GENOME SEQUENCE OF TSETSE POLYDNAVIRUS: INSIGHTS INTO SYMBIOTIC VIRUS EVOLUTION

KELVIN MUTERU KIMENYI, BSc.

I56/70112/2013

A THESIS SUBMITTED TO THE UNIVERSITY OF NAIROBI,

CENTER FOR BIOTECHNOLOGY AND BIOINFORMATICS

IN PARTIAL FULFILLMENT FOR AWARD OF THE DEGREE OF

MASTER OF SCIENCE IN BIOINFORMATICS

OCTOBER 2015

DECLARATION

I declare that this research thesis is entirely my work and has not been submitted for a degree in any other University

Kelvin Muteru Kimenyi,

I56/70112/2013

Signature: _____ Date: _____

APPROVED BY

This research thesis has been submitted with our approval as supervisors:

Dr. Benard Kulohoma,

Center for Biotechnology and Bioinformatics, University of Nairobi

Signature:  Date: 17th March 2016

Dr. Daniel Masiga,

International Center for Insect Physiology and Ecology, Nairobi, Kenya

Signature: _____ Date: _____

DEDICATION

I dedicate this thesis to my family.

ACKNOWLEDGEMENTS

First and foremost, I want to thank the Almighty God for His grace that has enabled me to successfully finish this work.

Secondly, I want to acknowledge my study supervisors who have stood by me, provided technical insights and mentored me throughout the project. Dr. Benard Kulohoma, Dr. Daniel Masiga and Prof. Enock Matovu, I am greatly indebted to you and may the good Lord bless you and give you prosperity in your future endeavors.

I also wish to acknowledge the enormous contribution of my colleagues and friends for their much needed suggestions and comments. Many thanks to Winnie Okeyo for guiding me through the laboratory work, and Muna Abry for her comments and suggestions on writing.

I take this opportunity to thank the University of Nairobi for the award of a scholarship that supported me financially. I also acknowledge the staff and students at the Center for Biotechnology and Bioinformatics (CEBIB) where I was registered as a Masters student, their support and cooperation throughout the project period was invaluable. Special thanks to CEBIB Director, Dr. George Obiero for recommending me to various scientific workshops and conferences.

Many thanks to *icipe* for providing resources for building my capacity as an upcoming scientist. In particular, I acknowledge the Molecular Biology and Bioinformatics Unit (MBBU) for hosting me. The suggestions and comments during journal club really helped to make this work a success.

Finally, I gratefully acknowledge my parents, Mr. and Mrs. Muteru and my sisters, Naomi and Esther. Despite their limited understanding of this work, their prayers, encouragement and financial support kept me going. I also acknowledge anybody not mentioned here who contributed in one-way or another to the success of this project.

ABSTRACT

Tsetse flies are the sole vectors of African trypanosomiasis in humans (sleeping sickness) and animals (nagana). The whole genome sequence of the tsetse fly *Glossina morsitans morsitans* revealed the presence of putative bracoviral sequences (n = 310) widely spread in the genome. These sequences are similar to those identified in parasitic braconid wasps. Bracoviruses (family *Polydnaviridae*) encode proteins that lower host immunity allowing for the development of parasitoid larvae in the host. The study aimed to determine the presence, prevalence and genetic diversity of Polydnaviruses (PDVs) across five tsetse fly species genomes (*Glossina austeni, Glossina brevipalpis, Glossina fuscipes fuscipes, Glossina morsitans morsitans* and *Glossina pallidipes*) found in East Africa. Four genes were unique to PDVs and served as good evolutionary models. These genes are components of mobile elements called Mavericks that are found in the PDVs genomes. They include DNA polymerase B2, Parvovirus coat protein, retroviral-like integrase and poxvirus A32 protein. The genes were observed to be present in multiple copies. *G. austeni* had the highest number (n = 18) and *G. brevipalpis* the least (n = 6). Phylogenetic reconstruction of each gene revealed two major clades which represent the two types of Mavericks. Selection pressure acting on these genes and their flanking regions was evaluated using dN/dS ratio. The study confirmed the presence of PDVs in tsetse fly genomes. The four PDV associated genes used as evolutionary modes were observed to be under varying magnitudes of purifying selection except for the poxvirus A32 which was under positive selection. These genes were showed to be inserted at conserved regions and co-evolve at similar rates with the host genomes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PDV	Polydnavirus
AAT	Animal African trypanosomiasis
HAT	Human African trypanosomiasis
SAT	Sequential Aerosol Spraying Technique
SIT	Sterile Insect Technique
Mbp	Mega base pairs
BLAST	Basic Local Alignment Tool
dsDNA	double stranded Deoxyribonucleic Acid
dNTPs	deoxy Nucleotide Triphosphate
NCBI	National Center for Biotechnology Information
UV	Ultra-Violet
DDT	Dichlorodiphenyltrichloroethane
TSP	Teratocyte Secreting Protein
TIR	Terminal Inverted Repeat

1.0 INTRODUCTION

1.1 Background information

Tsetse flies are Dipteran flies belonging to the genus, *Glossina*. They are the sole vectors of trypanosomes, which cause sleeping sickness or human African trypanosomiasis (HAT) in human and nagana or animal African trypanosomiasis (AAT) in livestock in sub-Saharan Africa (Yang et al. 2010). Both male and female adult tsetse flies feed exclusively on blood and transmit the disease during feeding. There are over 30 species and sub-species of tsetse, most of which can transmit trypanosomiasis. However, only 8-10 are of medical and agricultural importance. The most important tsetse vectors are the *Morsitans* or savannah species (*Glossina morsitans morsitans, Glossina pallidipes* and *Glossina austeni*) in Eastern and Southern Africa and the *Palpalis* or riverine species (*Glossina palpalis and Glossina fuscipes fuscipes*) in Western and Central Africa (Wellde et al. 1989). Tsetse flies are normally confined to Africa although an isolated population has been reported on the Arabian Peninsula (Elsen, Amoudi, and Leclercq 1991).

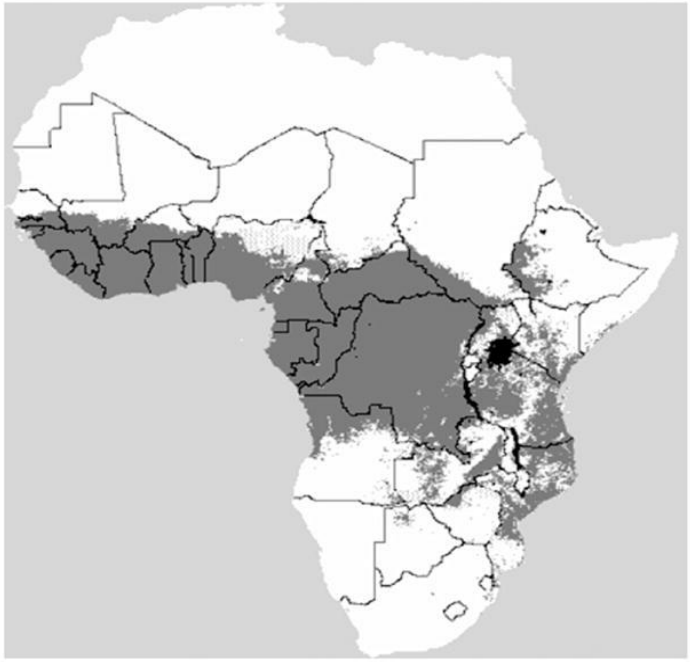

Figure 1: Distribution of tsetse flies in Africa. A map of Africa showing distribution of tsetse flies (grey) in the Sub-Saharan region. This image is adopted from (Abd-Alla et al. 2013)

HAT is a chronic, slow wasting disease that ultimately causes death if untreated. Approximately, 70 million people worldwide were at risk of infection in 1995. In 2012, 7,216 cases were reported and approximately 20,000 people across Africa were estimated to be infected (World Health Organization 2013). Around 50 million cattle in tsetse fly infested areas are at risk of AAT infection and approximately 3 million cattle deaths per year are reported (Gooding and Krafsur 2005). Farmers administer about 35 million doses of trypanocidal drugs adding to the total agriculture loss of $4.75 billion per year (Benoit et al. 2015a). In Africa, more than a third of the continent, mostly south of Sahara, is inhabited by tsetse flies exposing approximately 66 million people and 150 million cattle to the risk of contracting HAT and AAT respectively (Maudlin 2006). The people at the highest risk of tsetse fly bites, and of contracting HAT are the rural populations that primarily depend on small-scale agriculture. HAT epidemics are associated with economic decline, civil wars and population movements (Wellde et al. 1989).

1.2 Tsetse biology

Tsetse flies are members of higher *Diptera* and belong to the superfamily *Hippoboscoidea*. This superfamily contains four families namely *Glossinidae* (tsetse flies), *Hippoboscidae* (louse flies) and the *Streblidae* and *Nycteribiidae* (two families of bat flies). The family *Glossinidae* contains only one Genus called *Glossina* (Aksoy 2010). This genus has 23 species and 8 sub-species. The species are categorized into three groups based on the ecological niche they occupy namely *Fusca, Palpalis* and *Morsitans* (Benoit et al. 2015a). The *Fusca* flies mostly inhabit the rain forests of West Africa. Common members of *Fusca* include *G. fusca. fusca, G. schweitz, G. brevipalpis* and *G. longipennis*. Apart from *G. brevipalpis*, other members of this group have no medical or veterinary importance (Leak 1998). *Palpalis* group flies are riverine and include, *G. p. palpalis, G. f. fuscipes* and *G. p. gambiensis*. Members of this group are common vectors of *T. b. gambiense* and *T. b. congolense* that cause HAT. On the other hand, *Morsitans* group flies are largely found in the savanna and woodland environment. They include, *G. longipalpis, G. pallidipes, G. m. morsitans, G. swynnertoni* and *G. austeni*, and are major vectors of *Trypanosoma brucei* which causes nagana (Leak 1998).

1.3 Life cycle of tsetse fly

Tsetse flies reproduce by giving birth to live offsprings, a phenomena called adenotropic viviparity. Female tsetse live an average of ~90 days and are capable of producing an average of 8-10 off springs during the entire course of their lifetime. The females also mate only once in their lifetime storing the sperms in their spermathecal while the males can mate more than once (Yang et al. 2010).

Female flies release eggs periodically from the ovaries. Only one egg is fertilized at a time and hatches into a larva. The larva is retained *in utero*, where it is nourished with a milky secretion secreted by a modified accessory gland connected to the uterus and expands as a series of bifurcating tubules throughout the abdominal cavity. The secretion mainly contains nutrients and tsetse symbionts (Benoit et al. 2015b). After a period of development and molting, a third instar larva leaves the uterus and is deposited on the ground approximately 16 days after fertilization. The larva immediately crawls into the soil and pupates within 1-2 hours while relying on food resources derived from the mother. After 22-60 days, adult flies emerge in a temperature dependent fashion. This marks the end of the first cycle. Warmer the temperatures reduce the period it takes for the adult to emerge and vice versa (Leak 1998). The second cycle begins with oogenesis, which takes about 6-7 days to complete. This is followed by ovulation, after which a sperm from the spermathecae fertilizes an egg.

