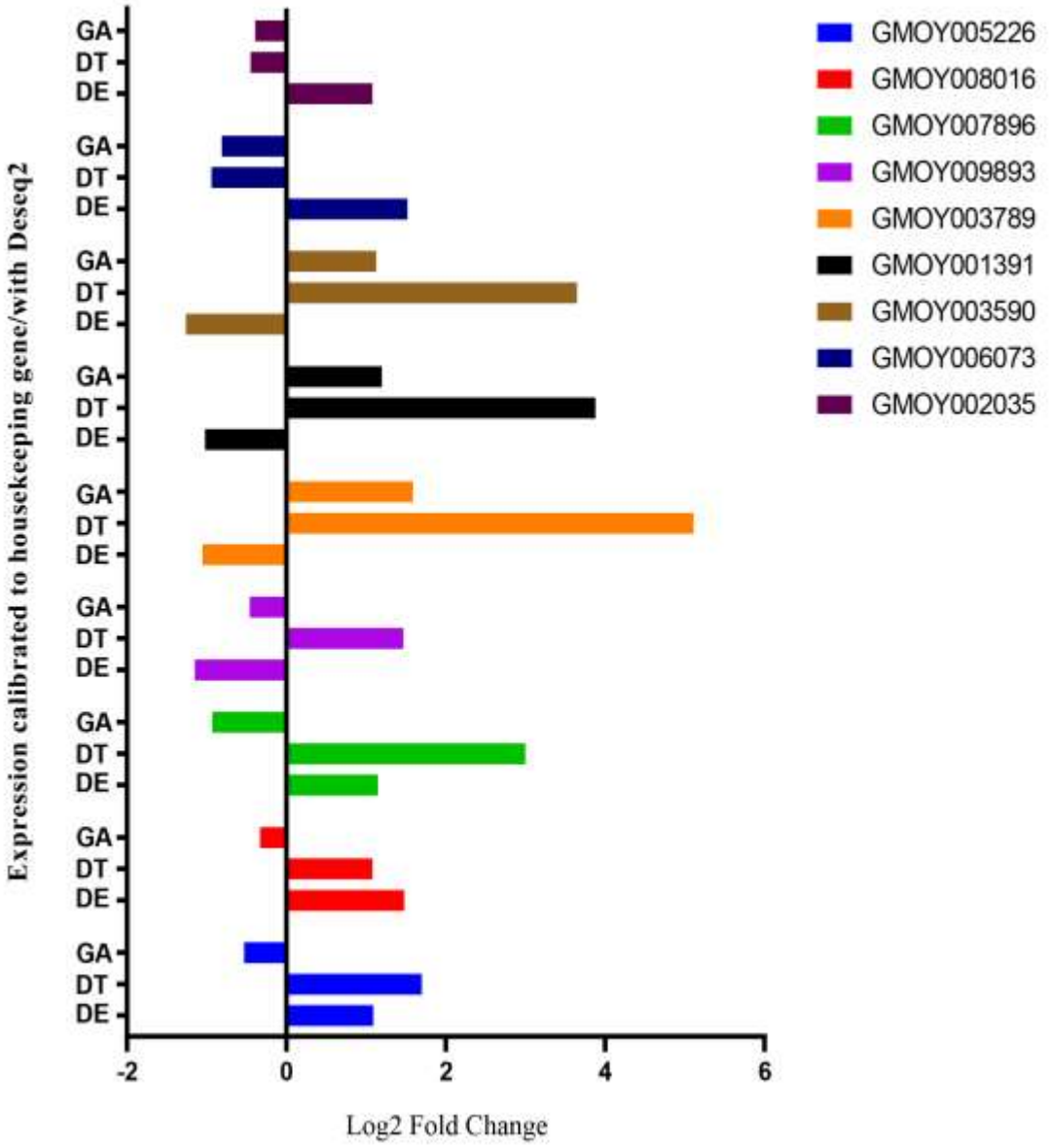

Figure 4.7. Expression quantification summary of globally differentially expressed transcripts through quantitative PCR. The fold-changes were obtained from comparing transcript expression due to odor exposures relative to similar expressions in their non-exposed controls. Expression quantifications were presented from computational analysis with Deseq2 (DE), molecular calibration with *G. m. morsitans* housekeeping genes *beta-tubulin* (DT) and GAPDH (GA). The negative and positive fold changes connote down-regulation or up-regulation of individual genes respectively by the attractant (ε-nonalactone) or repellent (δ-nonalactone) relative to non-exposed paraffin controls.

CHAPTER FIVE

DISCUSSION, CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

5.1 Discussion

Tsetse flies use olfactory cues to distinguish a suitable host from a non-suitable host before obtaining a blood meal. These decisions are mediated by genes in the tsetse fly antennae. This study sought to identify, characterize and establish putative functional roles of novel chemosensory genes in the antennae of male *G. m. morsitans*. The composition and concentration of the odor cues emanating from mammalian hosts in nature drive tsetse fly preferences. From choice experiments, it was established that ε-nonalactone attracted *G. m. morsitans* tsetse flies as previously reported by Wachira *et al*. (2016). However, there was no association between the odor concentrations and the responses observed. This could be attributed to the ability of the flies to respond to minute as well as high concentrations of the host odors (Hargrove and Vale, 1978). Studies in Drosophila also show that flies respond to minute concentrations in choice experiments (Giang *et al*., 2017). The effort of estimating the exact concentrations of these odorants that induce responses in these tsetse flies experimentally is a challenge since their distribution in nature is influenced by many biotic factors such as temperature and relative humidity (Pannunzi and Nowotny, 2019). Future field experiments in the tsetse fly natural environment with advanced tools and technologies might be key in determining this aspect of host odor.

More than 587 million reads were obtained as a result of sequencing all the eleven RNA libraries from the antennae of adult male *G. m. morsitans.* About 550 million good quality clean paired transcriptome reads were pooled from the different adult male tsetse fly treatments and replicates. More than 96% of the new transcriptome reads mapped onto the *G. m. morsitans* genome but not to the annotated gene set sequences available at the VectorBase (see Figure 3.4). This suggested an existing gap in the annotations of active genes in the genome because a majority of the reads mapped onto genomic regions that were hitherto uncharacterized. Importantly, the mappings also indicate that the reads were of good quality and are not artifacts. In conclusion, from these results, the characterization of the *G. m. morsitans* genome is incomplete and this can be remedied by using targeted tissue-specific transcriptome data to recover the missing gene coding regions.

Six genomes of tsetse fly species have been sequenced and annotated by the VectorBase community (Giraldo-Calderón *et al*., 2015). However, these initial annotations of the draft genomes relied on non-specific transcriptome existent at the time leading to underrepresentation of especially the lowly and rarely tissue-specific expressed genes. The choice of generating the antennae-specific data for this study was informed by the fact that the antennae is the principal chemoreception organ in tsetse fly, and it was hypothesized that most of the expressed genes are likely to be involved in chemosensation. Therefore, antennae-specific transcriptome from the male *G. m. morsitans* was generated and used to interrogate the completeness of the community-based *G. m. morsitans* genome annotations at the VectorBase.

Studies show that 13,018 genes have been computationally and manually annotated in the *G. m. morsitans* genome (Giraldo-Calderón *et al*., 2015). Of these, 12,308 were identified in the first ever published draft genome (IGGI, 2014) and an additional 710 subsequently annotated by the VectorBase community (Giraldo-Calderón *et al*., 2015). This study has added 202 novel genes to the gene repertoire from 592 gene models whose structures were annotated as complete with sufficient evidence for validation and adoption as per the criterion set by the VectorBase (Giraldo-Calderón *et al.*, 2015). The remaining 390 genes were already previously annotated but their structures seemed incomplete; evidences supporting improvements of their structural annotations was availed by recovering additional exons or introns and modified some the exon/intron boundaries to generate bonafide gene structures. The low number of novel genes annotated in this study might be a result of fewer genes in eukaryotic genomes and low transcriptional expression of some, depicting lack of association with any function (Pertea *et al*., 2018).

Notably, about 30% of the transcriptome were not utilized by MAKER prediction pipeline. This potentially indicates that more crucial genes expressed in the tsetse fly antennae probably remain unidentified in the genome. Hypothetically, coupling the gene finding pipeline with a more inclusive reference evidence data like the metazoan-wide non-redundant (nr) protein database as was previously used by IGGI (2014), would recover more genes. Alternatively, the unutilized transcripts probably constitute products of non-coding RNA (ncRNA) genes in the genome, and employment of The Encyclopedia of DNA Elements (ENCODE) Program (Hüttenhofer *et al.*, 2005) would offer an improved method to finding these elusive genes.

Nonetheless, this study improved the annotations of the *G. m. morsitans* genome by 4.55% and revealed that elusive gene regions can be uncovered by utilizing the otherwise divergent reads from the tissue-specific transcriptome like the ones from the antennae of the adult male *G. m. morsitans* in this study. Previous works indicate that, so far, the total number of coding genes annotated in the *G. m. morsitans* genome is considerably smaller by more than 50% relative to the dipteran *D. melanogaster* genome which is extensively annotated (IGGI, 2014). Similar observations in the number of coding genes were made in the avian and tetrapod genomes (Lovell *et al.,* 2014; Hughes and Friedman, 2008; Zhang *et al.,* 2014). In both studies on avian and tetrapod genomes, the disparities were attributed to the genes that remained unidentified and unannotated in the genomes.

Recently, divergent reads from ovarian-specific RNA-seq libraries were used to recover and annotate many novel genes in the genome of *Rhodnius prolixus* (Coelho *et al*., 2021). Similarly, this study utilized the otherwise divergent reads from antennae-specific expressed genes to characterize the gene-gap regions in the *G. m. morsitans* draft genome. The genes recovered in these genomic regions lend additional insight into critical molecular processes that underpin the physiological responses and related phenotype of the fly. Consequently, continued efforts to search and characterize the orphan regions of draft genomes like those of the *G. m. morsitans* remain critical.

The dimorphic behavior of tsetse fly in their ecological niches is partly explained by the diverse genes expressed in the antennae as the principal olfactory organ (Hallberg and Hansson, 1999; Shields, 2010). The canonical chemosensory active genes that are mostly expressed in the antennae are of odorant binding proteins (OBPs), chemosensory proteins (CSPs), gustatory receptors (GRs), odorant receptors (ORs), sensory neuron membrane proteins (SNMPs), ionotropic receptors (IRs) and odorant degrading enzymes gene families (Obiero *et al.,* 2014; Liu *et al.,* 2010; Liu *et al.,* 2012; Macharia *et al*., 2016; Kabaka *et al.,* 2020). In this study, no member of the canonical chemosensory active genes was recovered (Appendix 3).

The absence of the chemosensory genes among the novel genes could be due to previous efforts that specifically focused on annotation of chemosensory active genes, and biased towards supportive transcript evidences. This finding suggests a likelihood of successful annotation of almost all canonical antennae-associated chemosensory active genes. However, we caution that this study's data were generated from laboratory reared colony of adult male flies that were

also presented with specific treatment odors and are not active in search of a host for a blood meal.

None of the orthologs and homologs of the novel genes in selected genomes (*M. domestica, D. melanogaster* or *An. gambiae*) and in Uniprot database genes were putative canonical chemosensory genes. At functional level, the novel genes included those putatively associated with critical but general regulatory roles in the expression of these antennae-specific genes in the male *G. m. morsitans*. The credence of this argument is the fact that the homologs and orthologs to the novel genes (Appendix 3 and 4) were not directly associated with chemosensory activity but were linked to metabolic process, intracellular responses to extracellular signals, stress, regulation of cell cycle/growth, water homeostasis and diuresis point to yet undescribed complex molecular processes in the antennae of male *G. m. morsitans* in response to transient chemical and physical environment. However, how these processes specifically modulate the behavior of the tsetse fly in response to different odors remains unclear.

Two major protein domain; ribosomal protein and zinc finger protein were observed in the protein domain analysis for the novel gene residues. Besides, some of the genes showed no association to any conserved protein domains considered as tsetse fly specific, a phenomenon common to tsetse fly genomes (Attardo *et al*., 2019). The ribosomal protein domain shows greater variations in diverse organisms, with its proteins functionally indicted in ribosomal biogenesis and proper ribosomal protein folding (Melnikov *et al*., 2018).

Proteins under the zinc finger domain mediate transcriptional regulation, protein degradation and signal transduction among other functions (Lennarz and Lane 2013). Since proteins in these conserved domains cannot be directly associated with any olfactory functions up to current studies, hypothetically the expression of the parent novel genes might be. Hence further investigation on their olfactory or other roles in fulfilling the functionality of the antennae will be beneficial to understanding tsetse fly olfaction and communication biology.

Differential expression assessment of transcripts in the already existing *G. m. morsitans* gene-set revealed significant induction or suppression of different transcripts in response to the odor. All transcripts significantly modulated by exposure to either odor were not typically associated with canonical chemosensory roles except Or45 (Table 4.4). Functional role(s) of Or45 in *G. m. morsitans* have not been elucidated. However, its Or67d ortholog is involved in detection

and modulation of sensitivity to 11-cis vaccenyl acetate (cVA) pheromone (usually in concert with Snmp and lush) to prevent male-male mating and promote courtship in *D. melanogaster* (Wang & Anderson, 2010). However, concurrent inductions of the associated transcripts (Snmp and lush) were not observed in any library, which suggests that the *G. m. morsitans* Or45, potentially respond to this attractant through different mechanisms and may have different roles in different species.

The attractant also enhanced expressions of coat protein epsilon and cyclin-dependent kinase putatively associated with inducing biosynthetic protein transport within the Endoplasmic Reticulum (ER) and cell proliferation (Malumbres, 2014; Watson *et al*., 2004). On the other hand, the attractant putatively suppresses local immunity by down-regulating expressions immunoglobulin superfamily protein amalgam (Zeev-Ben-Mordehai *et al*., 2009) and hemolymph coagulation factor hemolectin (Lesch *et al*., 2007). The attractant suppressed expression of regulatory particle triple-A ATPase 4-related (Sauer and Baker, 2011), scavenger receptor class A, member 5 (Li *et al*., 2009) and vesicular monoamine transporter (Simon *et al*., 2009) transcripts that deregulate degradation of unneeded protein, iron metabolism and transport of monoamine neurotransmitters respectively. The repellent enhanced L-phenylalanine amino acid degradation by significant induction of Homogentisate 1,2-dioxygenase (Amaya *et al*., 2004).