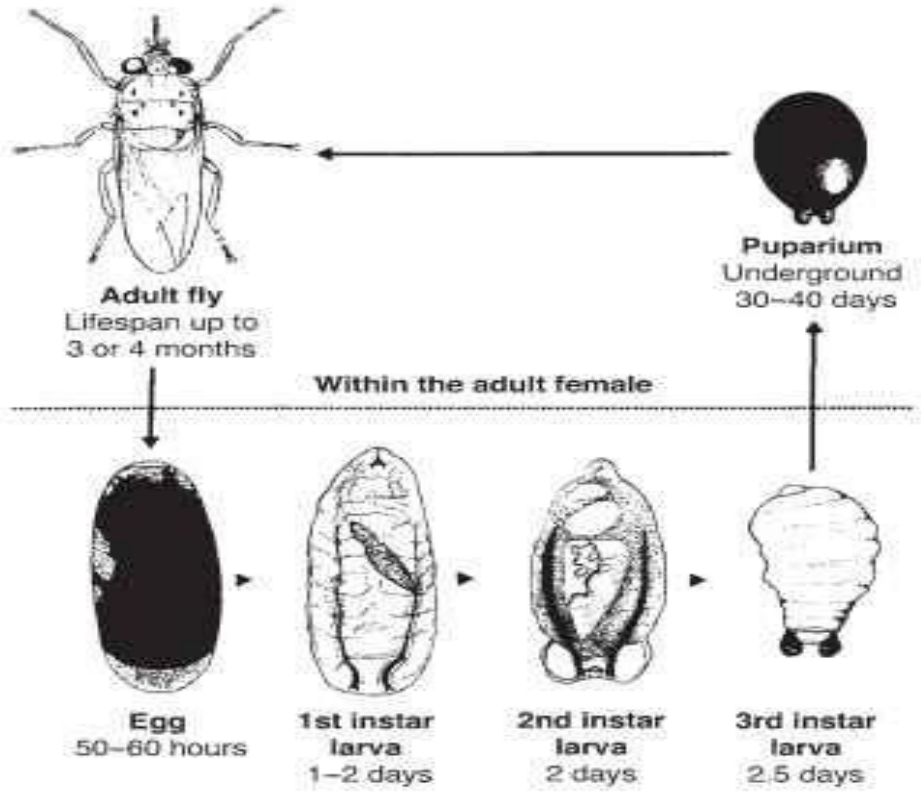

Figure 2: Life cycle of tsetse fly. The egg of a mature female fly is fertilized *in utero* and develops into a 1st instar larva. After about 2 days, the larva develops into a 2nd instar followed by a 3rd instar, which is deposited on the ground where it pupates. It then develops into an adult fly. This image was adapted from (Crankshaft Publishing n.d.)

1.4.0 Tsetse fly control

Current tools for trypanosomiasis control are limited as trypanocidal drugs display undesirable side effects, there are emerging reports of drug resistance and it is difficult to develop vaccines due to parasitic antigenic variation (Brun et al. 2010). Consequently, vector control is the preferred alternative (Aksoy 2003).

In the past, tsetse flies control methods were based on elimination of game-reservoir hosts of the parasite, creation of fly barriers and clearing of bush to destroy tsetse breeding sites (Aksoy 2000). However, majority of these methods were dropped as they contributed to environmental pollution (Brightwell, Dransfield, and Kyorku 1991). Other commonly used control interventions include sterile insect technique (SIT), push-pull method and the use of insecticides either through sequential aerosol spraying technique (SAT); ground

spraying; insecticide-treated targets. Successful application of these methods depends on their sustainability.

1.4.1 Insecticides

This involves use of chemicals to control tsetse flies. Commonly used insecticides include Dichlorodiphenyltrichloroethane **(**DDT), dieldrin and endosulfan. Extensive insecticide ground spraying was used to control tsetse in Zimbabwe and Nigeria (Gooding and Krafsur 2005). SAT is also used and mainly target adult flies. Its main advantage is that it can effectively clear large areas in a relatively short time although it is capital intensive. Spraying cycles vary between 16-18 days depending on prevailing temperatures (Kappmeier and Nevill 1999). Pour-ons or selective application of insecticides to backs, legs or belly of cattle on which tsetse feed is another effective means of tsetse control that minimises pollution of the environment with the insecticide (World Health Organization 2013).

1.4.2 Traps and targets

Both methods use attractants to attract and kill tsetse flies. Traps consist of blue and black colored cloths that attracts tsetse flies. Black and blue colors are the best attractant colors for tsetse flies. Usually, blue is used to attract tsetse to an object while black to get them land on it. (Green 1994). The traps also have other attractants like urine and chloroform to enhance their effectiveness. Once they are trapped, the flies are contained in special cages where they die of heat or starvation (Brightwell et al. 1991). The effectiveness of these methods can be improved by use of appropriate odour bait. Although these methods are inexpensive and simple to use, they cannot be used for large scale control programs (Kappmeier and Nevill 1999).

1.4.3 Live bait techniques

This involves treating cattle with appropriate insecticide formulations. It is achieved by means of cattle dips, pour-ons or spraying with veterinary formulations. These methods are highly effective against tsetse ticks, and other flies (de Beer, Venter, and Vreysen 2015).

1.4.4 Sterile insect technique (SIT)

This technique exploits the unique mating biology of tsetse, whereby female flies mate only once in their entire lifetime. Male flies are usually sterilized by irradiation in the laboratory and released to mate with wild females. Consequently, females mated with sterile males are unable to reproduce. This ultimately leads to reduced tsetse populations and reduces transmission of trypanosomes. Unlike other tsetse control techniques, SIT does not affect non-target organisms. It was used for the successful elimination of *G. austeni* in Unguja island of Zanzibar in 1995. As a result, trypanosomiasis was kicked out of this geographical area (Vreysen et al. 2000). Mass production of tsetse flies in the laboratory is a major constraint for this method due to bacterial symbionts and salivary gland hypertrophy virus that regulate the reproductive capacity of infected flies (Abd-Alla et al. 2013).

1.5 Hypothesis

Bracoviral DNA is only present in the *Glossina morsitans morsitans* genome not in other tsetse fly (*Glossina*) species genomes found in East Africa.

1.6.0 Objectives

1.6.1 General objective

To identify PDV DNA in the genomes of five tsetse fly species (*G. austeni, G. brevipalpis, G. f. fuscipes, G. m. morsitans* and *G. pallidipes*).

1.6.2 Specific objectives

1. To determine the presence of PDVs in the four tsetse fly species (*G. austeni, G. brevipalpis, G. f. fuscipes* and *G. pallidipes*) relative *to G. m. morsitans* genome.

2. To determine the prevalence and genetic diversity of PDVs in *G. austeni, G. brevipalpis, G. f. fuscipes, G. m. morsitans and G. pallidipes* species found in East Africa.

1.7 Justification

True mutualism exists between polydnaviruses and parasitoid wasps as parasitoid survival depends on viral infection of the wasp's host while viral transmission depends on parasitoid survival. The discovery of putative bracoviral genes in *Glossina morsitans morsitans* raises the question of whether there exist mutualism between tsetse fly (*Glossina* species) and polydnaviruses. Characterization of PDVs in *Glossina* species has the potential of revealing more knowledge on tsetse biology as well as providing novel tsetse fly control strategies. Current remedy for trypanosomiasis relies on trypanocidal drugs that were developed about three decades ago. These drugs cause undesirable side effects and the emergence of multi-drug resistant trypanosome strains has been reported in most parts of sub-Saharan Africa, where the disease is endemic. Hence, there is an urgent need for research on alternative strategies to combat the disease. Proper understanding of tsetse biology and genetics holds the promise of improved and sustainable vector and disease control strategies.

2.0 LITERATURE REVIEW

2.1 The tsetse fly genome

The recently sequenced *G. m. morsitans* revealed a size of 366 Megabase pairs (Mb) with 12,308 predicted protein encoding genes (International Glossina Genome Initiative 2014). This genome is more than twice the size of the *Drosophilla melanogaster* genome (116.8 Mbp) (Misra et al. 2002) and slightly bigger than that of *Anopheles gambiae* (278 Mbp) (Mongin et al. 2004). The bigger *Glossina* genome is due to larger introns and increased size of intergenic sequences, mostly transposons and other repetitive sequences (International Glossina Genome Initiative 2014). The bigger genome may explain *Glossina'*s unique characteristics that include obligate microbial symbioses, viviparous reproduction, lactation and an obligate hematophagy lifestyle. Interestingly, tsetse fly is the only insect known to lactate and nourish its young ones with milk, a characteristic of mammalian lifestyle (Yang et al. 2010). Availability of tsetse genome provides foundation for research into trypanosomiasis prevention, control and better understanding of tsetse biology. Four other *Glossina* genomes, that is *G. austeni*, *G. brevipalpis*, *G. f. fuscipes* and *G. pallidipes*, and two related Dipterans, a non-vector obligate blood feeder (*Stomoxys calcitran*) and a non-blood feeding mechanical vector (*M. domestica*), have since been sequenced by the International Glossina Genomics Initiative (IGGI) (Aksoy 2010).

Analysis of a group of genes lacking insect relatives in whole genome sequence of the tsetse fly *Glossina morsitans morsitans* has revealed the presence of many bracoviral genes that are spread over 151 genomic scaffolds, in addition to a large DNA hytrosavirus, the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) (International Glossina Genome Initiative 2014). Homology search of these genes [Basic Local Alignment Search Tool (BLAST), E values of <1e-50] reveals that they are similar to those identified in the parasitic braconid wasps; *Cotesia congregata*, *Microplitis demolitor* and *Glyptapanteles flavicoxis*. The presence of these genes that are homologous to the braconid wasp's genes suggests that tsetse fly may have been parasitized by an anonymous braconid wasp or lost PDV mutualism to adapt to larviparity (International Glossina Genome Initiative 2014).

2.2 Origin and evolution of insect dsDNA viruses

Viruses are the most common biological entities on earth and together with transposons and plasmids, parasitizes on all cellular organisms (Koonin, Krupovic, and Yutin 2015). This implies that virus-host coevolution has been the mode of evolution of life ever since its origin (Koonin, Dolja, and Krupovic 2015). Viral genomes are about 2 kilobases in size and are smaller compared to cellular life forms. The recent discovery of several groups of giant viruses about 2 megabases in size has expanded the viral genome size range. Viruses with large genomes possess many genes acquired from the hosts at different stages of evolution (M Krupovic and Koonin 2015).

An evolutionary connection has recently been shown to exist between virophages and eukaryotic dsDNA transposons of the Maverick (Polinton) family. Maverick origin dates back to the origin of eukaryotes. They seem to combine features of both viruses and transposons (Koonin, Krupovic, et al. 2015). Under certain conditions, they can produce virions that could infect new hosts as they encode viral capsid proteins. Interestingly, evolutionary connections have confirmed that Mavericks evolved directly from bacteriophages and evolved further to give rise to most of the large dsDNA viruses of eukaryotes, plasmids as well as transposons (Mart Krupovic and Koonin 2015).

2.3 Mutualism between dsDNA viruses and eukaryotes

Mutualism between eukaryotes and viruses is a rare biological event as viruses mostly have parasitic associations with their hosts (Roossinck 2015). However, ancient relationships between viruses and host have resulted in the adaption of the virus by the host in a process called symbiogenesis (Roossinck 2011). Several mutualistic virus host relationships have been discovered in different hosts including bacteria, insects, plants etc. For example, Polydnaviruses that are required for the survival of parasitoid wasp egg in insect hosts (Espagne, Dupuy, Huguet, Cattolico, Provost, Martins, Poirie, et al. 2004), Retroviruses are involved in mammalian placental evolution (Zhang, Shi, and Liu 2015) and Pararetroviruses that protect plants against pathogenic viruses (Hohn 2015). It has been postulated that modern symbiotic viruses are remnants of ancient viruses. Availability of genome sequence information has revealed many symbiotic viruses. Mutualism between Polydnaviruses (PDVs) and parasitoid wasps is the most studied. These viruses have been shown to form a symbiotic association with thousands of parasitoid wasps that parasitize lepidopteran hosts (Strand and Burke 2012). Therefore, PDVs presents an opportunity for

in depth analysis of the genome rearrangements involved in the evolution of viruses from pathogenic ancestors to mutualists (Espagne, Dupuy, Huguet, Cattolico, Provost, Martins, Poirié, et al. 2004).

2.4 Polydnaviruses

The family *Polydnaviridae* consists of unique insect viruses, in terms of their genomic molecular features and their obligate association with parasitoid wasps (Jancek et al. 2013). It is divided into two genera, Bracoviruses and Ichnoviruses that are associated with braconid and ichneumonid wasps respectively (Louis et al. 2013). The two species evolved from distinct evolutionary origins whereby the bracovirus group has a common ancestor and form a monophyletic group known as microgastroid complex (Theze et al. 2011). It has been hypothesized that the bracoviral ancestor, referred to as nudivirus - a sister group of baculoviruses, was initially integrated into the genome of the ancestor wasp approximately 100 million years ago (Bezier et al. 2013). Symbiosis between the wasp and PDVs was assembled over time from genes of both wasp and viral origin. It is postulated that the functional association between the symbionts dates around 73.7 ± 10 million years ago (Whitfield 2002).The PDVs have functionally been annotated in the wasp *Cotesia congregata* where they are essential for the successful parasitism of the lepidopteran larvae, *Manduca sexta* (Chevignon et al. 2014). Viral replication and particle production occurs only in females in a region of the wasp ovary called calyx, during pupal and adult stages. Virions containing multiple segmented double stranded DNA and encoding no viral structural proteins, constitute the major component of calyx fluid that is injected together with one or more eggs into the parasitized caterpillar host during wasp oviposition. The viral particles are replication deficient and are vertically transmitted as endogenous “provirus" integrated in the wasp genome (Espagne 2014). Their dsDNA is only expressed by their host's cellular replication machinery (Burke et al. 2013). The viral particles encode proteins that compromise the host immune defenses thus preventing recognition, encapsulation and destruction of the parasitoid eggs and larvae. This allows the wasp eggs to survive and develop, while the host ultimately dies (Burke et al. 2013). However, the lack of genes coding for structural proteins has elicited a debate on whether bracoviruses are of viral origin or are a 'genetic secretion' of the wasps (Bezier et al. 2009). An example is the bracoviral virion DNA in the wasp *Cotesia congregata* that consists of cellular genes of wasp origin, several viral genes and transposable elements(Drezen et al. 2006).

Phylogenetic analysis of its functional bracoviral genes has highlighted sugar transporters of wasp origin (Desjardins et al. 2007). Transfer of these wasp genes into the provirus was facilitated by transposable elements, and subsequently followed by co-evolution with the host's genome, to become more specialized (Espagne, Dupuy, Huguet, Cattolico, Provost, Martins, Poirie, et al. 2004; Jancek et al. 2013).

2.5 Summary

Recently, successful delivery of bracoviral genes by wasps into lepidopteran larvae has inspired notable agricultural applications. Currently, TSP14-producing transgenic plants are being used to effectively reduce *Manduca sexta* growth and development, thus protecting the plants from insect damage (Maiti et al. 2003). In the present study, a total of 310 genes from *G. m. morsitans* that bear highest similarity to bracoviral genes were analyzed and their orthologs investigated in *G. austeni*, *G. brevipalpis*, *G. f. fuscipes* and *G. pallidipes* and *M. domestica*. Bracoviral genes identified in *G. austeni*, *G. brevipalpis*, *G. f. fuscipes* and *G. pallidipes* were investigated in terms of their prevalence, population dynamics/genetics. Also the origin, replication machinery, genetic diversity relative to wasp bracoviruses and whether the DNA is genetically active was examined. This newfound knowledge is useful in better understanding tsetse biology, and novel possible intervention strategies.

3.0 METHODOLOGY

3.1 Identification of bracoviral sequences in *G. m. morsitans*

Bracoviral related sequences (n=310) present in the *G. m. morsitans* proteome, previously described by the International Glossina Genome Initiative (International Glossina Genome Initiative 2014), were retrieved using a Perl script. These sequences were aligned against NCBI's Conserved Domain Database (CDD) (28 may 2015 release) (Marchler-Bauer et al. 2011) using BLASTP and DELTA BLAST algorithms, to identify sequences showing highest similarity to bracoviral genes.

Identified sequences were confirmed by querying Pfam-A.hmm (13 March 2013 release) downloaded from Pfam 27.0 (Finn et al. 2014) protein families database. Searches against the Pfam-A.hmm were done using hmmscan command (cut off e-value of <1.0e-10) in HMMER version 3.1b1 (Finn et al. 2014).

3.2 Homology search across *Glossina spp* and *Musca domestica*

Deduced *G. m. morsitans* bracoviral protein sequences were aligned against five *Glossina* spp. (*G. austeni, G. brevipalpis, G. f. fuscipes*, *G. m. morsitans* and *G. pallidipes*) and *M. domestica* scaffold datasets using tBLASTn algorithm. A cut off e-value of 1.0e-10 was set.

3.3 Domain analysis

Next the resulting hits in each species were aligned against NCBI's non-redundant protein database using BLASTX to search for the presence of conserved domains. Sequences lacking domain were well interrogated to avoid false negatives. The domain sequences were extracted using Perl scripts and used for further analysis.

3.4 Phylogenetic reconstruction of bracoviral sequences

DNA pol B2, parvo coat N, pox A32, and an integrase-like sequences from *Glossina spp.* (*G. austeni, G. brevipalpis, G. f. fuscipes, G. m. morsitans* and *G. pallidipes*), *M. domestica*, and orthologs from parasitoid wasps (*Cotesia congregata, Glyptapanteles flavicoxis* and *Nasonia vitripennis*) and a beetle (*Tribolium castaneum*) (Dupuy et al. 2011), were used for phylogeny reconstruction. Sequences were aligned using Multiple

Sequence Comparison by Log Expectation (MUSCLE) (Edgar 2004). Maximum likelihood (ML) phylogenetic analysis of the multiple aligned sequences with bootstrap values of 100 replicates was performed using PHYML version 3.5 (Dereeper et al. 2008)

3.5 Magnitude and direction of selection pressure

The magnitude and direction of selection pressure on the bracoviral sequences was tested based on the ratio (ω) of non-synonymous to synonymous rates (ω = dN/dS), where dN is the average number of non-synonymous substitution per non-synonymous site and dS is the average number of synonymous substitutions per synonymous site. If ω = 1, amino acid substitution are assumed to be largely neutral, ω > 1 is indicative of positive selection whereas ω < 1 is evidence of negative or purifying selection. A sequence alignment was generated for each loci consisting of *Glossina* and *Musca* bracoviral sequences as well as sequences from parasitoid wasps (*Cotesia congregata*, *Glyptapanteles flavicoxis* and *Nasonia vitripennis*) and a beetle (*Tribolium castaneum*, and each alignment uploaded to the SNAP (Synonymous Non-synonymous Analysis Program) program (Korber, Allen G., and Gerald H. 2000) (www.hiv.lanl.gov). Sequences (length = 50bp) flanking either side of each sequence were extracted using Perl scripts and the magnitude selection pressure examined. This was also evaluated for each clade per loci under examination.

3.6.0 Gene validation

To confirm the existence of the sequences identified using bioinformatics analysis, polymerase chain reaction (PCR) amplification of the genes in the DNA of the different species was performed. The insects used were collected from Kenya (*G. brevipalpis, G. f. fuscipes* and *G. pallidipes)* and Malawi (*G. m. morsitans*). PCR products were sequenced and the sequences compared to those in the database. A list of primers used is provided in Table 1. Insect DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Sanger sequencing on both strands of the PCR products was outsourced from Macrogen (Seoul, South Korea). The sequences obtained were processed and compared with the original sequences to determine genetic diversity.

Table 1: List of genes, primers and the expected product size.

'For' stands for Forward primer and 'Rev' stands for Reverse primer.

Gene	Primer sequence	Product size
Pox A32	For: TTCGAAATCTTTATTTCAACCAAAA	270
	Rev:TTATCACGTATTAAATGTTTGGGAAT	
Integrase	For: AAGCCGATCTGGTTGAAATG	328
	Rev: CCACATCATTCCTTTTAATGTCC	
DNA pol B2	For: TGAAACGTAAGGGAGTCTTTCCT	537
	Rev: AGGGTAAACATTGCGACATGG	
Parvo coat N	For: TGAAACGTAAGGGAGTCTTTCCT	244
	Rev: AGGGTAAACATTGCGACATGG	

3.6.1 DNA Extraction

Each insect sample was first divided into small parts and transferred to a labeled 1.5ml Eppendorf tube. Whole DNA was extracted using the alcohol precipitation method as described in the QIAGEN's "Purification of total DNA from insects using the DNeasy Blood & Tissue Kit" (Qiagen, Hilden, Germany).

The samples in the Eppendorf tubes were lysed using 180µl of ATL buffer and protein contaminants removed using 20µl of proteinase K. The tubes were incubated overnight at 56 ^{0}C to allow for complete lysis. The mixture was precipitated using 200 µl of absolute ethanol, transferred to spin columns in collection tubes and washed with AW1 and AW2 buffers. The mixture was centrifuged at 8,000 rpm at each wash for 1 minute and 3 minutes

respectively. The DNA was finally eluted using 40µl Elution buffer into 1.5ml Eppendorf tubes.

3.6.2 Polymerase chain reaction (PCR)

Polymerase chain reaction was done using DreamTaq mastermix already contains DNA polymerase enzyme, the dNTPs and the magnesium chloride ions. Four sets of primers were designed to amplify the sequences across all the samples. Conserved flanking regions of the target regions were identified across the species and used for primer design. The reaction mix consisted of 25 µl of Dreamtaq Mastermix 2x, 2.5 µl of 10 mM of each primer and 15 µl of PCR water to give a total volume of 50 µl. The list of primers used is provided in Table 1. The amplification involved initial DNA template denaturation for 3 minutes at 95^0C, followed by 40 cycles of 30 seconds denaturation at 95^0C, 30 seconds of annealing (55^0C for DNA Pol B2, 51^0C for Parvo coat N, 54^0C for Pox A32 and 55^0C for Integrase), and 30 seconds of elongation at 72^0C. The final extension was 7 minutes at 72^0C.

3.6.3 Gel electrophoresis and Gel extraction of PCR products

1% agarose gel was prepared by dissolving 1 gram of agarose powder into 100ml of 1 x TAE buffer. The mixture was stirred, boiled on a microwave and allowed to cool before 5µl of 0.5 µg/mL ethidium bromide was added. It was then poured into a casting tray in order to polymerize. The PCR products were then loaded into the gel alongside a molecular weight DNA marker to confirm the size. Electrophoresis was set at 80 volts for 1 hour, followed by visualization under an Ultra Violet illuminator.

3.6.4 Sequencing

Sanger sequencing on both strands of the PCR products was outsourced from Macrogen (Seoul, South Korea). The sequences obtained were processed and compared with the annotated sequences (sequences obtained from bioinformatics analysis) to determine genetic diversity.