The attractant and repellent seemed to modulate five novel and uncharacterized transcripts whose further functional characterization might provide insight and understanding of the male *G. m. morsitans* specific and distinct responses to these diverse odor cues. Associations between these transcriptional changes in non-canonical chemosensory genes and the behavioral responses due to the attractant or repellent cue are not obvious but could be due to their potential intermediary roles in chemosensory responses in the fly.

The mining of chemosensory transcripts with at least two-fold changes in response to odors established induction of Or67d, Clumsy, Ir60a, Gr2a, Gr28b, Obp83c-d and Obp19b) or suppression of Or83a, Or45b, Ir84a and Obp8a by the attractant. The repellent also specifically induced Or7a, Obp19b, Obp19d and Phk-3 or suppressed Or33a-c, Or83a, Ir84a, Clumsy, Gr66a and Obp83g) *D. melanogaster* chemosensory transcript orthologs in the *G. m. morsitans* (Table 4.5). While DREAM was more applicable to Ors, this approach nevertheless provided insight on relative performance of other chemosensory gene families. Odor stimuli decreased transcription levels of most of the Ors, consistent DREAM principle (Von Der Weid *et al*.,

2015). However, induction of *G. m. morsitans* Or8 by the repellent as observed is probably due to odor-receptor specific responses.

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) validation of these responses partially enhanced the robustness of the observations. Molecular quantification using RT-qPCR of three transcripts; Or45, coat protein epsilon and a hypothetical protein induced by the attractant was concordant with the computational analysis but not of other transcripts. Replication of expression quantities of some transcripts through RT- qPCR as identified by profiling technologies is challenged by low sensitivity despite the stringency of the technologies hence fewer genes validated as demonstrated in this and other studies (Rajkumar *et al*., 2015; Karthikeyan *et al*., 2021). Additionally, validation of non-canonical chemosensory genes using established RNAi technologies will be most appropriate given the discrepancy revealed by comparative assessments of DREAM (Von Der Weid *et al*., 2015) and other competing technologies such as phosphorylated ribosome immunoprecipitation of mRNA approach (Jiang *et al*., 2015).

Upon priming the fly to respond to either the attractant or repellent relative to odor control, there was an apparent induction of novel gene transcripts GMOY014237.R1396, GMOY014112.R1263 and GMOY014071.R1219 against the attractant and GMOY014158.R1315 (only one with significant induction) against the repellent. Characterizing such non-canonical genes will provide information of other antennal genes that will facilitate further understanding of other important roles of the antennae (De Bruyne and Baker, 2008; Boxshall and Jaume 2013) and other regulatory processes that co-ordinate responses of tsetse fly to external stimuli, including odor. With the analysis capturing only these four genes, the inductions fell short of statistical significance to enable experimental validation. Technically, validation of a gene expression independently using RT-qPCR requires at least four reference transcripts for reliable normalization of the analysis (Udvardi *et al.*, 2008) and ten randomly selected differentially expressed transcript in the RNA-seq library for appropriate assessment of correlations between RT-qPCR and RNA-Seq data (Bateta *et al.*, 2017).

5.2 Conclusions

1. A total of 202 novel genes in male *G. m. morsitans* genome were annotated from divergent reads that would otherwise be discarded in the canonical annotation pipelines currently used by the VectorBase community.
2. Putative functions of the 202 novel genes were also identified with potential involvement in general regulatory roles on expression in the antennae. Hence, these annotations provide new insights into hitherto unknown novel genes that potentially mediate molecular functions in the antennae of male *G. m. morsitans*.
3. Novel gene GMOY014237.R1396 was significantly differentially expressed in response to the attractant while other novel genes were only slightly expressed. This result confirms that the used ε-nonalactone attractant and δ-nonalactone repellant are able to elicit gene responses in the male *G. m. morsitans* tsetse fly antennae.

5.3 Limitations

1. While six tsetse fly genomes have been sequenced and made available in VectorBase, this study utilized the *G. m. morsitans* genome as the genome of interest.
2. While potential functions of the novel genes were predicted from homology analysis with the curated proteins from the UniProt database that provides information from diverse organism database including humans, this study made comparisons with dipteran relatives of tsetse fly, namely *Musca domestica*, *Anopheles gambiae* and *D. melanogaster* to elucidate the putative functions.
3. This study used Maker pipeline with limited UniProt protein database as a source of reference evidence for protein coding gene prediction.

5.4 Recommendations

1. Since not all the divergent reads were utilized in the annotation indicating a possibility of missing to identify all the antennae-specific genes, development and employment of better algorithms that might capture the unincorporated reads is needed.
2. Further functional annotation through gene knockout/silencing of the novel genes to determine their definitive roles in the response of the tsetse flies to host odors will be a great milestone to improving the control technologies in place.

3. Further annotations using transcriptomes from tsetse flies of different ages and physiological states to confirm our hypothesis on complete annotation of chemosensory genes in *G. m. morsitans* is required.

REFERENCES

Aksoy, S., Buscher, P., Lehane, M., Solano, P., & Abbeele, J. (2017). Human African trypanosomiasis control : Achievements and challenges. *Bulletin of the geological society of China*,1–6.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Alves-Silva, J., Ribeiro, J. C., Abbeele, J. D., Attardo, G., Hao, Z., Haines, L. R., & Lehane, M. J. (2010). An insight into the sialome of *Glossina morsitans morsitans*. *BMC genomics*, *11*(1).

Amaya, A. A., Brzezinski, K. T., Farrington, N., & Moran, G. R. (2004). Kinetic analysis of human homogentisate 1,2-dioxygenase. *Archives of biochemistry and biophysics*, *421*(1), 135–142.

Andersson, M. N., Löfstedt, C., & Newcomb, D. (2015). Insect olfaction and the evolution of receptor tuning. *Frontiers in ecology and evolution*, *3*(53), 1–14.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed February, 2020.

Attardo, G. M., Abd-Alla, A. M. M., Acosta-Serrano, A., Allen, J. E., Bateta, R., Benoit, J. B., Bourtzis, K., Caers, J., Caljon, G., Christensen, M. B., Farrow, D.W., Friedrich, M., Hua-Van, A., Jennings, E. C., Larkin, D. M., Lawson, D., Lehane, M. J., Lenis, V. P., Lowy-Gallego, E., … & Aksoy, S. (2019). Comparative genomic analysis of six Glossina genomes, vectors of African trypanosomes. *Genome Biology*, *20*(1), 1–31.

Attardo, G. M., Guz, N., Strickler-Dinglasan, P., & Aksoy, S. (2006). Molecular aspects of viviparous reproductive biology of the tsetse fly (*Glossina morsitans morsitans*): regulation of yolk and milk gland protein synthesis. *Journal of insect physiology*, *52*(11-12), 1128–1136.

Bateta, R., Wang, J., Wu, Y., Weiss, B. L., Warren, W. C., Murilla, G. A., Aksoy, S., & Mireji, P. O. (2017). Tsetse fly (Glossina pallidipes) midgut responses to Trypanosoma brucei challenge. *Parasites and Vectors*, *10*(1), 1–12.

Bogitsh, B. J., Carter, C. E., & Oeltmann, T. N. (2018). Arthropods as Vectors. *Human parasitology*, 331–360.

Bohbot, J. D., & Pitts, R. J. (2015). The narrowing olfactory landscape of insect odorant receptors. *Frontiers in ecology and evolution*, *3*, 1–39.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bourn, D., Reid, R., Rogers, D., Snow, B., & Wint, W., (2001). Environmental change and the autonomous control of tsetse and trypanosomiasis in Sub- Saharan Africa. *Environmental Research Group Oxford Limited, Oxford*, 1–15.

Boxshall, G., & Jaume, D. (2013). Antennules and antennae in the Crustacea. (3rd ed.). Functional Morphology and Diversity. Oxford University Press. England, UK. 199–236.

Büscher, P., Cecchi, G., Jamonneau, V., & Priotto, G. (2017). Human African Trypanosomiasis. *Lancet*, 1–13.

Caers, J., Boonen, K., Abbeele, J., Rompay, L., Schoofs, L., & Hiel, M. B. (2015). Peptidomics of neuropeptidergic tissues of the tsetse fly *Glossina morsitans morsitans*. *Journal of the American society for mass spectrometry*, *26*(12), 2024–2038.

Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. In *Current Protocols in Bioinformatics*, 2014(12), 1–39.

Carey, A. F., & Carlson, J. R. (2011). Insect olfaction from model systems to disease control. *Proceedings of the national academy of sciences*, *108*(32), 12987–12995.

Centers for Disease Control and Prevention. (2012). African trypanosomiasis. *Global Health, Division of parasitic diseases*, 1–8.

Chahda, J. S., Soni, N., Sun, J. S., Ebrahim, S. M., Weiss, B. L., & Carlson, J. R. (2019). The molecular and cellular basis of olfactory response to tsetse fly attractants. *PLoS genetics*, *15*(3), 1–22.

Chertemps, T., François, A., Durand, N., Rosell, G., Dekker, T., Lucas, P., & Maïbèche-Coisne, M. (2012). A carboxylesterase, Esterase-6, modulates sensory physiological and behavioral response dynamics to pheromone in *Drosophila*. *BMC biology*, *10*(1), 56.

Chertemps, T., Younus, F., Steiner, C., Durand, N., Coppin, C. W., Pandey, G., & Maïbèche, M. (2015). An antennal carboxylesterase from *Drosophila melanogaster*, esterase 6, is a candidate odorant-degrading enzyme toward food odorants. *Frontiers in physiology*, *6*, 315.

Chou, S., Jensen, B. C., Parsons, M., Alber, T., & Grundner, C. (2010). The *Trypanosoma brucei* life cycle switch TbPTP1 is structurally conserved and dephosphorylates the nucleolar protein NOPP44/46. *Journal of Biological Chemistry*, *285*(29), 22075–22081.

Clark, J. T., & Ray, A. (2016). Olfactory mechanisms for discovery of odorants to reduce insect-host contact. *Journal of chemical ecology*, *42*(9), 919–930.

Coelho, V. L., de Brito, T. F., de Abreu Brito, I. A., Cardoso, M. A., Berni, M. A., Araujo, H. M. M., Sammeth, M., & Pane, A. (2021). Analysis of ovarian transcriptomes reveals thousands of novel genes in the insect vector *Rhodnius prolixus*. *Scientific Reports*, *11*(1), 1–17.

De Bruyne, M., & Baker, T. C. (2008). Odor detection in insects: Volatile codes. *Journal of Chemical Ecology*, *34*(7), 882–897.

Diallo, S., Shahbaaz, M., Makwatta, J. O., Muema, J. M., Masiga, D., Christofells, A., & Getahun, M. N. (2021). Antennal Enriched Odorant Binding Proteins Are Required for

Odor Communication in *Glossina f. fuscipes*. *Biomolecules*, *11*(4), 541–556.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., & Jha, S. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 29: 15–21.

Duan, J., Ladd, T., Doucet, D., Cusson, M., Frankenhuyzen, K., Mittapalli, O., & Quan, G. (2015). Transcriptome analysis of the emerald ash borer (EAB), *Agrilus planipennis*: de novo assembly, functional annotation, and comparative analysis. *PLoS one*, *10*(8), 1–19.

Dunn, N., & Adigun, R. (2018). African Trypanosomiasis (Sleeping Sickness). Treasure island, UK: StartPearls,1–30.

Dunn, N., Unni, D., Diesh, C., Munoz-Torres, M., Harris, N. L., Yao, E., Rasche, H., Holmes, I. H., Elsik, C. G., & Lewis, S. E. (2019). Apollo: Democratizing genome annotation. *PLoS Computational Biology*, *15*(2), 1–14.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 1–14.

Food and Agricultural Organization. (2018). African Animal Trypanosomiasis; Nagana, tsetse disease, tsetse fly disease, African animal trypanosomiasis. *Factsheets*.

Franco, J. R., Simarro, P. P., Diarra, A., & Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clinical epidemiology*, *6*, 257–275.

Gadenne, C., Barrozo, R. B., & Anton, S. (2016). Plasticity in insect olfaction: to smell or not to smell. *Annual review of entomology*, *61*, 317–333.

Gakii, C., Bwana, B. K., Mugambi, G. G., Mukoya, E., Mireji, P. O., & Rimiru, R. (2021). In silico-driven analysis of the *Glossina morsitans morsitans* antennae transcriptome in response to repellent or attractant compounds. *PeerJ*, *9*, 1–20.

Giang, T., He, J., Belaidi, S., & Scholz, H. (2017). Key Odorants Regulate Food Attraction in *Drosophila melanogaster*. *Frontiers in behavioral neuroscience*, *11*(160), 1–13.

Gibson, W., & Bailey, M. (2003). The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid biology and disease*, *2*(1), 1–13.

Gikonyo, N. K., Hassanali, A., Njagi, P. G. N., & Saini, R. K. (2000). Behaviour of *Glossina morsitans morsitans* Westwood (Diptera: Glossinidae) on waterbuck *Kobus defassa* Ruppel and feeding membranes smeared with waterbuck sebum indicates the presence of allomones. *Acta Tropica*, *77*(3), 295–303.

Gikonyo, N. K., Hassanali, A., Njagi, P. N., Gitu, P. M., & Midiwo, J. O. (2002). Odor composition of Preferred and non preferred host of some savanna Tsetse flies. *Journal of chemical ecology*, *28*(5), 969–970.

Gikonyo, N. K., Hassanali, A., Njagi, P. G. N., & Saini, R. K. (2003). Responses of *Glossina m. moristans* to blends of EAG active compounds in the odours of its preferred (Buffalo and Ox) and Non Preferred (waterbuck) hosts. *Journal of Chemical Ecology*, *29*(10), 2331–2345.

Giraldo-Calderón, G. I., Emrich, S. J., MacCallum, R. M., Maslen, G., Emrich, S., Collins, F., Dialynas, E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F. H., Lawson, D., Kersey, P., Allen, J., Christensen, M., Hughes, D., Koscielny, G., Langridge, N., … Wieck, R. (2015). VectorBase: An updated Bioinformatics Resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research*, *43*(D1), 707–713.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.

Haas, B. (2018). De novo RNA - Seq Assembly, Annotation, and Analysis Using Trinity and Trinotate. https://github.com/trinityrnaseq/KrumlovTrinityWorkshopJan2018/wiki.

Hallberg, E., & Hansson, B. S. (1999). Arthropod sensilla: Morphology and phylogenetic considerations. *Microscopy Research and Technique*, *47*(6), 428–439.

Hargrove, J., & Vale, G. (1978). The effect of host odour concentration on catches of tsetse flies (Glossinidae) and other Diptera in the field. *Bulletin of Entomological Research*,68(4), 607-612.

Holt, H. R., Selby, R., Mumba, C., Napier, G. B., & Guitian, J. (2016). Assessment of animal African trypanosomiasis (AAT) vulnerability in cattle-owning communities of sub-Saharan Africa. *Parasites & Vectors*, 1–12.

Hu, P., Wang, J., Cui, M., Tao, J., & Luo, Y. (2016). Antennal transcriptome analysis of the Asian longhorned beetle *Anoplophora glabripenni*s. *Scientific Reports*, *6* (November 2015), 26652.

Hughes, A. L., & Friedman, R. (2008). Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Molecular Biology and Evolution*, *25*(12), 2681–2688.

Hüttenhofer, A., Schattner, P., & Polacek, N. (2005). Non-coding RNAs: Hope or hype? *Trends in Genetics*, *21*(5), 289–297.

International Glossina Genome Initiative; Attardo, G. M., Abila, P. P., Auma, J. E., Baumann, A. A., Benoit, J. B., Brelsfoard, C. L., & Zhou, J.-J. (2014). Genome sequence of the tsetse fly (*Glossina morsitans*): Vector of african trypanosomiasis. *Science*, *344*(6182), 380–386.

Jiang, Y., Gong, N.N., Hu, X.S., Ni, M.J., Pasi, R., & Matsunami H. (2015). Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nature Neuroscience*. 18, 1446–1454.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.

Kabaka, J. M., Wachira, B. M., Mang'era, C. M., Rono, M. K., Hassanali, A., Okoth, S. O., Oduol, V. O., Macharia, R. W., Murilla, G. A., & Mireji, P. O. (2020). Expansions of

chemosensory gene orthologs among selected tsetse fly species and their expressions in *Glossina morsitans morsitans* tsetse fly. *PLoS Neglected Tropical Diseases*, *14*(6), 1–23.

Karthikeyan, A., Pathak, S.K., Kumar, A., Kumar, S.B.A.A., Bashir, A., Singh, S., Sahoo, N. R., & Mishra, B. P. (2021). Selection and validation of differentially expressed metabolic and immune genes in weaned Ghurrah versus crossbred piglets. *Tropical Animal Health and Production, 53(1),* 1–9.

Kuzoe, F. A., Schofield, C. & UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. (2005). Strategic review of traps and targets for tsetse and African trypanosomiasis control. *World Health Organization*, 1–58.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359.

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V., Goodman, J. L., Sian Gramates, L., Millburn, G., Strelets, V. B., Tabone, C. J., & Thurmond, J. (2021). FlyBase: Updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Research*, *49*(D1), 899–907.

Leak, S. G. (1999). Tsetse biology and ecology: their role in the epidemiology and control of trypanosomiasis. Wallingford, UK: CAB International, ILRI, 228–237.

Lennarz, W. J., & Lane, M. D. (2013). Encyclopedia of biological chemistry: Zinc finger. (2nd ed.). San Diego, SD; Academic Press, 575–579.

Lesch, C., Goto, A., Lindgren, M., Bidla, G., Dushay, MS., & Theopold, U. (2007). A role for Hemolectin in coagulation and immunity in *Drosophila melanogaster*. *Developmental Computational Immunology*. 31: 1255–1263.

Li, J. Y., Paragas, N., Ned, R. M., Qiu, A., Viltard, M., Leete, T., Drexler, I. R., Chen, X., Sanna-Cherchi, S., Mohammed, F., Williams, D., Lin, C. S., Schmidt-Ott, K. M., Andrews, N. C., & Barasch, J. (2009). Scara5 is a ferritin receptor mediating non-transferrin iron delivery. *Developmental cell*, *16*(1), 35–46.

Li, B., & Dewey, C. N. (2014). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *Bioinformatics: The Impact of Accurate Quantification on Proteomic and Genetic Analysis and Research*, 41–74.

Liu, R., He, X., Lehane, S., Lehane, M., Berriman, M., Field, L. M., & Zhou, J. (2012). Expression of chemosensory proteins in the tsetse fly *Glossina morsitans morsitans* is related to female host-seeking behavior. *Insect molecular biology*, *21*(11), 41–48.

Liu, R., Lehane, S., He, X., Lehane, M., Hertz-Fowler, C., Berriman, M., & Zhou, J. J. (2010). Characterizations of odorant-binding proteins in the tsetse fly *Glossina morsitans morsitans*. *Cellular and molecular life sciences*, *67*(6), 919–929.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, *25*(4), 402–408.

Love, M.I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*.1–15.

Lovell, P. V., Wirthlin, M., Wilhelm, L., Minx, P., Lazar, N. H., Carbone, L., Warren, W. C., & Mello, C. V. (2014). Conserved syntenic clusters of protein-coding genes are missing in birds. *Genome Biology*, *15*(12), 565.

Macharia, R., Mireji, P., Murungi, E., Murilla, G., Christoffels, A., Aksoy, S., & Masiga, D. (2016). Genome-Wide comparative analysis of five chemosensory genes in five tsetse fly species. *PLoS neglected tropical diseases*, *10*(2), 1–30.

Malumbres, M. (2014). Cyclin-dependent kinases. *Genome Biology*. 1–15.

Masiga, D., Obiero, G., Macharia, R., Mireji, P., & Christoffels, A. (2014). Chemosensory receptors in tsetse flies provide link between chemical and behavioral ecology. *Trends in parasitology*, *30*(9), 426–428.

Matthews, K. R. (2005). The developmental cell biology of *Trypanosoma brucei*. *Journal of cell science*, *118*(2), 283-290.

Melnikov, S., Manakongtreecheep, K., & So¨l, D. (2018). Revising the Structural Diversity of Ribosomal Proteins Across the Three Domains of Life. *Molecular Biology Evolution* 35(7):1588–1598.

Menuz, K., Larter, N. K., Park, J., & Carlson, J. R. (2014). An RNA-Seq screen of the *Drosophila* antenna identifies a transporter necessary for ammonia detection. *PLoS Genetics*, *10*(11), 1–20.

Moloo, S. K. (1971). An artificial feeding technique for glossina. *Parasitology*, *63*(3), 507–512.

Moindi, A. O., Tare, C., Ochieng, P. J., Wamunyokoli, F., & Nyanjom, S. G. (2018). Expression of odorant co-receptor Orco in tissues and development stages of *Glossina morsitans morsitans*, *Glossina fuscipies fuscipies* and *Glossina pallidipies*. *Scientific African*, *1*(11), 1–12.

Montagné, N., Fouchier, A. De, Newcomb, R. D., & Jacquin-joly, E. (2015). Advances in the identification and characterization of olfactory receptors in insects. *Progress in molecular biology and translational science*, *130*, 55–80.

Mullen, G. R., & Durden, L. A. (2019). Medical and veterinary entomology. (3rd ed.). San Diego, SD: Academic press, 171–190.

National Center for Biotechnology Information (n.d). Accessed January 30, 2020. https://www.ncbi.nlm.nih.gov/sra.

Nakazato, T., Ohta, T., & Bono, H. (2013). Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive. *PLoS one*, *8*(10), 1–6.

Ngari, N. N., Gamba, O. D., Olet, A. P., Zhao, W., Paone, M., & Cecchi, G. (2020). Developing a national atlas to support the progressive control of tsetse - transmitted animal trypanosomosis Ain Kenya. *Parasites and vectors*, 286(13), 1–12.

Nyanjom, S. G., Tare, C., Wamunyokoli, F., & Obiero, G. F. (2018). Expression levels of odorant receptor genes in the savanna tsetse fly, *Glossina morsitans morsitans*. *Journal of Medical Entomology*, *55*(4), 855–861.

Obiero, G. O., Mireji, P. O., Nyanjom, S. G., Christoffels, A., Robertson, H. M., & Masiga, D. K. (2014). Odorant and gustatory receptors in the tsetse fly *Glossina morsitans morsitans*. *PLoS neglected tropical diseases*, *8*(4), 1–8.

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*(5), e435.

Ooi, C. P., & Bastin, P. (2013). More than meets the eye: understanding *Trypanosoma brucei* morphology in the tsetse. *Frontiers in cellular and infection microbiology*, *3*(11), 1–12.

Pannunzi M & Nowotny T (2019) Odor Stimuli: Not Just Chemical Identity. *Frontiers in Physiology*. 10:1428, 1–20.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*,14, 417–419.

Pelosi, P., Zhou, J.J., Ban, L.P., & Calvello, M. (2006). Soluble proteins in insect chemical communication. *Cellular and Molecular life sciences*, 63,1658–1676.

Percoma, L., Sow, A., Pagabeleguem, S., Dicko, A. H., Serdebéogo, O., Ouédraogo, M., & Sidibé, I. (2018). Impact of an integrated control campaign on tsetse populations in Burkina Faso. *Parasites and vectors*, 1–13.

Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Breitwieser, F.P., Chang, Y.C., Madugundu, A.K., Pandey, A., & Salzberg, S.L. (2018). CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biology*, 19(208), 1–8.

Pilson, R. D., & Pilson, B. M. (1967). Behaviour studies of Glossina morsitans Westw. in the field. *Bulletin of Entomological Research*, *57*(2), 227–257.

Rajkumar, A.P., Qvist, P., Lazarus, R., Lescai F., Ju, J., Nyegaard, M., Mors, O., Børglum, D. A., Li Q., & Christensen H. J. (2015). Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. *BMC Genomics* 548 (16), 1–9.

Robertson, H. M. (2018). Molecular evolution of the major arthropod chemoreceptor gene families. *Annual review of entomology*, *64*(1), 227–242.

Saini, R. K., Orindi, B. O., Mbahin, N., Andoke, J. A., Muasa, N., Mbuvi, D. M., & Borgemeister, W. (2017). Protecting cows in smallholder farms in East Africa from tsetse flies by mimicking the odor profile of a non-host bovid. *PLoS neglected tropical diseases*, 1–27.

Savage, A. F., Kolev, N. G., Franklin, J. B., Vigneron, A., Aksoy, S., & Tschudi, C. (2016). Transcriptome profiling of *Trypanosoma brucei* development in the tsetse fly vector *Glossina morsitans*. *PLoS one*, *11*(12), 1–20.

Sauer, R.T., and Baker, T.A. (2011). AAA+ Proteases: ATP-fueled machines of protein destruction. *Annual Reviews Biochemistry*. 80, 587–612.

Shields, D. C. (2010). High resolution ultrastructural investigation of insect sensory organs using field emission scanning electron microscopy. *Microscopy: Science, Technology,*

Applications and Education, 321–328.

Shaw, A.P.M., Cecchi, G., Wint, G.R.W., Mattioli, R.C., & Robinson, T.P. (2014). Mapping and economic benefits to livestock keepers from intervening against bovine trypanosomiasis in Eastern Africa. *Preventive veterinary medicine*,113, 197–210.

Simarro, P.P., Diarra, A., Ruiz, P.J.A., Franco, J.R., & Jannin, J.G. (2011). The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000–2009: the way forward. *PLoS Neglected Tropical Diseases*, 5(2), 1–15.

Simon, A.F., Daniels, R., Romero-Calderón, R., Grygoruk, A., Chang, H.Y., Najibi, R.,.. & Krantz, E.D (2009). Drosophila vesicular monoamine transporter mutants can adapt to reduced or eliminated vesicular stores of dopamine and serotonin. *Genetics*. 181, 525–541.

Steiner, C., Bozzolan, F., Montagné, N., Maïbèche, M., & Chertemps, T. (2017). Neofunctionalization of "Juvenile Hormone Esterase Duplication" in *Drosophila* as an odorant-degrading enzyme towards food odorants. *Scientific reports*, *7*(1), 18–20.

Steve, R., & Helen, J. S. (2000) Primer3 on the www for general users and for biologist programmers, bioinformatics methods and protocols. *Methods in molecular biology*. Totowa, NJ: *Humana Press*, 365–386.

Taparia, T., Ignell, R., & Hill, S. R. (2017). Blood meal-induced regulation of the chemosensory gene repertoire in the southern house mosquito. *BMC Genomics*, *18*(1), 1–9.

Terblanche, J. S., Clusella-Trullas, S., Deere, J. A., & Chown, S. L. (2008). Thermal tolerance in a south-east African population of the tsetse fly Glossina pallidipes (Diptera, Glossinidae): implications for forecasting climate change impacts. *Journal of insect physiology*, *54*(1), 114–127.

Thuita, J. K., Kagira, J. M., Mwangangi, D., Matovu, E., Turner, C. M. R., & Masiga, D. (2008). *Trypanosoma brucei rhodesiense* transmitted by a single tsetse fly bite in vervet monkeys as a model of human African trypanosomiasis. *PLoS neglected tropical diseases*, *2*(5), 1–7.