4.0 RESULTS

4.1 Annotation of *G. m. morsitans* bracoviral sequences

Previous work identified bracoviral homologues (n=310) in *G. m. morsitans* (International Glossina Genome Initiative 2014). In our analysis, most of these bracoviral sequences had orthologs in other *Glossina* species *and M. domestica* (BLAST E-value <1e-50). In this dataset, most double stranded endogenous viral sequences identified had genes ubiquitously present in the host insect's genome and other metazoans in general. An example is the protein tyrosine phosphatase gene family, which plays a key role in regulating signal transduction pathways. (Espagne, Dupuy, Huguet, Cattolico, Provost, Martins, Poirie, et al. 2004). This complicated unbiased detailed analysis of bracoviral evolutionary biology. This challenge was circumvented by focusing on PDV sequences with virus-related genes previously described as good evolutionary biology models (Pritham, Putliwala, and Feschotte 2007). This work used transposable elements designated Mavericks (Herniou et al. 2013) to underpin the genetic diversity of PDVs in the recently sequenced Dipteran genomes. These Maverick sequences consist of four genes: a DNA polymerase B2 involved in DNA excision repair and initiation of replication (DNA Pol B 2) (Drezen et al. 2006) the N terminal region of the parvovirus coat protein VP1 (Parvo coat N) that is important for virion retention and transduction by insect vectors (Spitzer, Parrish, and Maxwell 1997); a retroviral-like integrase; and poxvirus A32 protein (Pox A32) that encodes an ATPase involved in virion DNA packaging (Cassetti et al. 1998). Eight sequences that are highly similar to bracoviral maverick sequences were identified (Table 2). Domain searches against Conserved Domain Database (CDD) and Pfam-A database confirmed that the sequences had the maverick related domains; DNA Pol B2, Parvo_coat_N and Pox_A32 (Table 2). Search for more bracoviral genes in sequences neighboring the 8 identified sequences revealed the presence of the retroviral-like-integrase.

Table 2: BLASTP results of bracoviral sequences in NCBI. The score, identity and E-value results were derived from *Cotesia congregata* bracovirus (CcBv) and *Glyptapanteles flavicoxis* hits.

Sequence id	Domain (CDD)	Score	Identity	E-value	Domain (Pfam)
GMOY004963	DNA_Pol_B2	249	47%	3e-71	#
GMOY010042	#	140	42%	6e-33	#
GMOY010043	Parvo_coat_N	164	53%	2e-43	Parvo_coat_N
GMOY004962	Parvo_coat_N	167	44%	8e-46	Parvo_coat_N
GMOY009082	Parvo_coat_N	160	54%	2e-45	Parvo_coat_N
GMOY000040	Pox_A32	267	60%	1e-82	Pox_A32
GMOY011252	#	248	66%	2e-73	#
GMOY009081	#	112	52%	2e-26	#

Domain was not detected

4.2 Homology searches and sequence analysis

Multiple maverick associated genes were detected across Glossina spp. (*G. austeni, G. brevipalpis, G. f. fuscipes, G. m. morsitans* and *G. pallidipes*) and *M. domestica* scaffold datasets. Previous studies have shown that the maverick associated genes are found in close vicinity to each other (Dupuy et al. 2011). Hence, only sequences occurring on the same scaffold were assumed to be maverick associated genes and considered for downstream analysis [Table 3]. Domain analysis of this hits using BLASTX revealed numerous domains across the datasets under study (Figure 3). DNA polymerase B2 sequences did not contain the domain sequences and thus full sequences were used for downstream analysis. Analysis of the average length of each gene across all the species and within each species showed that parvo coat N gene had the most conserved length across all the species (Figure 4). *G. austeni* was shown to have the highest number of Maverick associated genes (n=18) and *G. brevipalpis* the least (n=6).

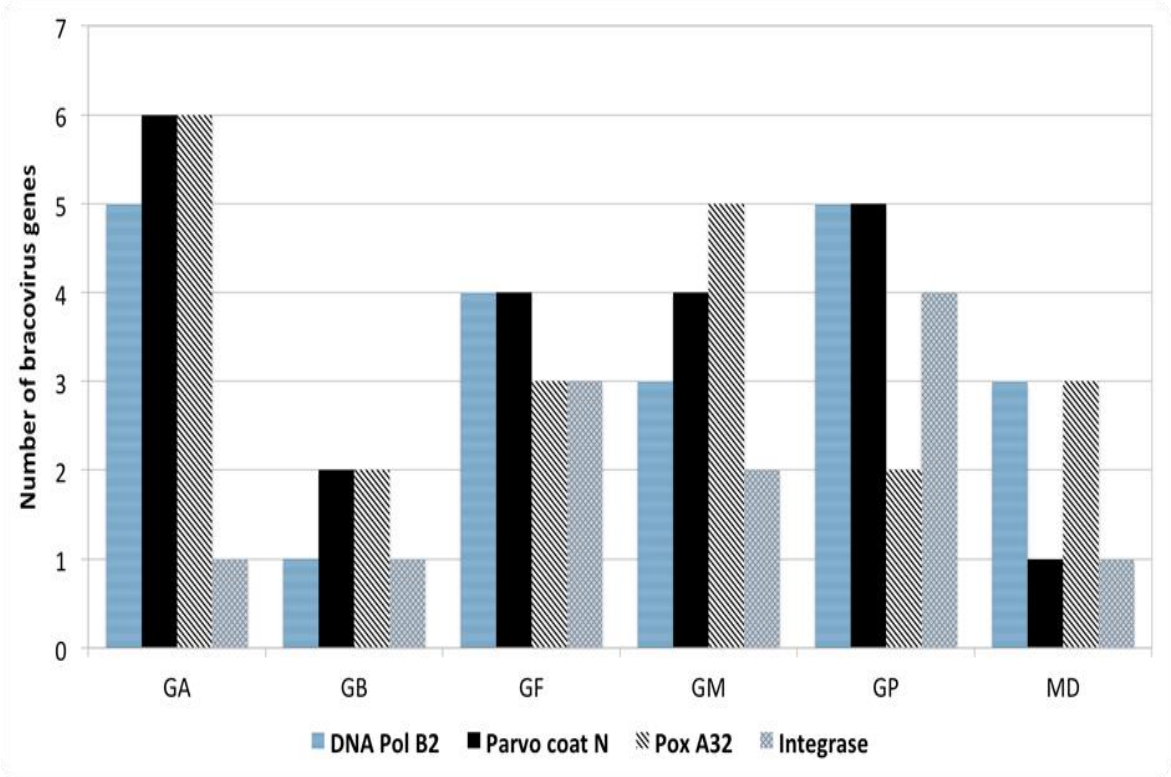

Figure 3: **Distribution of bracoviral associated maverick genes across *G. austeni, G. brevipalpis, G. f. fuscipes, G. m. morsitans* and *G. pallidipes* and *M. domestica.***

These genes are widely distributed in the respective species genomes (Table 3). Parvo coat N gene has the most conserved length across all species, emphasizing the importance of understanding the role of selection on genetic diversity at this locus.

Table 3. The distribution of Maverick associated genes in different Dipteran species genomes. The description is as follows: species abbreviations, genome scaffold, and start and end base pair position for each gene. The abbreviations describe genes from the sis different species: *Glossina austeni* (GA), *Glossina brevipalpis* (GB), *Glossina fuscipes fuscipes* (GF), *Glossina morsitans morsitans* (GM), *Glossina pallidipes* (GP), and *Musca domestica* (MD). Genes present on the same scaffold are in the same row. The stars indicate the number of genes present per scaffold, and if all four Maverick associated gene loci are present are highlighted in red.

ATPase	DNA Pol B2	Integrase	Parvo|coat
GA_Scaffold123_267546_267791			GA_Scaffold123_267022_267135
GA_Scaffold128_439833_440204	GA_Scaffold128_442748_443433	GA_Scaffold128_436565_436849	
GA_Scaffold145_537273_537644			GA_Scaffold145_536899_537012
GA_Scaffold489_25753_26124	GA_Scaffold489_22366_22876		GA_Scaffold489_25340_25453
GA_Scaffold490_168892_169263	GA_Scaffold490_171851_172525		GA_Scaffold490_169560_169673
	GA_Scaffold598_6228_6889		GA_Scaffold598_7751_7864
GA_Scaffold616_15810_16181	GA_Scaffold616_18759_19445		GA_Scaffold616_16481_16594
GB_JFJS01016125_913_1281			GB_JFJS01016125_486_599
GB_Scaffold12_3288037_3288387	GB_Scaffold12_3291037_3291712	GB_Scaffold12_3285226_3285546	GB_Scaffold12_3288708_3288821
GF_Scaffold248_381724_382092	GF_Scaffold248_378410_379091		GF_Scaffold248_381308_381421
	GF_Scaffold27_1373917_1374596		GF_Scaffold27_1371730_1371843
GF_Scaffold287_101580_101951		GF_Scaffold287_97841_98170	
GF_Scaffold425_99135_99485	GF_Scaffold425_96436_97107	GF_Scaffold425_102216_102545	
		GF_Scaffold61_279873_280202	GF_Scaffold61_276223_276336
	GF_Scaffold849_8684_9366		GF_Scaffold849_6563_6676
GM_scf7180000638930_1245_1616			GM_scf7180000638930_1917_2033
GM_scf7180000648346_202377_202748	GM_scf7180000648346_204796_205482		GM_scf7180000648346_203050_203166
GM_scf7180000650811_18055_18423		GM_scf7180000650811_21003_21287	
GM_scf7180000652025_44734_45102		GM_scf7180000652025_48275_48411	GM_scf7180000652025_44318_44431
GM_scf7180000652170_14995173_14995511	GM_scf7180000652170_14983389_14984075		GM_scf7180000652170_14986278_14986391
	GP_Scaffold191_79111_79387	GP_Scaffold191_88004_88324	
	GP_Scaffold356_175533_176159	GP_Scaffold356_181878_182162	GP_Scaffold356_178330_178443
	GP_Scaffold443_148653_149279		GP_Scaffold443_151348_151461
	GP_Scaffold553_1361_1841	GP_Scaffold553_7428_7694	GP_Scaffold553_3941_4054
GP_Scaffold577_68834_68989			GP_Scaffold577_67756_67869
GP_Scaffold72_1110090_1110458	GP_Scaffold72_1106846_1107522	GP_Scaffold72_1113448_1113732	GP_Scaffold72_1109674_1109787
MD_KB854650_33444_33812	MD_KB854650_28137_28460		
MD_KB857874_8388_8777	MD_KB857874_20994_21181		
MD_KB860254_44736_45077	MD_KB860254_45831_46154		
		MD_KB861321_912_1055	MD_KB861321_3188_3301

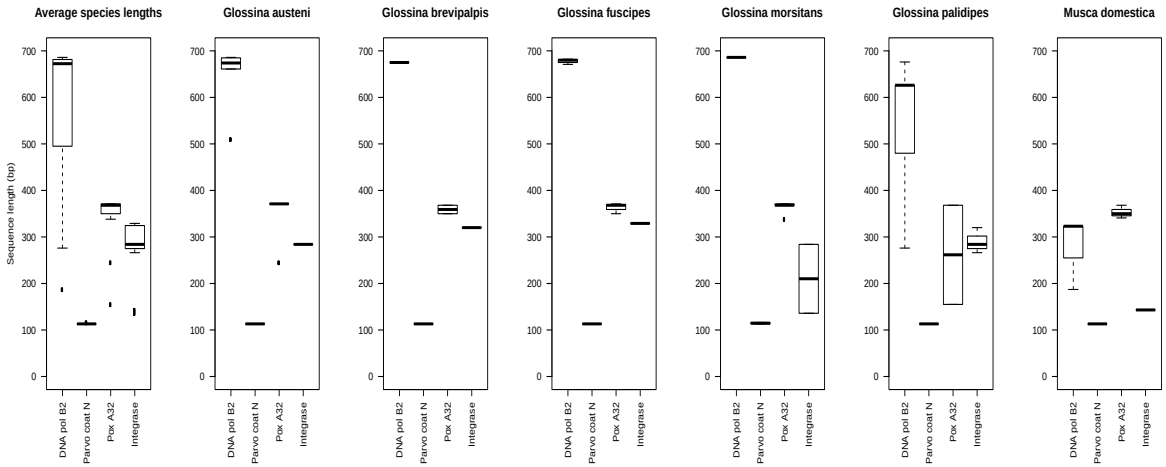

Figure 4: Maverick associated genes sequence lengths. Box plots depicting the entire range of gene lengths and the average gene length for each of the Maverick associated gene: DNA polymerase B2, parvo coat N, pox A32 and integrase. The plots show the average length for each gene across all the species, and within each species. The parvo coat N gene has the most conserved length across all species.