Tian, Z., Sun, L., Li, Y., Quan, L., Zhang, H., Yan, W., & Qiu, G. (2018). Antennal transcriptome analysis of the chemosensory gene families in *Carposina sasakii* (Lepidoptera: Carposinidae). *BMC genomics*, *19*(1), 1–16.

Udvardi, M. K., Czechowski, T., & Scheible, W. R. (2008). Eleven golden rules of quantitative RT-PCR. *Plant Cell*, *20*(7), 1736–1737.

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. *47*(1),506–515.

Von Der Weid, B., Rossier, D., Lindup, M., Tuberosa, J., Widmer, A., Col, J. D., Kan, C., Carleton, A., & Rodriguez, I. (2015). Large-scale transcriptional profiling of chemosensory neurons identifies receptor-ligand pairs in vivo. *Nature Neuroscience*, *18*(10), 1455–1463.

Wachira, B. M., Mireji, P. O., Okoth, S., William, J. M., Murilla, G. A., & Hassanali, A. (2016). Responses of *Glossina pallidipes* and *Glossina morsitans morsitans* tsetse flies to

analogues of δ-octalactone and selected blends. *Acta tropica*, *160*, 53–57.

Wang, L., & Anderson, D.J. (2010). Identification of an aggression-promoting pheromone and its receptor neurons in Drosophila. *Nature*,463, 227–231.

Watson, P.J., Frigerio, G., Collins, B.M., Duden, R., & Owen, D.J.(2004). γ-COP appendage domain - Structure and function. *Traffic.* 5, 79–88.

Willemse, L. (1991). A trial of odour baited targets to control the tsetse fly, Glossina morsitans centralis (Diptera: Glossinidae) in west Zambia. Bullet Entomological Research, 81, 351–357.

World Health Organization. (2022). Trypanosomiasis, Human African trypanosomiasis (sleeping sickness). *Factsheet*. Accessed on 2022. https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness).

Younus, F., Fraser, N. J., Coppin, C. W., Liu, J. W., Correy, G. J., Chertemps, T., & Oakeshott, J. G. (2017). Molecular basis for the behavioral effects of the odorant degrading enzyme Esterase 6 in *Drosophila*. *Scientific reports*, *7*, 1–12.

Yuvaraj, J. K., Andersson, M. N., Zhang, D. D., & Löfstedt, C. (2018). Antennal transcriptome analysis of the chemosensory gene families from Trichoptera and Basal Lepidoptera. *Frontiers in Physiology*, *9*, 1–16.

Zeev-Ben-Mordehai T, Paz A, Peleg Y, Toker L, Wolf SG, Rydberg EH, (2009). Amalgam, an axon guidance Drosophila adhesion protein belonging to the immunoglobulin superfamily: Over-expression, purification and biophysical characterization. *Protein Expr Purif*. 63, 147–157.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Zeng, Y., Xiong, Z., Liu, S., Zhou, L., Huang, Z., An, N., Wang, J., Zheng, Q., Xiong, Y., & Ganapathy, G. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, *346(6215)*, 1311–1321.

Zhang, S. F., Liu, H., Kong, X., Wang, H., Liu, F., & Zhang, Z. (2017). Identification and expression profiling of chemosensory genes in *Dendrolimus punctatus* walker. *Frontiers in Physiology*, *8*, 1–11.

Zwiebel, L. J., & Takken, W. (2004). Olfactory regulation of mosquito – host interactions. *Insect biochemistry and molecular biology*, *34*, 645–652.

APPENDICES

Appendix 1: Homologs of trinity transcripts in Uniprot

https://docs.google.com/spreadsheets/d/1i239JFnXI2lXoNuHdBt79m919ovX2T8q/edit?usp=sharing&ouid=118222947320670999711&rtpof=true&sd=true

Appendix 2: Homologs of trinity transcripts in *G. m. morsitans*

https://docs.google.com/spreadsheets/d/1CdLzxNRaCZJdNOjzIvzqH4jZ8aC15FFy/edit?usp=sharing&ouid=118222947320670999711&rtpof=true&sd=true

Appendix 3: Novel genes orthologs

https://docs.google.com/spreadsheets/d/1DDZWZEY4rY-mHHhGytNaBSCze6QP9bEc/edit?usp=sharing&ouid=118222947320670999711&rtpof=true&sd=t

Note: The supplementary tables can also be accessed on Bwana, B. K., Mireji, P. O., Obiero, G. F., Gakii, C., Akoth, M. O., Mugweru, J. N., Nyabuga, F. N., Wachira, B. M., Bateta, R., Ng'ang'a, M. M., & Hassanali, A. (2022). Annotations of novel antennae-expressed genes in male Glossina morsitans morsitans tsetse flies. *PloS one*, *17*(8), e0273543. https://doi.org/10.1371/journal.pone.0273543

Appendix 4: Novel genes homologs in Uniprot

G. m. morsitans ID	Uniprot ID*	Gene Identity	Species	% Identity	% Query Coverage	evalue	bitscore
GMOY014062.R1207	P12370	cAMP-dependent protein kinase catalytic subunit 1	Drosophila melanogaster	99.15	99	0	730
GMOY014062.R1208	P12370	cAMP-dependent protein kinase catalytic subunit 1	Drosophila melanogaster	99.15	99	0	730
GMOY014070.R1216	Q9I7H9	Prolyl 3-hydroxylase sudestada1	Drosophila melanogaster	62.524	88	0	644
GMOY014070.R1218	Q9I7H9	Prolyl 3-hydroxylase sudestada1	Drosophila melanogaster	63.269	93	0	659
GMOY014079.R1227	Q80XK6	Autophagy-related protein 2 homolog B	Mus musculus	29.528	87	0	624
GMOY014086.R1234	Q9VXA0	GATOR complex protein NPRL2	Drosophila melanogaster	85.676	96	0	682
GMOY014086.R1235	Q9VXA0	GATOR complex protein NPRL2	Drosophila melanogaster	86.072	88	0	653
GMOY014093.R1242	P48612	Protein pelota	Drosophila melanogaster	86.792	95	0	691
GMOY014113.R1264	Q295Y7	Mannose-1-phosphate guanyltransferase beta	Drosophila pseudoobscura pseudoobscura	88.14	99	0	684
GMOY014137.R1290	O75600	2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial	Homo sapiens	67.513	92	0	561
GMOY014148.R1302	Q8WSF3	Probable beta-hexosaminidase fdl	Drosophila melanogaster	63.429	97	0	953

GMOY014148.R1303	Q8WSF3	Probable beta-hexosaminidase fdl	Drosophila melanogaster	63.429	99	0	953
GMOY014158.R1315	A1ZA47	PDZ and LIM domain protein Zasp	Drosophila melanogaster	80.818	51	0	633
GMOY014175.R1332	Q5ZKN1	Cyclin-dependent kinase 9	Gallus gallus	74.932	90	0	565
GMOY014205.R1363	Q96BP3	Peptidylprolyl isomerase domain and WD repeat-containing protein 1	Homo sapiens	57.576	99	0	772
GMOY014213.R1371	B4M357	Ubiquitin-like modifier-activating enzyme 5	Drosophila virilis	74.378	99	0	591
GMOY014169.R1326	Q8BHS3	Pre-mRNA-splicing factor RBM22	Mus musculus	57.845	99	2.46E-180	511
GMOY014109.R1260	Q9W5U2	Probable chitinase 10	Drosophila melanogaster	35.434	88	2.34E-173	562
GMOY014088.R1237	Q5LJP9	Asparagine synthetase domain-containing protein CG17486	Drosophila melanogaster	44.732	99	3.29E-166	487
GMOY014163.R1320	Q6YP21	Kynurenine--oxoglutarate transaminase 3	Homo sapiens	51.708	92	9.35E-164	473
GMOY014149.R1304	O94268	25S rRNA (cytosine-C(5))-methyltransferase nop2	Schizosaccharomyces pombe	60.526	49	1.98E-163	489
GMOY014112.R1263	P0CG60	Polyubiquitin-B	Pongo pygmaeus	99.127	99	2.23E-162	450
GMOY014145.R1299	Q94CD5	Ubiquitin-like modifier-activating enzyme atg7	Arabidopsis thaliana	40	98	4.48E-160	481
GMOY014197.R1355	Q8BZW8	NHL repeat-containing protein 2	Mus musculus	40.762	87	1.65E-153	466

GMOY01404.R1232	Q68FU4	Succinate--hydroxymethylglutarate CoA-transferase	Rattus norvegicus	51.084	95	3.42E-151	439
GMOY01421.R1373	B4MGF8	Transmembrane emp24 domain-containing protein bai	Drosophila virilis	89.604	96	6.6E-134	377
GMOY01406.R1215	Q7SXR3	E3 ubiquitin-protein transferase MAEA	Danio rerio	48.724	99	9.58E-131	384
GMOY01417.R1313	Q16983	Diuretic hormone receptor	Acheta domesticus	49.742	77	1.56E-127	381
GMOY01408.R1228	Q8WTC1	28S ribosomal protein S15, mitochondrial	Drosophila melanogaster	65.299	96	5.42E-122	352
GMOY01417.R1314	Q16983	Diuretic hormone receptor	Acheta domesticus	50	75	1.21E-121	365
GMOY01407.R1219	O62518	Brix domain-containing protein ZK795.3	Caenorhabditis elegans	55.401	95	4.79E-119	346
GMOY01416.R1317	Q61753	D-3-phosphoglycerate dehydrogenase	Mus musculus	53.681	98	5.85E-119	356
GMOY01415.R1305	P52429	Diacylglycerol kinase epsilon	Homo sapiens	38.554	92	5.18E-116	357
GMOY01415.R1306	Q94518	Nascent polypeptide-associated complex subunit alpha	Drosophila melanogaster	83.556	99	3.16E-113	325
GMOY01415.R1315	A1ZA47	PDZ and LIM domain protein Zasp	Drosophila melanogaster	89.326	51	9.35E-110	386
GMOY01420.R1358	Q9WU56	tRNA pseudouridine synthase A	Mus musculus	44.032	84	2.71E-104	319
GMOY01441.R1401	P52013	Peptidyl-prolyl cis-trans isomerase 5	Caenorhabditis elegans	75.41	88	2.75E-98	286
GMOY01417.R1376	P41092	60S ribosomal protein L27a	Drosophila melanogaster	87.248	99	4.47E-90	261

GMOY014187.R1345	P51521	Transcriptional regulator ovo	Drosophila melanogaster	96.324	99	6.26E-89	284
GMOY014105.R1256	P37889	Fibulin-2	Mus musculus	28.402	98	2.03E-88	310
GMOY014114.R1265	Q9GQN0	Ran-binding protein 16	Drosophila melanogaster	82.895	99	1.42E-78	254
GMOY014109.R1260	Q9W5U2	Probable chitinase 10	Drosophila melanogaster	39.798	88	6.44E-78	284
GMOY014153.R1308	Q6NWZ9	Cysteine dioxygenase type 1	Danio rerio	52.941	84	2.21E-77	235
GMOY014153.R1309	Q6NWZ9	Cysteine dioxygenase type 1	Danio rerio	52.941	82	2.5E-77	235
GMOY014092.R1241	Q962Q6	40S ribosomal protein S24	Spodoptera frugiperda	82.576	99	2.63E-76	225
GMOY014194.R1352	Q8BX05	Putative glycerol kinase 5	Mus musculus	36.308	90	4.87E-74	241
GMOY014099.R1249	Q7JR71	Extracellular superoxide dismutase [Cu-Zn]	Drosophila melanogaster	61.111	95	7.07E-73	220
GMOY014183.R1341	Q9V597	60S ribosomal protein L31	Drosophila melanogaster	91.129	99	1.45E-72	215
GMOY014109.R1260	Q9W5U2	Probable chitinase 10	Drosophila melanogaster	33.469	88	4.41E-72	266
GMOY014099.R1250	Q7JR71	Extracellular superoxide dismutase [Cu-Zn]	Drosophila melanogaster	60.894	80	5.28E-72	219
GMOY014162.R1319	Q9VH95	Uncharacterized protein CG16817	Drosophila melanogaster	69.643	90	8.61E-72	218
GMOY014109.R1260	Q9W5U2	Probable chitinase 10	Drosophila melanogaster	38.421	88	9.44E-71	262

GMOY014152.R1307	Q701R2	Beta-galactoside alpha-2,6-sialyltransferase 2	Danio rerio	42.308	56	9.45E-69	230
GMOY014100.R1251	P84045	Histone H4	Tigriopus californicus	100	99	3.61E-67	199
GMOY014109.R1260	Q9W5U2	Probable chitinase 10	Drosophila melanogaster	36.364	88	1.13E-65	246
GMOY014073.R1221	Q2KID9	28S ribosomal protein S5, mitochondrial	Bos taurus	42.751	79	1.81E-65	215
GMOY014217.R1375	P41092	60S ribosomal protein L27a	Drosophila melanogaster	82.883	99	2.26E-62	189
GMOY014081.R1229	P10674	Fasciclin-1	Drosophila melanogaster	42.491	86	3.04E-61	208
GMOY014109.R1260	Q9W5U2	Probable chitinase 10	Drosophila melanogaster	35.509	88	1.78E-60	230
GMOY014067.R1213	Q9P2W9	Syntaxin-18	Homo sapiens	34.021	99	1.02E-54	186
GMOY014070.R1217	Q5XI31	GPI transamidase component PIG-S	Rattus norvegicus	30.147	92	8.78E-53	191
GMOY014066.R1212	Q9VJ33	NEDD8	Drosophila melanogaster	100	95	2.42E-49	153
GMOY014182.R1340	P53777	Muscle LIM protein 1	Drosophila melanogaster	91.304	98	6.43E-48	150
GMOY014219.R1378	B4NP05	Integrator complex subunit 3 homolog	Drosophila willistoni	62.121	71	2.06E-45	161
GMOY014155.R1311	Q9CY16	28S ribosomal protein S28, mitochondrial	Mus musculus	51.592	83	1.76E-43	145
GMOY014127.R1279	Q6XIM7	60S ribosomal protein L38	Drosophila yakuba	94.286	99	9.17E-42	133

GMOY01407 9.R1227	Q80XK6	Autophagy-related protein 2 homolog B	Mus musculus	30.199	87	2.61E-41	171
GMOY01418 4.R1342	O77027	Protein cubitus interruptus (Fragment)	Drosophila yakuba	43.986	98	2.98E-41	150
GMOY01407 2.R1220	P22978	Rhodanese domain-containing protein CG4456	Drosophila melanogaster	53.153	99	1.29E-40	133
GMOY01414 4.R1297	Q6PAT0	Probable inactive tRNA-specific adenosine deaminase-like protein 3	Mus musculus	30.312	95	1.58E-39	145
GMOY01414 4.R1298	Q6PAT0	Probable inactive tRNA-specific adenosine deaminase-like protein 3	Mus musculus	30.312	95	1.58E-39	145
GMOY01406 8.R1214	P22978	Rhodanese domain-containing protein CG4456	Drosophila melanogaster	54.054	99	1.14E-37	125
GMOY01413 3.R1285	P29859	Cytochrome c oxidase subunit 2	Drosophila bifasciata	58.879	97	2.1E-35	122
GMOY01408 2.R1230	Q9VF36	Acylphosphatase-2	Drosophila melanogaster	53.125	97	4.85E-33	112
GMOY01406 5.R1211	P10674	Fasciclin-1	Drosophila melanogaster	37.374	67	7.24E-32	127
GMOY01415 4.R1310	A4QNC6	Protein FAM136A	Xenopus tropicalis	38.849	94	9.93E-32	113
GMOY01421 2.R1370	P22978	Rhodanese domain-containing protein CG4456	Drosophila melanogaster	49.558	63	1.91E-31	112
GMOY01419 2.R1350	Q9NII1	Double-stranded RNA-specific editase Adar	Drosophila melanogaster	60	63	5.06E-30	118

GMOY014216.R1374	Q99N92	39S ribosomal protein L27, mitochondrial	Mus musculus	39.583	99	6.68E-29	106
GMOY014094.R1243	Q6TAW2	Stress-associated endoplasmic reticulum protein 2	Mus musculus	70.968	95	2.73E-27	96.3
GMOY014234.R1393	Q00871	Chymotrypsin BI	Penaeus vannamei	36.095	99	3.34E-25	100
GMOY014145.R1299	O44126	32 kDa beta-galactoside-binding lectin	Haemonchus contortus	30.627	84	5.45E-25	104
GMOY014173.R1330	P29872	Cytochrome c oxidase subunit 2	Ctenocephalides felis	85.417	96	1.14E-21	85.5
GMOY014242.R1402	Q9M2U3	Protein ALP1-like	Arabidopsis thaliana	26.174	70	2.64E-19	92.4
GMOY014158.R1315	A1ZA47	PDZ and LIM domain protein Zasp	Drosophila melanogaster	51.587	51	2.1E-17	92.4
GMOY014233.R1392	P26228	Protease inhibitor	Sarcophaga bullata	58.824	28	2.94E-17	74.7
GMOY014090.R1239	O76879	Circadian clock-controlled protein	Drosophila melanogaster	23.404	90	6.66E-16	78.2
GMOY014172.R1329	Q6WV19	Polypeptide N-acetylgalactosaminyltransferase 2	Drosophila melanogaster	68.085	96	8.66E-15	68.9
GMOY014240.R1399	Q02085	Zinc finger protein SNAI1	Mus musculus	34.677	22	5.91E-14	75.5
GMOY014123.R1275	P43248	DNA mismatch repair protein spellchecker 1	Drosophila melanogaster	53.061	91	6.12E-12	61.2
GMOY014190.R1348	Q31696	NADH-ubiquinone oxidoreductase chain 5 (Fragment)	Anopheles quadriannulatus	81.818	75	1.26E-11	59.3
GMOY014061.R1206	Q9W2H9	Protein panoramix	Drosophila melanogaster	28.261	38	1.27E-10	68.2

GMOY01424 0.R1400	Q6NS86	Zinc finger protein 366	Mus musculus	32.609	26	8.35E-10	65.1
GMOY01410 6.R1257	Q8VIG6	E3 ubiquitin-protein ligase TRAIP	Mus musculus	28.877	24	1.17E-09	64.7
GMOY01419 6.R1354	Q8CIV7	Transcription factor Ovo-like 2	Mus musculus	34.694	25	1.36E-08	58.9
GMOY01424 6.R1406	Q7JQ07	Mariner Mos1 transposase	Drosophila mauritiana	38.71	85	0.00000631	45.1
GMOY01422 7.R1386	O97177	Enhancer of split M2 protein	Drosophila melanogaster	25.767	75	0.0000229	47
GMOY01405 8.R1203	-	-	-	-	-	-	-
GMOY01405 9.R1204	-	-	-	-	-	-	-
GMOY01406 0.R1205	-	-	-	-	-	-	-
GMOY01406 3.R1209	-	-	-	-	-	-	-
GMOY01406 4.R1210	-	-	-	-	-	-	-
GMOY01407 4.R1222	-	-	-	-	-	-	-
GMOY01407 5.R1223	-	-	-	-	-	-	-
GMOY01407 6.R1224	-	-	-	-	-	-	-
GMOY01407 7.R1225	-	-	-	-	-	-	-
GMOY01407 8.R1226	-	-	-	-	-	-	-
GMOY01408 3.R1231	-	-	-	-	-	-	-
GMOY01408 5.R1233	-	-	-	-	-	-	-
GMOY01408 7.R1236	-	-	-	-	-	-	-
GMOY01408 9.R1238	-	-	-	-	-	-	-
GMOY01409 1.R1240	-	-	-	-	-	-	-
GMOY01409 5.R1244	-	-	-	-	-	-	-
GMOY01409 6.R1245	-	-	-	-	-	-	-
GMOY01409 6.R1246	-	-	-	-	-	-	-

GMOY0140 97.R1247	-	-	-	-	-	-	-
GMOY0140 98.R1248	-	-	-	-	-	-	-
GMOY0141 01.R1252	-	-	-	-	-	-	-
GMOY0141 02.R1253	-	-	-	-	-	-	-
GMOY0141 03.R1254	-	-	-	-	-	-	-
GMOY0141 04.R1255	-	-	-	-	-	-	-
GMOY0141 07.R1258	-	-	-	-	-	-	-
GMOY0141 08.R1259	-	-	-	-	-	-	-
GMOY0141 10.R1261	-	-	-	-	-	-	-
GMOY0141 11.R1262	-	-	-	-	-	-	-
GMOY0141 15.R1266	-	-	-	-	-	-	-
GMOY0141 16.R1267	-	-	-	-	-	-	-
GMOY0141 17.R1268	-	-	-	-	-	-	-
GMOY0141 18.R1269	-	-	-	-	-	-	-
GMOY0141 19.R1270	-	-	-	-	-	-	-
GMOY0141 20.R1271	-	-	-	-	-	-	-
GMOY0141 20.R1272	-	-	-	-	-	-	-
GMOY0141 21.R1273	-	-	-	-	-	-	-
GMOY0141 22.R1274	-	-	-	-	-	-	-
GMOY0141 24.R1276	-	-	-	-	-	-	-
GMOY0141 25.R1277	-	-	-	-	-	-	-
GMOY0141 26.R1278	-	-	-	-	-	-	-
GMOY0141 28.R1280	-	-	-	-	-	-	-
GMOY0141 30.R1282	-	-	-	-	-	-	-
GMOY0141 31.R1283	-	-	-	-	-	-	-
GMOY0141 32.R1284	-	-	-	-	-	-	-

GMOY0141 34.R1286	-	-	-	-	-	-	-
GMOY0141 35.R1287	-	-	-	-	-	-	-
GMOY0141 36.R1288	-	-	-	-	-	-	-
GMOY0141 36.R1289	-	-	-	-	-	-	-
GMOY0141 38.R1291	-	-	-	-	-	-	-
GMOY0141 39.R1292	-	-	-	-	-	-	-
GMOY0141 40.R1293	-	-	-	-	-	-	-
GMOY0141 41.R1294	-	-	-	-	-	-	-
GMOY0141 42.R1295	-	-	-	-	-	-	-
GMOY0141 43.R1296	-	-	-	-	-	-	-
GMOY0141 47.R1301	-	-	-	-	-	-	-
GMOY0141 56.R1312	-	-	-	-	-	-	-
GMOY0141 59.R1316	-	-	-	-	-	-	-
GMOY0141 61.R1318	-	-	-	-	-	-	-
GMOY0141 64.R1321	-	-	-	-	-	-	-
GMOY0141 65.R1322	-	-	-	-	-	-	-
GMOY0141 66.R1323	-	-	-	-	-	-	-
GMOY0141 67.R1324	-	-	-	-	-	-	-
GMOY0141 68.R1325	-	-	-	-	-	-	-
GMOY0141 70.R1327	-	-	-	-	-	-	-
GMOY0141 71.R1328	-	-	-	-	-	-	-
GMOY0141 74.R1331	-	-	-	-	-	-	-
GMOY0141 76.R1333	-	-	-	-	-	-	-
GMOY0141 77.R1334	-	-	-	-	-	-	-
GMOY0141 78.R1335	-	-	-	-	-	-	-
GMOY0141 79.R1336	-	-	-	-	-	-	-

GMOY0141 80.R1337	-	-	-	-	-	-	-
GMOY0141 80.R1338	-	-	-	-	-	-	-
GMOY0141 81.R1339	-	-	-	-	-	-	-
GMOY0141 85.R1343	-	-	-	-	-	-	-
GMOY0141 86.R1344	-	-	-	-	-	-	-
GMOY0141 88.R1346	-	-	-	-	-	-	-
GMOY0141 89.R1347	-	-	-	-	-	-	-
GMOY0141 91.R1349	-	-	-	-	-	-	-
GMOY0141 93.R1351	-	-	-	-	-	-	-
GMOY0141 95.R1353	-	-	-	-	-	-	-
GMOY0141 98.R1356	-	-	-	-	-	-	-
GMOY0142 01.R1359	-	-	-	-	-	-	-
GMOY0142 02.R1360	-	-	-	-	-	-	-
GMOY0142 03.R1361	-	-	-	-	-	-	-
GMOY0142 04.R1362	-	-	-	-	-	-	-
GMOY0142 06.R1364	-	-	-	-	-	-	-
GMOY0142 07.R1365	-	-	-	-	-	-	-
GMOY0142 08.R1366	-	-	-	-	-	-	-
GMOY0142 09.R1367	-	-	-	-	-	-	-
GMOY0142 10.R1368	-	-	-	-	-	-	-
GMOY0142 11.R1369	-	-	-	-	-	-	-
GMOY0142 14.R1372	-	-	-	-	-	-	-
GMOY0142 18.R1377	-	-	-	-	-	-	-
GMOY0142 20.R1379	-	-	-	-	-	-	-
GMOY0142 21.R1380	-	-	-	-	-	-	-
GMOY0142 22.R1381	-	-	-	-	-	-	-

GMOY0142 23.R1382	-	-	-	-	-	-	-
GMOY0142 24.R1383	-	-	-	-	-	-	-
GMOY0142 25.R1384	-	-	-	-	-	-	-
GMOY0142 26.R1385	-	-	-	-	-	-	-
GMOY0142 28.R1387	-	-	-	-	-	-	-
GMOY0142 29.R1388	-	-	-	-	-	-	-
GMOY0142 30.R1389	-	-	-	-	-	-	-
GMOY0142 31.R1390	-	-	-	-	-	-	-
GMOY0142 32.R1391	-	-	-	-	-	-	-
GMOY0142 35.R1394	-	-	-	-	-	-	-
GMOY0142 36.R1395	-	-	-	-	-	-	-
GMOY0142 37.R1396	-	-	-	-	-	-	-
GMOY0142 38.R1397	-	-	-	-	-	-	-
GMOY0142 39.R1398	-	-	-	-	-	-	-
GMOY0142 43.R1403	-	-	-	-	-	-	-
GMOY0142 44.R1404	-	-	-	-	-	-	-
GMOY0142 45.R1405	-	-	-	-	-	-	-
GMOY0142 47.R1407	-	-	-	-	-	-	-
GMOY0141 46.R1300	-	-	-	-	-	-	-
GMOY0141 99.R1357	-	-	-	-	-	-	-
GMOY0142 48.R1408	-	-	-	-	-	-	-

* Uniprot homolog ID as defined in Bateman, 2019.