4.3 Phylogeny of Dipteran bracoviral sequences identifies two major clades

Phylogenetic reconstruction of each of the four Maverick associated Dipteran gene loci alongside orthologs from parasitoid wasps (*Cotesia congregata, Glyptapanteles flavicoxis and Nasonia vitripennis*) and a beetle (*Tribolium castaneum*) identified two major bracoviral clades, designated clade 1 and 2 respectively (Figure 5). Interestingly, Dipteran sequences within each of the two clades consistently cluster together across all four loci implying similarity in direction of selection pressure. There are notable differences in the branching times of the clades for each gene locus. Clade 1 is more recent for DNA polymerase B2 and parvo coat N compared to the retroviral-like integrase and the poxvirus A32 protein. The opposite is true for clade 2. The direction and magnitude of natural selection acting on these loci was evaluated using the dN/dS ratio. The genes were under varying magnitudes of purifying selection (Table 4). This is similar in the flanking sequences suggesting that maverick associated genes were inserted at conserved regions of the host's genome (Figure 6). Terminal inverted repeats which are associated with transposable elements were to be missing.

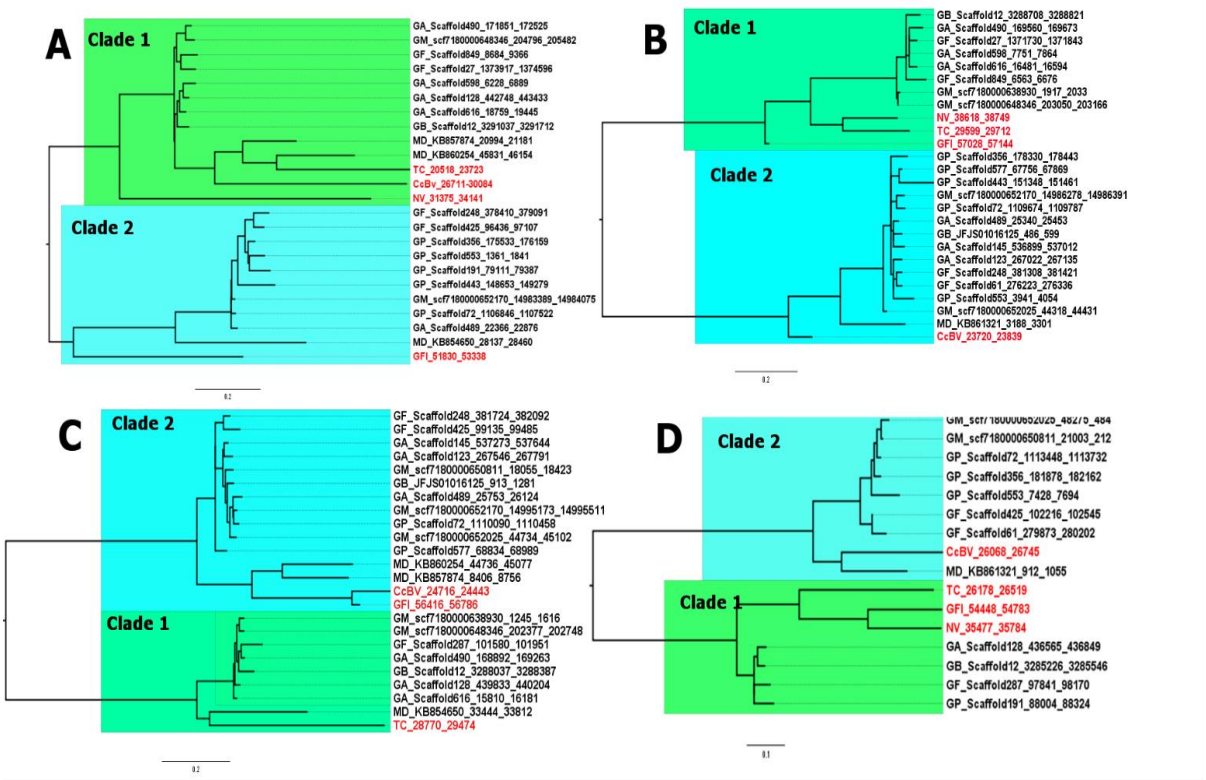

Figure 5: Phylogenetic reconstruction Maverick associated genes. Trees for each gene locus were reconstructed as follows: (A) DNA polymerase B2 genes, (B) parvo coat N genes, (C) pox A32 and (D) integrase. Orthologs at each loci separated in to two distinct clades (clade 1 and 2). Sequences that clustered together in each clade are highlighted in blue and green. Orthologs from parasitoid wasps (*Cotesia congregata* (CcBV), *Glyptapanteles flavicoxis* (GFl) and *Nasonia vitripennis* (NV)) and a beetle (*Tribolium castaneum* (TC)) are highlighted in red. The scale bar represents the number of substitution per site.

4.4 Direction and magnitude of selection pressure

Selection pressure acting on the maverick associated Dipteran loci was evaluated based on the ratios (ω) of non-synonymous to synonymous sites (dN/dS). The ω for integrase, parvo coat N and DNA pol B2 were 0.85, 0.77 and 0.66 respectively, suggesting that these loci were under purifying selection. However, the pox A32 loci had a ω of 1.1011, which suggests the possibility of positive selection (Figure 6). The ω values for the regions flanking the integrase and the DNA pol B2 loci were relatively higher than the values for the respective loci. This suggests that the flanking regions experienced different levels of selection as compared to the loci (Figure 6). Phylogenetic analysis revealed two distinct clades that had sequences from conserved scaffolds across the four loci (Figure 5). Although the pox A32 gene was under positive selection (ω = 1.1011), its two clades, i.e. clade 1 and clade 2 had a ω of 0.7807 and 0.1604 respectively (Table 4). In the parvo coat N loci, clade 2 had higher ω of 0.9363 as compared to the overall loci and clade 1 ω of 0.7695 and 0.1754 respectively. Clade 2 had the highest ω of 1.0058 in the DNA pox A32, as compared to the overall loci and clade 2 ω of 0.6639 and 1.0058 respectively (Table 4).

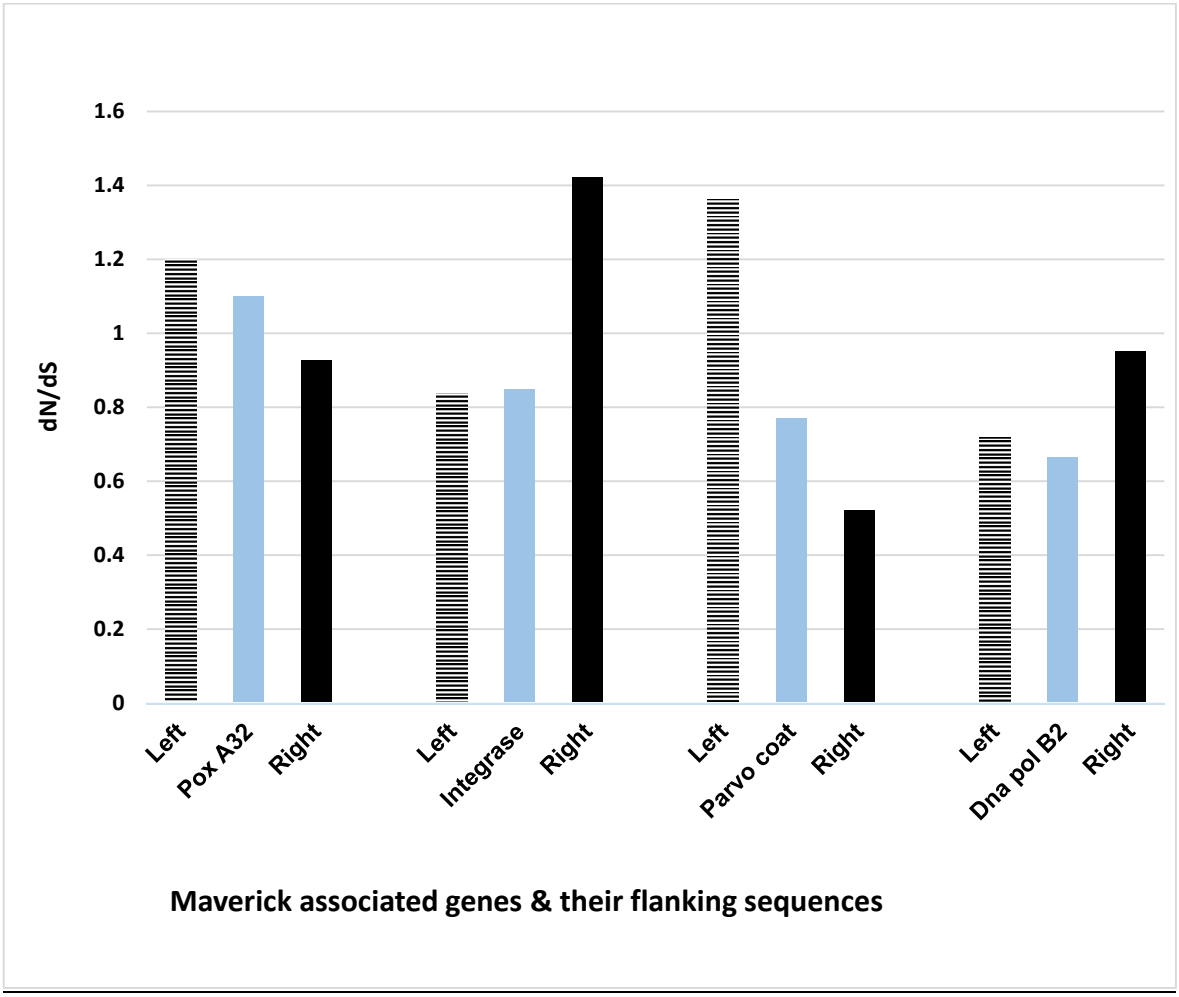

Figure 6: Magnitude and direction of selection pressure at Maverick associated loci and their flanking regions. Shows the magnitude and direction of selection pressure of the flanking regions and the overall loci for all the analyzed bracoviral sequences

Table 4. The direction and magnitude of selection pressure on Maverick associated genes. Shows the magnitude and direction per clade for a given loci, and the overall loci dN/dS values for all the analyzed bracoviral sequences.

Gene	Clade 1	Clade 2	Overall
DNA Pol B2	0.85	1	0.66
Parvo coat N	0.18	0.94	0.77
Pox A32	0.78	0.16	1.1
Integrase	1.21	0.17	0.85

4.5 Detection of Maverick sequences within recent African tsetse flies

Genomic DNA was extracted from a collection of male and female tsetse flies from Kenya (*G. pallidipes, G. brevipalpis and G. fuscipes*) and Malawi (*G. m. morsitans*). Maverick genes were amplified across this variety of tsetse fly species (Figure 8). The results confirmed that the annotated Maverick genes exist in tsetse flies collected from other geographic regions. The PCR products obtained were of the expected sizes (Table 1). The relationship between the amplified products and previously annotated gene sequences was tested using phylogenetic analysis. The amplified products clustered together with sequences from corresponding *Glossina* spp. apart from the pox A32 loci. They also clustered in the most recent clade except for the DNA polymerase B2 gene which clustered in clade 1 (Figure 7).

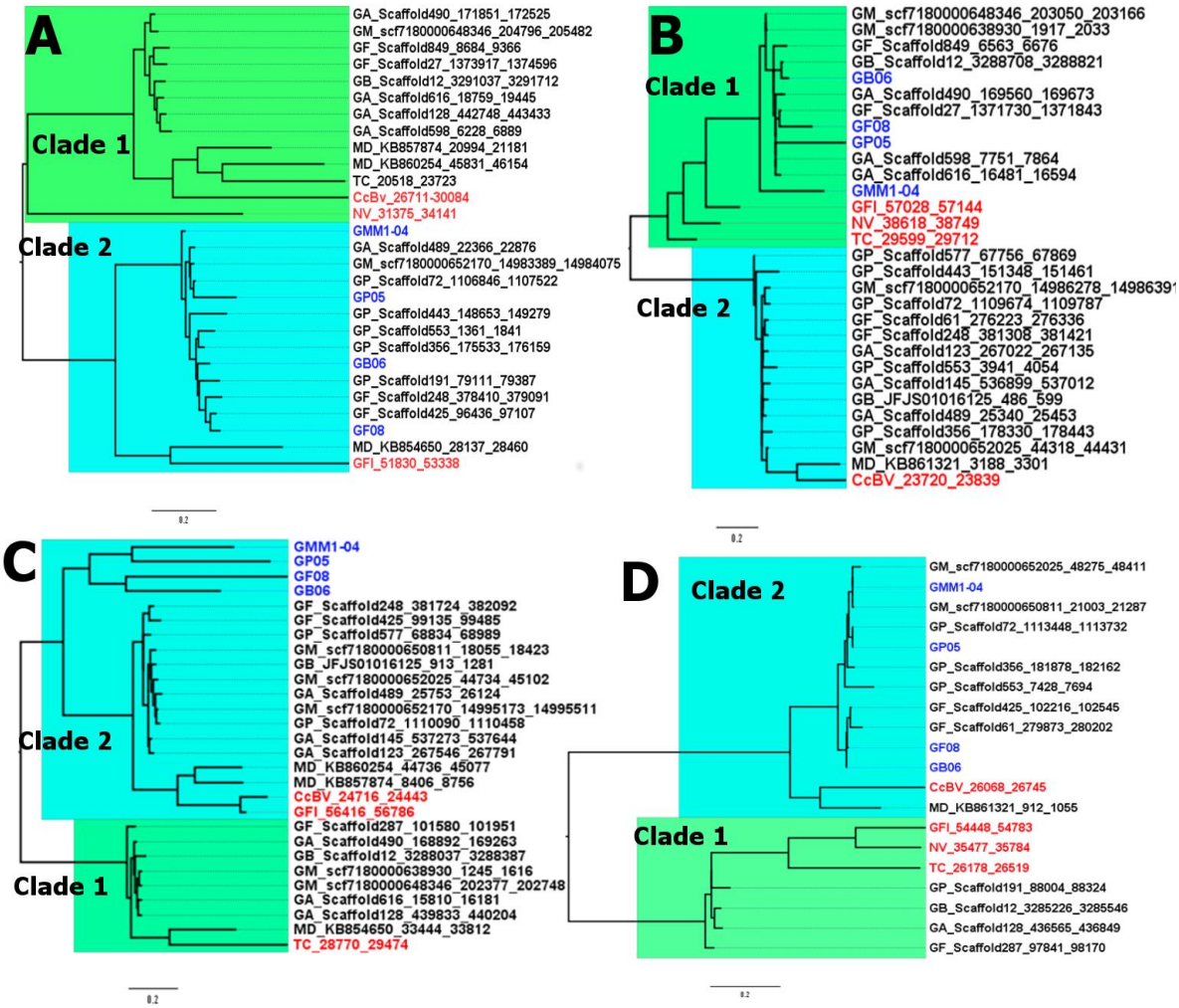

Figure 7: Phylogenetic reconstruction of amplified Maverick genes and the annotated Maverick associated gene sequences Trees for each gene locus were reconstructed as follows: (A) DNA polymerase B2 genes, (B) parvo coat N genes, (C) pox A32 and (D) integrase. Orthologs separated in to two distinct clades (clade 1 (green) and 2 (blue)). Amplified gene products sequences are highlighted in deep blue. Orthologs from parasitoid wasps (*Cotesia congregata* (CcBV), *Glyptapanteles flavicoxis* (GFl) and *Nasonia vitripennis* (NV)) and a beetle (*Tribolium castaneum* (TC)) are highlighted in red. The scale bar represents the number of substitution per site.

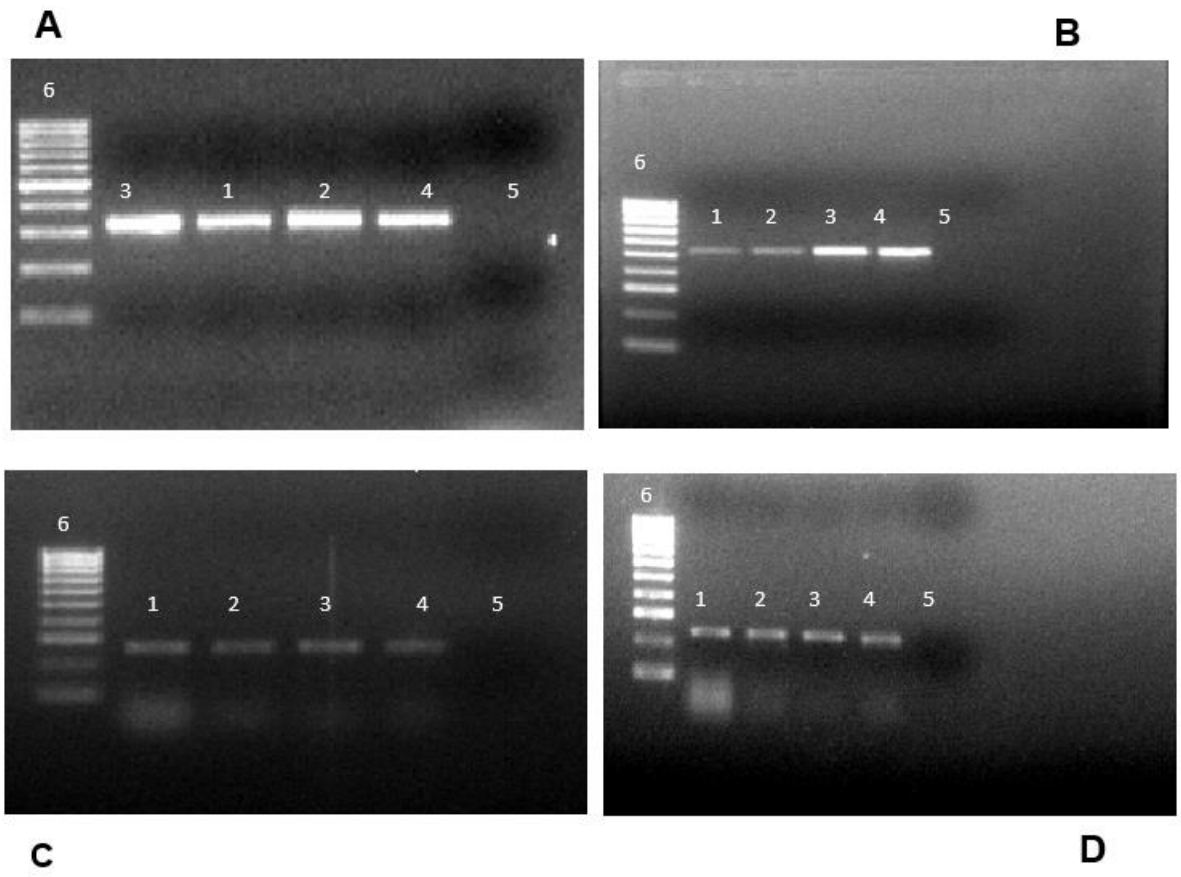

Figure 8: PCR amplification of the four Maverick associated genes. The maverick associated genes PCR amplifications are as follows. (A) Integrase, (B) DNA Pol B2, (C) Parvo coat N, (D) Pox A32. The expected size of the genes is 328, 537, 244 and 270 respectively. The bands represent different samples; (1) *Glossina brevipalpis*, (2) *Glossina fuscipes*, (3) *Glossina morsitans morsitans*, (4) *Glossina pallidipes*, (5) Negative control, (6) 100bp DNA ladder.

5.0 DISCUSSION

Previous studies identified endogenous bracoviral sequences in the genomes of parasitoid wasps and *Glossina morsitans morsitans* (International Glossina Genome Initiative 2014). Bracovirus present a unique symbiotic relationship of eukaryotes co-opting endogenous viruses, and understanding their organisation, may significantly contribute new vector control strategies and the development of vectors for gene therapy (Bezier et al. 2009). An example, is the polydnavirus *Oryctes rhinoceros nudivirus* (OrNV) that has been used as a biological control agent in palm tree farming against rhinocerous beetle (Bezier et al. 2009). However, the prevalence of PDVs in *Glossina* species is unknown, and it is still uncertain whether present day PDVs descended from a single ancestor after initial host integration, or whether it was a single random event in the *G. m. morsitans*. It is also still unclear whether they vary in size, and undergo intra-genomic rearrangements, in their host genomes. The hypothesis that a single integration event of the viral genome occurred before the evolutionary radiation of different insect orders was explored. This study aimed to improve our understanding on PDVs, and why they exist and are retained in tsetse flies, yet they do not exhibit parasitoid reproduction. The evolutionary biology of *Dipteran* bracoviruses was explored using Maverick genes. Mavericks (also known as Polintons) are a group of selfish elements, present in most eukaryotes - including parasitoid wasps (Dupuy et al. 2011), which combine features of bona fide viruses and transposable elements (Koonin, Dolja, et al. 2015). They can subsist as viruses or transposons depending on the combination of genes present (Koonin, Dolja, et al. 2015). Mavericks are thought to be the first group of double-stranded DNA viruses to evolve from bacteriophages (Mart Krupovic and Koonin 2015). They continuously evolve during vertical eukaryotic host transmission, and are actively been involved in shaping the genomes and biology of eukaryotes (Mart Krupovic and Koonin 2015).

Maverick genes were identified in *Glossina* species and *M. domestica*, suggesting that PDVs were integrated into the genomes of Dipterans prior to their diversification. It is unclear when exactly parasitoid wasp and Dipteran PDVs separated and diversified in their different hosts. The findings support suggestions that bracoviruses are descended from a common ancestor in the Paleozoic Era, and raise the possibility of integration of PDVs before the separation of *Hymenoptera*, *Coleoptera*, *Lepidoptera* and *Diptera* (Theze et al. 2011). It was observed that within *Diptera*, bracovirus from *Glossina* are more closely

related compared to those to *M. domestica*, suggesting that PDVs form evolutionary lineages.