- Homologs or domain of the transcript is not available

Appendix 5: Novel genes protein domains

G. m. morsitans ID	Domain description*	Interpro description
GMOY014070.R1216	2OG-Fe(II) oxygenase superfamily	Prolyl 3,4-dihydroxylase TPA1/OFD1, N-terminal domain
GMOY014070.R1218	2OG-Fe(II) oxygenase superfamily	Prolyl 3,4-dihydroxylase TPA1/OFD1, N-terminal domain
GMOY014147.R1301	32 KDA HEAT SHOCK PROTEIN	-
GMOY014183.R1341	60S ribosomal protein L31	Ribosomal protein L31e
GMOY014157.R1313	7 transmembrane receptor (Secretin family)	GPCR, family 2, secretin-like
GMOY014157.R1314	7 transmembrane receptor (Secretin family)	GPCR, family 2, secretin-like
GMOY014082.R1230	Acylphosphatase	Acylphosphatase
GMOY014137.R1290	Aminotransferase class I and II	Aminotransferase, class I/classII
GMOY014163.R1320	Aminotransferase class I and II	Aminotransferase, class I/classII
GMOY014088.R1237	Asparagine synthase	Asparagine synthase
GMOY014079.R1227	Autophagy-related protein C terminal domain	Autophagy-related, C-terminal
GMOY014113.R1264	Bacterial transferase hexapeptide (six repeats)	Hexapeptide repeat
GMOY014148.R1302	Beta-acetyl hexosaminidase like	Beta-hexosaminidase, eukaryotic type, N-terminal
GMOY014148.R1303	Beta-acetyl hexosaminidase like	Beta-hexosaminidase, eukaryotic type, N-terminal
GMOY014071.R1219	Brix domain	Brix domain
GMOY014105.R1256	Calcium-binding EGF domain	EGF-like calcium-binding domain
GMOY014100.R1251	Centromere kinetochore component CENP-T histone fold	CENP-T/Histone H4, histone fold
GMOY014109.R1260	Chitin binding Peritrophin-A domain	Chitin binding domain
GMOY014084.R1232	CoA-transferase family III	CoA-transferase family III
GMOY014099.R1249	Copper/zinc superoxide dismutase (SODC)	Superoxide dismutase, copper/zinc binding domain
GMOY014099.R1250	Copper/zinc superoxide dismutase (SODC)	Superoxide dismutase, copper/zinc binding domain
GMOY014162.R1319	CS domain	CS domain
GMOY014069.R1215	CTLH/CRA C-terminal to LisH motif domain	CTLH/CRA C-terminal to LisH motif domain
GMOY014205.R1363	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD	Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain
GMOY014241.R1401	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD	Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain
GMOY014153.R1308	Cysteine dioxygenase type I	Cysteine dioxygenase type I
GMOY014153.R1309	Cysteine dioxygenase type I	Cysteine dioxygenase type I
GMOY014144.R1297	Cytidine and deoxycytidylate deaminase zinc-binding region	Cytidine and deoxycytidylate deaminase domain
GMOY014144.R1298	Cytidine and deoxycytidylate deaminase zinc-binding region	Cytidine and deoxycytidylate deaminase domain

GMOY014133.R1285	Cytochrome C oxidase subunit II, periplasmic domain	Cytochrome c oxidase subunit II-like C-terminal
GMOY014173.R1330	Cytochrome C oxidase subunit II, periplasmic domain	Cytochrome c oxidase subunit II-like C-terminal
GMOY014242.R1402	DDE superfamily endonuclease	Harbinger transposase-derived nuclease domain
GMOY014195.R1353	Deltamethrin resistance	Deltamethrin resistance protein prag01
GMOY014150.R1305	Diacylglycerol kinase catalytic domain	Diacylglycerol kinase, catalytic domain
GMOY014160.R1317	D-isomer specific 2-hydroxyacid dehydrogenase, NAD binding domain	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding domain
GMOY014139.R1292	Domain of unknown function (DUF4813)	Protein of unknown function DUF4814
GMOY014192.R1350	Double-stranded RNA-specific editase 1	-
GMOY014214.R1372	Emp24/gp25L/p24 family/GOLD	GOLD domain
GMOY014215.R1373	emp24/gp25L/p24 family/GOLD	GOLD domain
GMOY014227.R1386	Enhancer of split M4 family	Enhancer of split M4 family
GMOY014093.R1242	eRF1 domain 1	eRF1 domain 1/Pelota-like
GMOY014154.R1310	Eukaryotic protein of unknown function (DUF842)	Protein of unknown function DUF842, eukaryotic
GMOY014114.R1265	Exportin-7	Exportin-7
GMOY014131.R1283	FANCI solenoid 1	FANCI solenoid 1 domain
GMOY014065.R1211	Fasciclin domain	FAS1 domain
GMOY014081.R1229	Fasciclin domain	FAS1 domain
GMOY014194.R1352	FGGY family of carbohydrate kinases, N-terminal domain	Carbohydrate kinase, FGGY, N-terminal
GMOY014206.R1364	Gag-polypeptide of LTR copia-type	-
GMOY014145.R1299	Galactoside-binding lectin	Galectin, carbohydrate recognition domain
GMOY014152.R1307	Glycosyltransferase family 29 (sialyltransferase)	Glycosyl transferase family 29
GMOY014090.R1239	Haemolymph juvenile hormone binding protein (JHBP)	Haemolymph juvenile hormone binding
GMOY014233.R1392	Kunitz/Bovine pancreatic trypsin inhibitor domain	Pancreatic trypsin inhibitor Kunitz domain
GMOY014210.R1368	LD39211P	-
GMOY014158.R1315	LIM domain	Zinc finger, LIM-type
GMOY014182.R1340	LIM domain	Zinc finger, LIM-type
GMOY014142.R1295	LITAF-like zinc ribbon domain	LPS-induced tumour necrosis factor alpha factor
GMOY014155.R1311	Mitochondrial ribosomal protein MRP-S35	Ribosomal protein S28, mitochondrial
GMOY014151.R1306	NAC domain	Nascent polypeptide-associated complex NAC domain
GMOY014197.R1355	NHL repeat	NHL repeat
GMOY014086.R1234	Nitrogen permease regulator 2	Nitrogen permease regulator 2

GMOY014086.R1235	Nitrogen permease regulator 2	Nitrogen permease regulator 2
GMOY014149.R1304	N-terminal domain of 16S rRNA methyltransferase RsmF	Ribosomal RNA small subunit methyltransferase F, N-terminal
GMOY014070.R1217	Phosphatidylinositol-glycan biosynthesis class S protein	Phosphatidylinositol-glycan biosynthesis class S protein
Gmoy014172.R1329	Polypeptide N-Acetylglucosaminyltransferase 2	-
GMOY014062.R1207	Protein kinase domain	Protein kinase domain
GMOY014062.R1208	Protein kinase domain	Protein kinase domain
GMOY014175.R1332	Protein kinase domain	Protein kinase domain
GMOY014115.R1266	Proton-conducting membrane transporter	NADH:quinone oxidoreductase/Mrp antiporter, membrane subunit
GMOY014068.R1214	Rhodanese-like domain	Rhodanese-like domain
GMOY014072.R1220	Rhodanese-like domain	Rhodanese-like domain
GMOY014212.R1370	Rhodanese-like domain	Rhodanese-like domain
GMOY014216.R1374	Ribosomal L27 protein	Ribosomal protein L27
GMOY014127.R1279	Ribosomal L38e protein family	Ribosomal protein L38e
GMOY014080.R1228	Ribosomal protein S15	Ribosomal protein S15
GMOY014092.R1241	Ribosomal protein S24e	Ribosomal protein S24e
GMOY014073.R1221	Ribosomal protein S5, N-terminal domain	Ribosomal protein S5, N-terminal
GMOY014217.R1375	Ribosomal proteins 50S-L15, 50S-L18e, 60S-L27A	Ribosomal protein L18e/L15P
GMOY014217.R1376	Ribosomal proteins 50S-L15, 50S-L18e, 60S-L27A	Ribosomal protein L18e/L15P
GMOY014094.R1243	Ribosome associated membrane protein RAMP4	Stress-associated endoplasmic reticulum protein
GMOY014106.R1257	Ring finger domain	Zinc finger, RING-type
GMOY014169.R1326	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	RNA recognition motif domain
GMOY014067.R1213	SNARE-complex protein Syntaxin-18 N-terminus	SNARE-complex protein Syntaxin-18, N-terminal
GMOY014213.R1371	ThiF family	THIF-type NAD/FAD binding fold
GMOY014184.R1342	Transcriptional activator cubitus interruptus	-
GMOY014200.R1358	tRNA pseudouridine synthase	Pseudouridine synthase I, TruA, alpha/beta domain
GMOY014248.R1408	tRNA pseudouridine synthase	Pseudouridine synthase I, TruA, alpha/beta domain
GMOY014234.R1393	Trypsin	Serine proteases, trypsin domain
GMOY014066.R1212	Ubiquitin family	Ubiquitin-like domain
GMOY014112.R1263	Ubiquitin family	Ubiquitin-like domain
GMOY014156.R1312	Ubuquitin-activating enzyme E1	-
GMOY014199.R1357	Ubiquitin-like modifier-activating enzyme ATG7 N-terminus	Ubiquitin-like modifier-activating enzyme Atg7, N-terminal
GMOY014187.R1345	Zinc finger, C2H2 type	Zinc finger C2H2-type

GMOY014196.R1354	Zinc-finger associated domain (zf-AD)	Zinc finger, AD-type
GMOY014240.R1399	Zinc-finger associated domain (zf-AD)	Zinc finger, AD-type
GMOY014240.R1400	Zinc-finger associated domain (zf-AD)	Zinc finger, AD-type
GMOY014061.R1206	-	-
GMOY014058.R1203	-	-
GMOY014059.R1204	-	-
GMOY014060.R1205	-	-
GMOY014074.R1222	-	-
GMOY014085.R1233	-	-
GMOY014087.R1236	-	-
GMOY014089.R1238	-	-
GMOY014091.R1240	-	-
GMOY014096.R1245	-	-
GMOY014096.R1246	-	-
GMOY014098.R1248	-	-
GMOY014101.R1252	-	-
GMOY014111.R1262	-	-
GMOY014120.R1271	-	-
GMOY014120.R1272	-	-
GMOY014121.R1273	-	-
GMOY014130.R1282	-	-
GMOY014135.R1287	-	-
GMOY014136.R1288	-	-
GMOY014136.R1289	-	-
GMOY014146.R1300	-	-
GMOY014161.R1318	-	-
GMOY014167.R1324	-	-
GMOY014168.R1325	-	-
GMOY014170.R1327	-	-
GMOY014174.R1331	-	-
GMOY014176.R1333	-	-
GMOY014178.R1335	-	-
GMOY014180.R1337	-	-
GMOY014180.R1338	-	-
GMOY014185.R1343	-	-
GMOY014186.R1344	-	-
GMOY014188.R1346	-	-
GMOY014191.R1349	-	-
GMOY014201.R1359	-	-
GMOY014224.R1383	-	-
GMOY014225.R1384	-	-

GMOY014228.R1387	-	-
GMOY014232.R1391	-	-
GMOY014235.R1394	-	-
GMOY014237.R1396	-	-
GMOY014244.R1404	-	-
GMOY014246.R1406	-	-
GMOY014063.R1209	-	-
GMOY014064.R1210	-	-
GMOY014075.R1223	-	-
GMOY014076.R1224	-	-
GMOY014077.R1225	-	-
GMOY014078.R1226	-	-
GMOY014083.R1231	-	-
GMOY014095.R1244	-	-
GMOY014097.R1247	-	-
GMOY014102.R1253	-	-
GMOY014103.R1254	-	-
GMOY014104.R1255	-	-
GMOY014107.R1258	-	-
GMOY014108.R1259	-	-
GMOY014110.R1261	-	-
GMOY014116.R1267	-	-
GMOY014117.R1268	-	-
GMOY014118.R1269	-	-
GMOY014119.R1270	-	-
GMOY014122.R1274	-	-
GMOY014123.R1275	-	-
GMOY014124.R1276	-	-
GMOY014125.R1277	-	-
GMOY014126.R1278	-	-
GMOY014128.R1280	-	-
GMOY014132.R1284	-	-
GMOY014134.R1286	-	-
GMOY014138.R1291	-	-
GMOY014140.R1293	-	-
GMOY014141.R1294	-	-
GMOY014143.R1296	-	-
GMOY014159.R1316	-	-
GMOY014164.R1321	-	-
GMOY014165.R1322	-	-
GMOY014166.R1323	-	-
GMOY014171.R1328	-	-
GMOY014177.R1334	-	-

GMOY014179.R1336	-	-
GMOY014181.R1339	-	-
GMOY014189.R1347	-	-
GMOY014193.R1351	-	-
GMOY014198.R1356	-	-
GMOY014202.R1360	-	-
GMOY014203.R1361	-	-
GMOY014204.R1362	-	-
GMOY014207.R1365	-	-
GMOY014208.R1366	-	-
GMOY014209.R1367	-	-
GMOY014211.R1369	-	-
GMOY014218.R1377	-	-
GMOY014219.R1378	-	-
GMOY014220.R1379	-	-
GMOY014221.R1380	-	-
GMOY014222.R1381	-	-
GMOY014223.R1382	-	-
GMOY014226.R1385	-	-
GMOY014229.R1388	-	-
GMOY014230.R1389	-	-
GMOY014231.R1390	-	-
GMOY014236.R1395	-	-
GMOY014238.R1397	-	-
GMOY014239.R1398	-	-
GMOY014243.R1403	-	-
GMOY014245.R1405	-	-
GMOY014247.R1407	-	-
GMOY014190.R1348		-

* Domain name as defined in Jones et al., 2014

- Domain of the transcript is not available

Appendix 6: List of modified genes

Vectorbase ID*	Vectorbase Gene	Vectorbase annotation#
GMOY014268.R1431	GMOY001413-RA	Acyl-CoA oxidase
GMOY012040.R1574	GMOY012040-RA	ACADS: butyryl-CoA dehydrogenase
GMOY001363.R1698	GMOY001363-RA	Activated Cdc42 kinase
GMOY014252.R1412	GMOY001505-RA	Adaptor Protein complex 2, alpha subunit
GMOY010690.R1876	GMOY010690-RA	Adaptor Protein complex 1%2C gamma subunit
GMOY010690.R1877	GMOY010690-RA	Adaptor Protein complex 1%2C gamma subunit
GMOY000853.R1749	GMOY000853-RA	Ance-5: angiotensin-converting enzyme 5
GMOY003831.R1533	GMOY003831-RA	Ankyrin repeat protein
GMOY003850.R1799	GMOY003850-RA	Arf1: ADP-ribosylation factor 1
GMOY003850.R1800	GMOY003850-RA	Arf1: ADP-ribosylation factor 1
GMOY003850.R1801	GMOY003850-RA	Arf1: ADP-ribosylation factor 1
GMOY003850.R1802	GMOY003850-RA	Arf1: ADP-ribosylation factor 1
GMOY003850.R1803	GMOY003850-RA	Arf1: ADP-ribosylation factor 1
GMOY003850.R1804	GMOY003850-RA	Arf1: ADP-ribosylation factor 1
GMOY001425.R1712	GMOY001425-RA	Aubergine
GMOY006001.R1652	GMOY006001-RA	Autophagy-related
GMOY003987.R1777	GMOY003987-RA	Axotactin
GMOY001346.R1519	GMOY001346-RA	Boudin
GMOY001349.R1894	GMOY001349-RA	Cabeza
GMOY001349.R1895	GMOY001349-RA	Cabeza
GMOY001349.R1896	GMOY001349-RA	Cabeza
GMOY004525.R1866	GMOY004525-RA	Calpain-A
GMOY004525.R1867	GMOY004525-RA	Calpain-A
GMOY003852.R1796	GMOY003852-RA	Chromator
GMOY003852.R1797	GMOY003852-RA	Chromator
GMOY003852.R1798	GMOY003852-RA	Chromator
GMOY002668.R1628	GMOY002668-RA	Circadian trip
GMOY012164.R1549	GMOY012164-RA	CSP1: Chemosensory protein 1
GMOY000155.R1644	GMOY000155-RA	Cubitus interruptus
GMOY000474.R1890	GMOY000474-RA	Cyclic-AMP response element binding protein B
GMOY000474.R1891	GMOY000474-RA	Cyclic-AMP response element binding protein B
GMOY002733.R1540	GMOY002733-RA	Daughterless
GMOY000994.R1572	GMOY000994-RA	Defective proboscis extension response
GMOY004933.R1776	GMOY004933-RA	DNA primase
GMOY001747.R1864	GMOY001747-RA	Dual-specificity tyrosine phosphorylation-regulated kinase
GMOY001747.R1865	GMOY001747-RA	Dual-specificity tyrosine phosphorylation-regulated kinase
GMOY001350.R1897	GMOY001350-RA	Easily shocked
GMOY001350.R1898	GMOY001350-RA	Easily shocked