Four maverick genes were identified in this study: DNA polymerase B2 involved in DNA excision repair and initiation of replication (DNA Pol B 2), the N terminal region of the parvovirus coat protein VP1 (Parvo coat N) that is important for virion retention and transduction by insect vectors, a retroviral-like integrase, and poxvirus A32 protein (Pox A32) that encodes an ATPase involved in virion DNA packaging. Although bracoviruses in wasps are co-opted to ensure their successful reproduction, their role in *Diptera* that do not share this mode of reproduction remains unclear. Protease gene was not found confirming that Dipteran Maverick lost the viral maturation ability but retained the transposition ability due to the presence of the integrase (Pritham et al. 2007). Maverick-associated protease genes, which play key roles in successful reproduction, are present in the bracoviral genomes of parasitoid wasps (Dupuy et al. 2011). *G. austeni* was shown to have the most abundant number of Maverick genes, and *G. brevipalpis* the least number. From the phylogenetic reconstruction of *Glossina* species, *Fusca* group (*G. brevipalpis*) formed the deepest branches, followed by *Morsitans* (*G. pallidipes* and *G. morsitans*) and *Palpalis* (*G. fuscipes*) groups (Chen, Li, and Aksoy 1999). *G. austeni* formed a separate fourth subgenus, *Machmadomyia*, which formed a sister-group relationship with the *Morsitans* group species (Chen et al. 1999). This was significant as it implied that *G. brevipalpis* is the least diverse species and has fewer genome rearrangement events.

Maverick genes were observed to branch into two distinct clades with notable differences in the branching times. This was significant as it concurred with previous findings that used different phylogenetic and taxonomic sampling methods to highlight the two types of Maverick sequences. The first type is thought to replicate in the mitochondria, and second in the cytoplasm (Mart Krupovic and Koonin 2015). Co-occurrence of two types of Maverick genes in each Dipteran genome was also showed. It was speculated that lineage specific duplications may have resulted in diverse sets of genes under different magnitudes of evolutionary pressures giving rise to the two clades. Previous findings showed that multiple copies of homologous PDVs have been identified in wasps at different loci of the insect genome (Bezier et al. 2013; Desjardins et al. 2008). This evolution of endogenous viral genes involves numerous rearrangements resulting from successive lineage-specific duplications, each creating genetic variants (Bezier et al. 2013).

Terminal inverted repeats (TIRs) were absent in the flanking regions of the Dipteran Mavericks. TIRs have also been shown to be absent in some parasitoid wasp bracoviral genomes, for example the *Cotesia congregata* bracoviral (CcBV) genome (Dupuy et al. 2011). This absence was ascribed to cumulative insertion and deletion events that blur genetic identification, and these events had minimal effects in the host's fitness (Langley and Charlesworth 1997). Transposable elements without TIRS are normally dysfunctional TIRs and are degraded (Dupuy et al. 2011). However, the retention of Dipteran Maverick genes suggested that they are under purifying selection, albeit the fact that their function was unknown (Desjardins et al. 2008). Due to lack of function this group of double-stranded DNA degrade over time and this makes it challenging to reconstruct macro-evolutionary history of viruses beyond 100 Mya (Theze et al. 2011).

The direction and magnitude of selection on the Maverick genes was also examined. Overall; these genes were under purifying selection, except pox A32, which was under positive selection. This was surprising because PDVs in *Diptera* appear to be non-functional viral fragments that are evolutionary dead ends and raised questions on the necessity to host's genome to retain them. Selection pressure of host genome sequences flanking these loci was examined to determine whether the bracoviral genes were inserted in conserved regions. It was observed that the flanking regions were relatively conserved; with the exception of the sequence upstream of parvo coat N, pox A32, and downstream of the integrase genes. Similarity in the magnitude and direction of selection pressure was significant as it suggested that bracoviral genes have lowered functional constraints to globally co-evolve with the host genome (Herniou et al. 2013).

Bracoviruses were also observed in different parts of the Africa (Kenya and Malawi) indicating that the genes are widely distributed and stably associated with *Diptera*. This was important as it confirmed that the presence of the virus in Dipteran genomes was not due to contamination of reference tsetse fly genome with a pathogenic virus or a single case of being parasitized by bracoviruses of wasp origin as earlier alleged. The genetic diversity of Mavericks from widely geographically distributed insect hosts was also assessed using phylogenetic analysis. It was observed that there is less diversity as corresponding sequences clustered together with the exception of pox A32 genes. This confirms that Mavericks are conserved and evolve together with the host genome.

6.0 CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

In conclusion, data generated in this study provides evidence of the presence of PDVs in tsetse fly genomes as well as in *M. domestica*. The study confirmed that PDVs are present in *Diptera* and share close homology. Phylogenetic reconstruction of the maverick genes showed that they can be grouped into two major groups each forming a distinct clade. Selection pressure analysis revealed that the integrase, parvo coat N and DNA pol B2 are under purifying selection while the pox A32 locus is under positive selection. The study also showed that the four loci were inserted at conserved regions and co-evolve with the host genome.

6.2 Recommendations

1. The PDVs identified in this study are non-functional although they were shown be under purifying selection. More elaborate studies should be carried out on the role of PDVs in tsetse fly genome to determine whether it can be exploited for vector control and gene therapy.
2. PDVs was observed in *Glossina* spp. and *M. domestica.* More studies should be conducted in other Dipteran genomes to determine the prevalence and evolution of PDVs genes across *Diptera*.
3. PDVs in the wasps has been shown to be assembled mainly from mobile elements, ancestral virus and the host DNA. Recombination events involving PDVs among Dipteran genomes should be investigated.

REFERENCES

Abd-Alla, Adly M. M. et al. 2013. 'Improving Sterile Insect Technique (SIT) for Tsetse Flies through Research on Their Symbionts and Pathogens.' *Journal of invertebrate pathology* 112(0):S2–10.

Aksoy, S. 2000. 'Tsetse--A Haven for Microorganisms.' *Parasitology today (Personal ed.)* 16(3):114–18.

Aksoy, Serap. 2003. 'Control of Tsetse Flies and Trypanosomes Using Molecular Genetics.' *Veterinary parasitology* 115(2):125–45.

Aksoy, Serap. 2010. 'A Proposal for Tsetse Fly (Glossina) Genome Projects.' (3):1–18.

Beer, Chantel J., Gert J. Venter, and Marc J. B. Vreysen. 2015. 'Determination of the Optimal Mating Age of Colonised Glossina Brevipalpis and Glossina Austeni Using Walk-in Field Cages in South Africa.' *Parasites & vectors* 8:467.

Benoit, Joshua B., Geoffrey M. Attardo, Aaron A. Baumann, Veronika Michalkova, and Serap Aksoy. 2015a. 'Adenotrophic Viviparity in Tsetse Flies : Potential for Population Control and as an Insect Model for Lactation.' *Annual review of entomology* 60:51–71.

Benoit, Joshua B., Geoffrey M. Attardo, Aaron A. Baumann, Veronika Michalkova, and Serap Aksoy. 2015b. 'Adenotrophic Viviparity in Tsetse Flies: Potential for Population Control and as an Insect Model for Lactation.' *Annual review of entomology* 60:351–71.

Bezier, Annie et al. 2009. 'Polydnaviruses of Braconid Wasps Derive from an Ancestral Nudivirus.' *Science (New York, N.Y.)* 323(5916):926–30.

Bezier, Annie et al. 2013. 'Functional Endogenous Viral Elements in the Genome of the Parasitoid Wasp Cotesia Congregata: Insights into the Evolutionary Dynamics of Bracoviruses.' *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1626):20130047.

Brightwell, R., R. D. Dransfield, and C. Kyorku. 1991. 'Development of a Low-Cost Tsetse Trap and Odour Baits for Glossina Pallidipes and G. Longipennis in Kenya.' *Medical and veterinary entomology* 5(2):153–64.

Brun, Reto, Johannes Blum, Francois Chappuis, and Christian Burri. 2010. 'Human African Trypanosomiasis.' *Lancet* 375(9709):148–59. Retrieved 22 July 2014.

Burke, Gaelen R., Sarah A. Thomas, Jai H. Eum, and Michael R. Strand. 2013. 'Mutualistic Polydnaviruses Share Essential Replication Gene Functions with Pathogenic Ancestors.' *PLoS pathogens* 9(5):e1003348.

Cassetti, M. C., M. Merchlinsky, E. J. Wolffe, a S. Weisberg, and B. Moss. 1998. 'DNA Packaging Mutant: Repression of the Vaccinia Virus A32 Gene Results in Noninfectious, DNA-Deficient, Spherical, Enveloped Particles.' *Journal of virology* 72(7):5769–80.

Chen, X., S. Li, and S. Aksoy. 1999. 'Concordant Evolution of a Symbiont with Its Host Insect Species: Molecular Phylogeny of Genus Glossina and Its Bacteriome-Associated Endosymbiont, Wigglesworthia Glossinidia.' *Journal of molecular evolution* 48(1):49–58.

Chevignon, Germain et al. 2014. ‘Functional Annotation of Cotesia Congregata Bracovirus: Identification of Viral Genes Expressed in Parasitized Host Immune Tissues.’ *Journal of virology* 88(16):8795–8812.

Crankshaft Publishing. n.d. ‘Tsetse Fly (Insects).’ Retrieved 29 June 2015 (http://what-when-how.com/insects/tsetse-fly-insects/).

Dereeper, A. et al. 2008. ‘Phylogeny.fr: Robust Phylogenetic Analysis for the Non-Specialist.’ *Nucleic acids research* 36(Web Server issue):W465–69.

Desjardins, Christopher A. et al. 2007. ‘Structure and Evolution of a Proviral Locus of Glyptapanteles Indiensis Bracovirus.’ *BMC microbiology* 7:61.

Desjardins, Christopher A. et al. 2008. ‘Comparative Genomics of Mutualistic Viruses of Glyptapanteles Parasitic Wasps.’ *Genome biology* 9(12):R183.

Drezen, J. M. et al. 2006. ‘The Few Virus-like Genes of Cotesia Congregata Bracovirus.’ *Archives of insect biochemistry and physiology* 61(3):110–22.

Dupuy, C. et al. 2011. ‘Transfer of a Chromosomal Maverick to Endogenous Bracovirus in a Parasitoid Wasp.’ *Genetica* 139(4):489–96.

Edgar, Robert C. 2004. ‘MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.’ *Nucleic Acids Research* 32(5):1792–97.

Elsen, P., M. A. Amoudi, and M. Leclercq. 1991. ‘[The discovery in Saudi Arabia of 2 species of tse-tse flies, vectors of human and animal trypanosomiasis].’ *Revue medicale de Liege* 46(4):225–31.

Espagne, Eric, Catherine Dupuy, Elisabeth Huguet, Laurence Cattolico, Bertille Provost, Nathalie Martins, Marylene Poirie, et al. 2004. ‘Genome Sequence of a Polydnavirus: Insights into Symbiotic Virus Evolution.’ *Science (New York, N.Y.)* 306(5694):286–89.

Espagne, Eric, Catherine Dupuy, Elisabeth Huguet, Laurence Cattolico, Bertille Provost, Nathalie Martins, Marylène Poirié, et al. 2004. ‘Genome Sequence of a Polydnavirus: Insights into Symbiotic Virus Evolution.’ *Science (New York, N.Y.)* 306(5694):286–89.

Espagne, Eric. 2014. ‘Evolution Genome Sequence of a Polydnavirus : Insights into Symbiotic Virus Evolution.’ 286(2004).

Finn, Robert D. et al. 2014. ‘Pfam: The Protein Families Database.’ *Nucleic acids research* 42(Database issue):D222–30.

Gooding, R. H. and E. S. Krafsur. 2005. ‘Tsetse Genetics: Contributions to Biology, Systematics, and Control of Tsetse Flies.’ *Annual review of entomology* 50:101–23.

Green, C. H. 1994. ‘Bait Methods for Tsetse Fly Control.’ *Advances in parasitology* 34:229–91.

Herniou, Elisabeth A. et al. 2013. ‘When Parasitic Wasps Hijacked Viruses: Genomic and Functional Evolution of Polydnaviruses.’ *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1626):20130051.

Hohn, Thomas. 2015. ‘RNA Based Viral Silencing Suppression in Plant Pararetroviruses.’ *Frontiers in plant science* 6:398.