GMOY001350.R1899	GMOY001350-RA	Easily shocked
GMOY001492.R1678	GMOY001492-RA	Echinus
GMOY002618.R1889	GMOY002618-RA	ELG3: elongase 3
GMOY000532.R1547	GMOY000532-RA	Elongator complex protein
GMOY005062.R1605	GMOY005062-RA	Enhanced level of genomic instability
GMOY005394.R1586	GMOY005394-RA	ER Membrane protein Complex
GMOY003827.R1587	GMOY003827-RA	Ester hydrolase C11orf54-like protein
GMOY006712.R1713	GMOY006712-RA	Eukaryotic translation initiation factor 2B subunit gamma
GMOY001991.R1566	GMOY001991-RA	Exo2: exocyst complex component 2
GMOY001230.R1627	GMOY001230-RA	Extra macrochaetae
GMOY000453.R1779	GMOY000453-RA	E3 ubiquitin-protein ligase
GMOY000453.R1780	GMOY000453-RA	E3 ubiquitin-protein ligase
GMOY003884.R1634	GMOY003884-RA	Fatty acyl-CoA reductase
GMOY003380.R1643	GMOY003380-RA	Female sterile (2) Ketel
GMOY012098.R1583	GMOY012098-RA	G protein-coupled receptor kinase
GMOY001205.R1561	GMOY001205-RA	Gamma-glutamylcysteine synthetase
GMOY001419.R1552	GMOY001419-RA	Gamma-tubulin complex component
GMOY006918.R1770	GMOY006918-RA	Glucose-6-phosphate 1-dehydrogenase
GMOY014349.R1514	GMOY012058-RA	Glucosamine-6-phosphate deaminase
GMOY014350.R1515	GMOY012058-RA	Glucosamine-6-phosphate deaminase
GMOY003770.R1520	GMOY003770-RA	Glutamine synthetase
GMOY004675.R1758	GMOY004675-RA	Glutathione S-transferase
GMOY001348.R1518	GMOY001348-RA	GTPase Rab21
GMOY014316.R1481	GMOY008525-RA	Glyoxylase
GMOY014317.R1482	GMOY008525-RA	Glyoxylase
GMOY003951.R1565	GMOY003951-RA	Hairy/E(spl)-related with YRPW motif
GMOY014302.R1466	GMOY004902-RA	HIB CoA deacylase
GMOY014302.R1467	GMOY004902-RA	HIB CoA deacylase
GMOY014303.R1468	GMOY004902-RA	HIB CoA deacylase
GMOY014277.R1440	GMOY002188-RA	Hig-anchoring scaffold protein
GMOY002677.R1774	GMOY002677-RA	Hormone-receptor-like in
GMOY005035.R1816	GMOY005035-RA	Increased minichromosome loss
GMOY005035.R1817	GMOY005035-RA	Increased minichromosome loss
GMOY014257.R1419	GMOY000424-RA	Imaginal discs arrested
GMOY014256.R1418	GMOY000424-RA	Imaginal discs arrested
GMOY001810.R1608	GMOY001810-RA	Ir25a: ionotropic receptor 25a
GMOY001620.R1673	GMOY001620-RA	Junctophilin
GMOY005940.R1651	GMOY005940-RA	Juvenile hormone-inducible protein
GMOY009515.R1723	GMOY009515-RA	Kallmann syndrome
GMOY001368.R1697	GMOY001368-RA	Kinesin-like protein
GMOY001887.R1716	GMOY001887-RA	Lambik
GMOY008918.R1560	GMOY008918-RA	Lanthionine synthetase C-like protein 1

GMOY005793.R1653	GMOY005793-RA	Lethal (1) G0255
GMOY010299.R1578	GMOY010299-RA	Lethal (2) k09913
GMOY006952.R1701	GMOY006952-RA	Major Facilitator Superfamily transporter
GMOY000833.R1677	GMOY000833-RA	Mediator complex subunit
GMOY000572.R1665	GMOY000572-RA	Meiotic P26
GMOY014288.R1452	GMOY002624-RA	Methionine aminopeptidase
GMOY014289.R1453	GMOY002624-RA	Methionine aminopeptidase
GMOY000033.R1828	GMOY000033-RA	Modifier of mdg4
GMOY000033.R1829	GMOY000033-RA	Modifier of mdg5
GMOY000034.R1830	GMOY000034-RA	Modifier of mdg6
GMOY000034.R1831	GMOY000034-RA	Modifier of mdg7
GMOY000034.R1832	GMOY000034-RA	Modifier of mdg8
GMOY010505.R1766	GMOY010505-RA	Mipp2: multiple inositol-polyphosphate phosphatase
GMOY001352.R1900	GMOY001352-RA	Mitochondrial F1F0-ATP synthase subunit epsilon/ATP15
GMOY001352.R1901	GMOY001352-RA	Mitochondrial F1F0-ATP synthase subunit epsilon/ATP16
GMOY009517.R1724	GMOY009517-RA	Mitogen-activated protein kinase
GMOY005627.R1750	GMOY005627-RA	Monensin sensitivity
GMOY002606.R1739	GMOY002606-RA	Myosin 81F
GMOY005034.R1603	GMOY005034-RA	N(alpha)-acetyltransferase
GMOY001029.R1638	GMOY001029-RA	NADH dehydrogenase (ubiquinone) 13 kDa B subunit
GMOY001824.R1591	GMOY001824-RA	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial
GMOY002975.R1610	GMOY002975-RA	Neither inactivation nor afterpotential B
GMOY009512.R1722	GMOY009512-RA	Niemann-Pick type C-2f
GMOY001427.R1521	GMOY001427-RA	Nucleoplasmin
GMOY001365.R1699	GMOY001365-RA	Or14: odorant receptor
GMOY012283.R1590	GMOY012283-RA	Or29: Odorant receptor
GMOY012282.R1589	GMOY012282-RA	Or30: Odorant receptor
GMOY013231.R1730	GMOY013231-RA	Or7: odorant receptor
GMOY007536.R1645	GMOY007536-RA	Orange
GMOY000882.R1569	GMOY000882-RA	Pathetic
GMOY009473.R1732	GMOY009473-RA	Phospholipase A2 group III
GMOY000544.R1921	GMOY000544-RA	Proliferation-related protein MLF
GMOY000544.R1922	GMOY000544-RA	Proliferation-related protein MLF
GMOY000544.R1923	GMOY000544-RA	Proliferation-related protein MLF
GMOY000544.R1924	GMOY000544-RA	Proliferation-related protein MLF
GMOY012181.R1658	GMOY012181-RA	Protein kinase C
GMOY001379.R1536	GMOY001379-RA	Proton-coupled amino acid transporter 1
GMOY012024.R1835	GMOY012024-RA	Pten: phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase

GMOY012024.R1836	GMOY012024-RA	Pten: phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase
GMOY001298.R1914	GMOY001298-RA	Putative D-lactate dehydrognease 2
GMOY001298.R1915	GMOY001298-RA	Putative D-lactate dehydrognease 2
GMOY000092.R1573	GMOY000092-RA	PPT: palmitoyl-protein thioesterase
GMOY009484.R1720	GMOY009484-RA	Pyrroline 5-carboyxlate reductase
GMOY006164.R1537	GMOY006164-RA	Rack1: guanine nucleotide-binding protein subunit beta-like protein
GMOY001202.R1769	GMOY001202-RA	Reaper
GMOY001201.R1534	GMOY001201-RA	RpS9: 40S ribosomal protein S9
GMOY000523.R1546	GMOY000523-RA	Rush hour
GMOY000466.R1744	GMOY000466-RA	Salivary C-type lectin
GMOY005863.R1925	GMOY005863-RA	Salivary mucin
GMOY005863.R1926	GMOY005863-RA	Salivary mucin
GMOY002054.R1738	GMOY002054-RA	SCARA5: Scavenger Receptor Class A, Member 5
GMOY005860.R1768	GMOY005860-RA	Ser/Thr-rich caspase
GMOY002729.R1856	GMOY002729-RA	Ser1: Serine protease 1
GMOY002729.R1857	GMOY002729-RA	Ser1: Serine protease 2
GMOY001316.R1862	GMOY001316-RA	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B
GMOY001316.R1863	GMOY001316-RA	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B
GMOY001030.R1639	GMOY001030-RA	Simjang
GMOY003854.R1794	GMOY003854-RA	Slif homolog 2: slimfast homolog 2
GMOY003854.R1795	GMOY003854-RA	Slif homolog 2: slimfast homolog 2
GMOY004050.R1874	GMOY004050-RA	SOD1: Superoxide Dismutase 1
GMOY004050.R1875	GMOY004050-RA	SOD1: Superoxide Dismutase 1
GMOY001351.R1558	GMOY001351-RA	SOK1 kinase
GMOY001280.R1548	GMOY001280-RA	Starvin
GMOY008180.R1694	GMOY008180-RA	Suppressor of Cytokine Signaling at 16D
GMOY003845.R1807	GMOY003845-RA	Tetraspan membrane protein in hair cell stereocilia
GMOY003845.R1808	GMOY003845-RA	Tetraspan membrane protein in hair cell stereocilia
GMOY008524.R1595	GMOY008524-RA	Tetraspanin
GMOY002896.R1622	GMOY002896-RA	TLG1: SNARE protein /syntaxin 6
GMOY002819.R1579	GMOY002819-RA	Transcription elongation factor spt6
GMOY001294.R1919	GMOY001294-RA	Translation factor GUF1 homolog, mitochondrial
GMOY001294.R1920	GMOY001294-RA	Translation factor GUF1 homolog, mitochondrial
GMOY001362.R1551	GMOY001362-RA	Transporter
GMOY001636.R1870	GMOY001636-RA	Trithorax-like
GMOY001636.R1871	GMOY001636-RA	Trithorax-like
GMOY002210.R1584	GMOY002210-RA	tRNA (guanine-N(7)-)-methyltransferase
GMOY006817.R1705	GMOY006817-RA	Troponin C-akin-1

GMOY013044.R1527	GMOY013044-RA	Tubulin-specific chaperone A
GMOY001369.R1868	GMOY001369-RA	Ubiquitin activating enzyme
GMOY001369.R1869	GMOY001369-RA	Ubiquitin activating enzyme
GMOY014275.R1438	GMOY002158-RA	Ubiquitin Ligase
GMOY014274.R1437	GMOY002158-RA	Ubiquitin Ligase
GMOY000177.R1785	GMOY000177-RA	Ubiquitinyl hydrolase 1
GMOY004044.R1872	GMOY004044-RA	WD repeat domain
GMOY004044.R1873	GMOY004044-RA	WD repeat domain
GMOY001431.R1568	GMOY001431-RA	Zeste
GMOY002523.R1707	GMOY002523-RA	4-hydroxybenzoate polyprenyltransferase, mitochondrial
GMOY001297.R1563	GMOY001297-RA	6-phosphogluconate dehydrogenase, decarboxylating
GMOY002614.R1743	GMOY002614-RA	Unspecified product
GMOY008257.R1792	GMOY008257-RA	Unspecified product
GMOY008257.R1793	GMOY008257-RA	Unspecified product
GMOY000184.R1676	GMOY000184-RA	Unspecified product
GMOY010009.R1765	GMOY010009-RA	Unspecified product
GMOY001296.R1763	GMOY001296-RA	Unspecified product
GMOY014253.R1413	GMOY001429-RA/GMOY001432-RA	Unspecified product
GMOY014254.R1414	GMOY000819-RA	Unspecified product
GMOY014254.R1415	GMOY000819-RA	Unspecified product
GMOY014254.R1416	GMOY000819-RA	Unspecified product
GMOY009853.R1708	GMOY009853-RA	Unspecified product
GMOY010473.R1596	GMOY010473-RA	Unspecified product
GMOY014352.R1517	GMOY013374-RA	Unspecified product
GMOY002804.R1745	GMOY002804-RA	Unspecified product
GMOY001242.R1532	GMOY001242-RA	Unspecified product
GMOY005352.R1719	GMOY005352-RA	Unspecified product
GMOY005659.R1756	GMOY005659-RA	Unspecified product
GMOY002962.R1577	GMOY002962-RA	Unspecified product
GMOY001320.R1751	GMOY001320-RA	Unspecified product
GMOY012158.R1664	GMOY012158-RA	Unspecified product
GMOY004993.R1704	GMOY004993-RA	Unspecified product
GMOY004641.R1656	GMOY004641-RA	Unspecified product
GMOY010898.R1660	GMOY010898-RA	Unspecified product
GMOY001175.R1904	GMOY001175-RA	Unspecified product
GMOY001175.R1905	GMOY001175-RA	Unspecified product
GMOY001175.R1906	GMOY001175-RA	Unspecified product
GMOY014261.R1423	GMOY001254-RA	Unspecified product
GMOY003858.R1753	GMOY003858-RA	Unspecified product
GMOY011816.R1575	GMOY011816-RA	Unspecified product

GMOY014273.R1436	GMOY002098-RA	Unspecified product
GMOY005569.R1630	GMOY005569-RA	Unspecified product
GMOY014310.R1475	GMOY007092-RA	Unspecified product
GMOY000804.R1654	GMOY000804-RA	Unspecified product
GMOY002331.R1601	GMOY002331-RA	Unspecified product
GMOY006848.R1754	GMOY006848-RA	Unspecified product
GMOY011690.R1787	GMOY011690-RA	Unspecified product
GMOY011690.R1788	GMOY011690-RA	Unspecified product
GMOY011690.R1789	GMOY011690-RA	Unspecified product
GMOY003846.R1809	GMOY003846-RA	Unspecified product
GMOY003846.R1810	GMOY003846-RA	Unspecified product
GMOY003846.R1811	GMOY003846-RA	Unspecified product
GMOY011987.R1609	GMOY011987-RA	Unspecified product
GMOY006607.R1625	GMOY006607-RA	Unspecified product
GMOY005931.R1539	GMOY005931-RA	Unspecified product
GMOY001358.R1747	GMOY001358-RA	Unspecified product
GMOY001729.R1818	GMOY001729-RA	Unspecified product
GMOY001729.R1819	GMOY001729-RA	Unspecified product
GMOY014335.R1500	GMOY009668-RA	Unspecified product
GMOY014294.R1458	GMOY003716-RA	Unspecified product
GMOY014314.R1479	GMOY007380-RA	Unspecified product
GMOY004423.R1767	GMOY004423-RA	Unspecified product
GMOY011981.R1629	GMOY011981-RA	Unspecified product
GMOY014262.R1424	GMOY001254-RA	Unspecified product
GMOY001493.R1564	GMOY001493-RA	Unspecified product
GMOY005954.R1604	GMOY005954-RA	Unspecified product
GMOY000904.R1592	GMOY000904-RA	Unspecified product
GMOY002402.R1582	GMOY002402-RA	Unspecified product
GMOY007190.R1593	GMOY007190-RA	Unspecified product
GMOY002120.R1781	GMOY002120-RA	Unspecified product
GMOY002120.R1782	GMOY002120-RA	Unspecified product
GMOY003974.R1580	GMOY003974-RA	Unspecified product
GMOY001897.R1600	GMOY001897-RA	Unspecified product
GMOY001899.R1599	GMOY001899-RA	Unspecified product
GMOY014292.R1456	GMOY003378-RA	Unspecified product
GMOY000525.R1927	GMOY000525-RA	Unspecified product
GMOY001469.R1655	GMOY001469-RA	Unspecified product
GMOY001299.R1912	GMOY001299-RA	Unspecified product
GMOY001299.R1913	GMOY001299-RA	Unspecified product
GMOY004974.R1650	GMOY004974-RA	Unspecified product
GMOY002396.R1675	GMOY002396-RA	Unspecified product
GMOY014278.R1441	GMOY002397-RA	Unspecified product
GMOY014278.R1442	GMOY002397-RA	Unspecified product

GMOY002392.R1674	GMOY002392-RA	Unspecified product
GMOY002187.R1538	GMOY002187-RA	Unspecified product
GMOY001882.R1717	GMOY001882-RA	Unspecified product
GMOY000420.R1755	GMOY000420-RA	Unspecified product
GMOY010219.R1607	GMOY010219-RA	Unspecified product
GMOY003913.R1567	GMOY003913-RA	Unspecified product
GMOY014271.R1434	GMOY001799-RA	Unspecified product
GMOY006266.R1633	GMOY006266-RA	Unspecified product
GMOY012130.R1679	GMOY012130-RA	Unspecified product
GMOY005259.R1709	GMOY005259-RA	Unspecified product
GMOY014315.R1480	GMOY007380-RA	Unspecified product
GMOY002408.R1571	GMOY002408-RA	Unspecified product
GMOY001360.R1559	GMOY001360-RA	Unspecified product
GMOY003772.R1773	GMOY003772-RA	Unspecified product
GMOY001309.R1762	GMOY001309-RA	Unspecified product
GMOY014266.R1428	GMOY001264-RA	Unspecified product
GMOY014267.R1429	GMOY001264-RA	Unspecified product
GMOY014267.R1430	GMOY001264-RA	Unspecified product
GMOY009989.R1714	GMOY009989-RA	Unspecified product
GMOY011166.R1621	GMOY011166-RA	Unspecified product
GMOY006525.R1812	GMOY006525-RA	Unspecified product
GMOY006525.R1813	GMOY006525-RA	Unspecified product
GMOY001073.R1602	GMOY001073-RA	Unspecified product
GMOY001891.R1715	GMOY001891-RA	Unspecified product
GMOY014312.R1477	GMOY007152-RA	Unspecified product
GMOY006778.R1783	GMOY006778-RA	Unspecified product
GMOY006778.R1784	GMOY006778-RA	Unspecified product
GMOY014311.R1476	GMOY007092-RA	Unspecified product
GMOY004191.R1700	GMOY004191-RA	Unspecified product
GMOY014265.R1427	GMOY001257-RA	Unspecified product
GMOY014306.R1471	GMOY006139-RA	Unspecified product
GMOY014301.R1465	GMOY004876-RA	Unspecified product
GMOY001522.R1837	GMOY001522-RA	Unspecified product
GMOY001522.R1838	GMOY001522-RA	Unspecified product
GMOY001522.R1839	GMOY001522-RA	Unspecified product
GMOY014296.R1460	GMOY003962-RA	Unspecified product
GMOY014297.R1461	GMOY003962-RA	Unspecified product
GMOY004761.R1778	GMOY004761-RA	Unspecified product
GMOY001076.R1529	GMOY001076-RA	Unspecified product
GMOY002573.R1681	GMOY002573-RA	Unspecified product
GMOY001018.R1680	GMOY001018-RA	Unspecified product
GMOY002577.R1598	GMOY002577-RA	Unspecified product
GMOY001321.R1752	GMOY001321-RA	Unspecified product

GMOY004176.R1833	GMOY004176-RA	Unspecified product
GMOY004176.R1834	GMOY004176-RA	Unspecified product
GMOY001311.R1689	GMOY001311-RA	Unspecified product
GMOY014309.R1474	GMOY006465-RA	Unspecified product
GMOY014264.R1426	GMOY001257-RA	Unspecified product
GMOY001004.R1659	GMOY001004-RA	Unspecified product
GMOY014300.R1464	GMOY004876-RA	Unspecified product
GMOY001416.R1711	GMOY001416-RA	Unspecified product
GMOY014293.R1457	GMOY003378-RA	Unspecified product
GMOY001197.R1666	GMOY001197-RA	Unspecified product
GMOY006556.R1648	GMOY006556-RA	Unspecified product
GMOY000611.R1667	GMOY000611-RA	Unspecified product
GMOY000332.R1649	GMOY000332-RA	Unspecified product
GMOY000104.R1597	GMOY000104-RA	Unspecified product
GMOY006910.R1641	GMOY006910-RA	Unspecified product
GMOY012062.R1718	GMOY012062-RA	Unspecified product
GMOY014307.R1472	GMOY006139-RA	Unspecified product
GMOY001490.R1581	GMOY001490-RA	Unspecified product
GMOY003972.R1902	GMOY003972-RA	Unspecified product
GMOY003972.R1903	GMOY003972-RA	Unspecified product
GMOY012077.R1635	GMOY012077-RA	Unspecified product
GMOY001619.R1849	GMOY001619-RA	Unspecified product
GMOY001619.R1850	GMOY001619-RA	Unspecified product
GMOY001619.R1851	GMOY001619-RA	Unspecified product
GMOY008575.R1688	GMOY008575-RA	Unspecified product
GMOY001390.R1623	GMOY001390-RA	Unspecified product
GMOY001528.R1663	GMOY001528-RA	Unspecified product
GMOY000512.R1695	GMOY000512-RA	Unspecified product
GMOY007638.R1759	GMOY007638-RA	Unspecified product
GMOY001300.R1916	GMOY001300-RA	Unspecified product
GMOY001300.R1917	GMOY001300-RA	Unspecified product
GMOY001300.R1918	GMOY001300-RA	Unspecified product
GMOY002727.R1702	GMOY002727-RA	Unspecified product
GMOY001374.R1657	GMOY001374-RA	Unspecified product
GMOY003848.R1805	GMOY003848-RA	Unspecified product
GMOY003848.R1806	GMOY003848-RA	Unspecified product
GMOY007535.R1535	GMOY007535-RA	Unspecified product
GMOY000402.R1693	GMOY000402-RA	Unspecified product
GMOY001551.R1562	GMOY001551-RA	Unspecified product
GMOY001295.R1764	GMOY001295-RA	Unspecified product
GMOY002603.R1884	GMOY002603-RA	Unspecified product
GMOY002603.R1885	GMOY002603-RA	Unspecified product
GMOY012036.R1824	GMOY012036-RA	Unspecified product

GMOY012036.R1825	GMOY012036-RA	Unspecified product
GMOY008842.R1550	GMOY008842-RA	Unspecified product
GMOY014318.R1483	GMOY008844-RA	Unspecified product
GMOY003844.R1524	GMOY003844-RA	Unspecified product
GMOY014299.R1463	GMOY004054-RA	Unspecified product
GMOY014298.R1462	GMOY004054-RA	Unspecified product
GMOY003291.R1840	GMOY003291-RA	Unspecified product
GMOY003291.R1842	GMOY003291-RA	Unspecified product
GMOY012114.R1706	GMOY012114-RA	Unspecified product
GMOY000570.R1909	GMOY000570-RA	Unspecified product
GMOY000570.R1910	GMOY000570-RA	Unspecified product
GMOY000570.R1911	GMOY000570-RA	Unspecified product
GMOY002072.R1761	GMOY002072-RA	Unspecified product
GMOY002602.R1691	GMOY002602-RA	Unspecified product
GMOY001393.R1624	GMOY001393-RA	Unspecified product
GMOY001354.R1892	GMOY001354-RA	Unspecified product
GMOY001354.R1893	GMOY001354-RA	Unspecified product
GMOY005674.R1710	GMOY005674-RA	Unspecified product
GMOY014250.R1410	GMOY012234-RA	Unspecified product
GMOY000108.R1814	GMOY000108-RA	Unspecified product
GMOY000108.R1815	GMOY000108-RA	Unspecified product
GMOY001355.R1746	GMOY001355-RA	Unspecified product
GMOY004435.R1606	GMOY004435-RA	Unspecified product
GMOY014295.R1459	GMOY003716-RA	Unspecified product
GMOY014263.R1425	GMOY001254-RA	Unspecified product
GMOY001119.R1757	GMOY001119-RA	Unspecified product
GMOY001216.R1854	GMOY001216-RA	Unspecified product
GMOY001216.R1855	GMOY001216-RA	Unspecified product
GMOY002722.R1703	GMOY002722-RA	Unspecified product
GMOY000078.R1642	GMOY000078-RA	Unspecified product
GMOY001218.R1852	GMOY001218-RA	Unspecified product
GMOY001218.R1853	GMOY001218-RA	Unspecified product
GMOY014270.R1433	GMOY001799-RA	Unspecified product
GMOY011964.R1790	GMOY011964-RA	Unspecified product
GMOY011964.R1791	GMOY011964-RA	Unspecified product
GMOY014272.R1435	GMOY002098-RA	Unspecified product
GMOY009500.R1721	GMOY009500-RA	Unspecified product
GMOY003291.R1841	GMOY003291-RA	Unspecified product
GMOY003826.R1523	GMOY003826-RA	Unspecified product
GMOY007381.R1760	GMOY007381-RA	Unspecified product
GMOY014319.R1484	GMOY008844-RA	Unspecified product
GMOY014308.R1473	GMOY006465-RA	Unspecified product
GMOY006920.R1771	GMOY006920-RA	Unspecified product

GMOY001494.R1772	GMOY001494-RA	Unspecified product
GMOY002754.R1626	GMOY002754-RA	Unspecified product
GMOY002641.R1646	GMOY002641-RA	Unspecified product
GMOY002412.R1690	GMOY002412-RA	Unspecified product
GMOY014279.R1443	GMOY002397-RA	Unspecified product
GMOY014280.R1444	GMOY002397-RA	Unspecified product
GMOY014313.R1478	GMOY007152-RA	Unspecified product
GMOY014351.R1516	GMOY013374-RA	Unspecified product
GMOY000323.R1528	GMOY000323-RA	Unspecified product
GMOY000571.R1907	GMOY000571-RA	Unspecified product
GMOY000571.R1908	GMOY000571-RA	Unspecified product
GMOY000884.R1594	GMOY000884-RA	Unspecified product
GMOY001184.R1826	GMOY001184-RA	Unspecified product
GMOY001184.R1827	GMOY001184-RA	Unspecified product
GMOY001307.R1696	GMOY001307-RA	Unspecified product
GMOY003736.R1522	GMOY003736-RA	Unspecified product
GMOY003898.R1847	GMOY003898-RA	Unspecified product
GMOY003898.R1848	GMOY003898-RA	Unspecified product
GMOY006482.R1640	GMOY006482-RA	Unspecified product
GMOY007206.R1737	GMOY007206-RA	Unspecified product
GMOY009525.R1725	GMOY009525-RA	Unspecified product
GMOY010375.R1748	GMOY010375-RA	Unspecified product

* Vectorbase ID as defined in (Giraldo-Calderón et al., 2015)

Vectorbase parent gene annotated function as defined in (Giraldo-Calderón et al., 2015)