International Glossina Genome Initiative. 2014. ‘Genome Sequence of the Tsetse Fly (Glossina Morsitans): Vector of African Trypanosomiasis.’ *Science (New York, N.Y.)* 344:380–86.

Jancek, Severine et al. 2013. ‘Adaptive Selection on Bracovirus Genomes Drives the Specialization of Cotesia Parasitoid Wasps.’ *PloS one* 8(5):e64432.

Kappmeier, K. and E. M. Nevill. 1999. ‘Evaluation of a Proposed Odour-Baited Target to Control the Tsetse Flies Glossina Brevipalpis and Glossina Austeni (Diptera: Glossinidae) in South Africa.’ *The Onderstepoort journal of veterinary research* 66(4):327–32.

Koonin, Eugene V, Valerian V Dolja, and Mart Krupovic. 2015. ‘Origins and Evolution of Viruses of Eukaryotes: The Ultimate Modularity.’ *Virology* 479–480:2–25.

Koonin, Eugene V, Mart Krupovic, and Natalya Yutin. 2015. ‘Evolution of Double-Stranded DNA Viruses of Eukaryotes: From Bacteriophages to Transposons to Giant Viruses.’ *Annals of the New York Academy of Sciences* 1341:10–24.

Korber, B., Rodrigo Allen G., and Learn Gerald H. 2000. ‘HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences.’ *Kluwer Academic Publishers* 55–72.

Krupovic, M. and E. V Koonin. 2015. ‘Polintons: A Hotbed of Eukaryotic Virus, Transposon and Plasmid Evolution.’ *Nat Rev Microbiol* 13(2):105–15.

Krupovic, Mart and Eugene V Koonin. 2015. ‘Polintons: A Hotbed of Eukaryotic Virus, Transposon and Plasmid Evolution.’ *Nature reviews. Microbiology* 13(2):105–15.

Langley, C. H. and B. Charlesworth. 1997. ‘Endogenous Proviruses as “Mementos”?’ *Nature* 388(6645):840.

Leak, S. G. A. 1998. *Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomosis.* CAB International, in association with the International Livestock Research Institute, Nairobi, Kenya. Retrieved 2 March 2015.

Louis, Faustine et al. 2013. ‘The Bracovirus Genome of the Parasitoid Wasp Cotesia Congregata Is Amplified within 13 Replication Units, Including Sequences Not Packaged in the Particles.’ *Journal of virology* 87(17):9649–60.

Maiti, Indu B. et al. 2003. ‘Antibiosis-Type Insect Resistance in Transgenic Plants Expressing a Teratocyte Secretory Protein (TSP14) Gene from a Hymenopteran Endoparasite (Microplitis Croceipes).’ *Plant biotechnology journal* 1(3):209–19.

Marchler-Bauer, Aron et al. 2011. ‘CDD: A Conserved Domain Database for the Functional Annotation of Proteins.’ *Nucleic acids research* 39(Database issue):D225–29.

Maudlin, I. 2006. ‘African Trypanosomiasis.’ *Annals of tropical medicine and parasitology* 100(8):679–701.

Misra, Sima et al. 2002. ‘Annotation of the Drosophila Melanogaster Euchromatic Genome: A Systematic Review.’ *Genome biology* 3(12):RESEARCH0083.

Mongin, Emmanuel, Christos Louis, Robert A. Holt, Ewan Birney, and Frank H. Collins. 2004. ‘The Anopheles Gambiae Genome: An Update.’ *Trends in parasitology* 20(2):49–52.

Pritham, Ellen J., Tasneem Putliwala, and Cédric Feschotte. 2007. ‘Mavericks, a Novel Class of Giant Transposable Elements Widespread in Eukaryotes and Related to DNA Viruses.’ *Gene* 390(1-2):3–17.

Roossinck, Marilyn J. 2011. ‘The Good Viruses: Viral Mutualistic Symbioses.’ *Nat Rev Micro* 9(2):99–108.

Roossinck, Marilyn J. 2015. ‘Move Over, Bacteria! Viruses Make Their Mark as Mutualistic Microbial Symbionts.’ *Journal of virology* 89(13):6532–35.

Spitzer, Austin L., Colin R. Parrish, and Ian H. Maxwell. 1997. ‘Tropic Determinant for Canine Parvovirus and Feline Panleukopenia Virus Functions through the Capsid Protein VP2.’ *Journal of General Virology* 78(4):925–28.

Strand, Michael R. and Gaelen R. Burke. 2012. ‘Polydnaviruses as Symbionts and Gene Delivery Systems.’ *PLoS pathogens* 8(7):e1002757.

Theze, Julien, Annie Bezier, Georges Periquet, Jean-Michel Drezen, and Elisabeth A. Herniou. 2011. ‘Paleozoic Origin of Insect Large dsDNA Viruses.’ *Proceedings of the National Academy of Sciences of the United States of America* 108(38):15931–35.

Vreysen, M. J. et al. 2000. ‘Glossina Austeni (Diptera: Glossinidae) Eradicated on the Island of Unguja, Zanzibar, Using the Sterile Insect Technique.’ *Journal of economic entomology* 93(1):123–35.

Wellde, B. T., D. A. Chumo, D. Waema, M. J. Reardon, and D. H. Smith. 1989. ‘A History of Sleeping Sickness in Kenya.’ *Annals of tropical medicine and parasitology* 83 Suppl 1:1–11.

Whitfield, James B. 2002. ‘Estimating the Age of the Polydnavirus/braconid Wasp Symbiosis.’ *Proceedings of the National Academy of Sciences of the United States of America* 99(11):7508–13.

World Health Organization. 2013. ‘Control and Surveillance of Human African Trypanosomiasis.’ *World Health Organization technical report series* (984):1–237.

Yang, Guangxiao, Geoffrey M. Attardo, Claudia Lohs, and Serap Aksoy. 2010. ‘Molecular Characterization of Two Novel Milk Proteins in the Tsetse Fly (Glossina Morsitans Morsitans).’ *Insect molecular biology* 19(2):253–62.

Zhang, Yufei, Jing Shi, and Shuying Liu. 2015. ‘Recent Advances in the Study of Active Endogenous Retrovirus Envelope Glycoproteins in the Mammalian Placenta.’ *Virologica Sinica* 30(4):239–48.

APPENDIX 1: BLASTP analysis of the 310 reported bracoviral sequences in *G. m. morsitans*

		Search BLASTP			
	G. m. morsitans	**NCBI acc. no.**	**E-value**	**Identity (%)**	**Description**
1	GMOY000028-PA	XP_011291826.1	0		PREDICTED: serine/threionine-protein kinase Genghis Khan [Musca domestica]
2	**GMOY000040-PA**	**YP_184883.1**	**1.00E-082**	**60**	**unnamed protein product [Cotesia congregata bracovirus]**
3	GMOY000102-PA	XP_005187509.1	0	72	PREDICTED: leucine-rich repeat neuronal protein_1 [Musca domestica]
4	GMOY000116-PA	XP_011878243.1	1.00E-043	50	PREDICTED: uncharacherized protein LOC105567743 [Vollenhovia emeryi]
5	GMOY000195-PA	XP_005175813.1	0	73	PREDICTED: putative serine pretease F56F10.a [Musca domestica]
6	GMOY000200-PA	XP_005175221.1	0	85	PREDICTED: probable ATP-dependent RNA helicase DDX47 [Musca domestica]
7	GMOY000220-PA	XP_011296230.1	0	73	PREDICTED: serine/threonine-protein kinase D3 [Musca domestica]
8	GMOY000243-PA	XP_001353143.1	0	59	GA19640 [Drosophila pseudoobscura pseudoobscura]
9	GMOY000268-PA	KKF20647.1	1.00E-063	98	Histone H4 [Larimichthys crocea]
10	GMOY000307-PA	XP_005176429.1	0	64	PREDICTED: uncharacherized protein LOC101900590 isoform

					X1 [Musca domestica]
11	GMOY000362-PA	XP_011294394.1	0	76	PREDICTED: serine/threionine-protein kinase GL21140 isoform X1 [Musca domestica]
12	GMOY000410-PA	XP_004519104.1	0	98	PREDICTED: serine/threonine-protein kinase tricomer [Ceratitis capitata]
13	GMOY000437-PA	XP_005187950.2	0	88	PREDICTED: rho-associated protein kinase 1 [Musca domestica]
14	GMOY000481-PA	XP_005183330.1	0	86	PREDICTED: probable ATP-dependent RNA helicase DDX46 [Musca domestica]
15	GMOY000482-PA	XP_011292439.1	0	56	PREDICTED:dystrophin-1 isoform X1 [Muscadomestica]
16	GMOY000488-PA	XP_011181033.1	0	63	PREDICTED: kinesin-like protein KIF16B [Bactrocera cucurbitae]
17	GMOY000502-PA	XP_005186272.1	0	86	PREDICTED: sodium-dependent neutral amin acid transporter [Musca domestica]
18	GMOY000525-PA	XP_005177622.1	2.00E-128	68	PREDICTED: trypsin eta [Musca domestica]
19	GMOY000526-PA	XP_011296597.1	3.00E-121	74	PREDICTED: ras-related protein Rab-274 [Musca domestica]
20	GMOY000556-PA	XP_005185071.1	0	58	PREDICTED: phosphatidylinositol phosphatase PTPRQ isoform X1 [Musca domestica]
21	GMOY000672-PA	XP_005186909.2	2.00E-063	49	PREDICTED: serine protease SP24D [Musca domestica]
22	GMOY000676-PA	XP_005189159.1	0	84	PREDICTED: sodium-dependent serotonin transporter [Musca

					domestica]
23	GMOY000684-PA	XP_005189502.1	2.00E-103	99	PREDICTED: protein mago nashi [Musca domestica]
24	GMOY000747-PA	XP_001981055.1	4.00E-108	88	GG11858[Drosophila erecta]
25	GMOY000748-PA	XP_011293272.1	9.00E-033	37	PREDICTED: uncharacterized protein LOC105261421 [Musca domestica]
26	GMOY000953-PA	XP_001959566.1	0	51	GF11982 [Drosophila ananassae]
27	GMOY001000-PA	XP_005180679.1	3.00E-127	99	PREDICTED: ADP-ribosylation factor-like protein 1 [Musca domestica]
28	GMOY001013-PA	XP_005175656.1	0	72	PREDICTED: protein 5NUC-like [Musca domestica]
29	GMOY001132-PA	XP_005181847.1	0	78	PREDICTED: calcium/calmodulin-dependent protein kinase type 1 [Musca domestica]
30	GMOY001164-PA	XP_00518027.1	5.00E-136	93	PREDICTED: ras-related protein Rab-2A [Musca domestica]
31	GMOY001193-PA	XP_005190818.1	0	87	PREDICTED: kinesin-like protein KIF3A [Musca domestica]
32	GMOY001293-PA	XP_011206365.1	0	72	PREDICTED: serine/threonine-protein kinase CST20 isoform X1 [Bactrocera dorsalis]
33	GMOY001348-PA	XP_005176344.1	1.00E-123	80	PREDICTED: ras-related protein Rab-21 [Musca domestica]
34	GMOY001362-PA	XP_005188545.1	0	75	PREDICTED: sodium-dependent nutrient amino acid transporter 1 [Musca domestica]
35	GMOY001368	XP_002063425.1	0	79	GK21901 [Drosophila

	-PA				wilistoni]
36	GMOY001560-PA	XP_011290547.1	0	77	PREDICTED: ATP-dependent RNA helicase DDX54 [Musca domestica]
37	GMOY001584-PA	XP_005188728.1	0	80	PREDICTED: SET and MYND domain-containing protein 4 [Musca domesitca]
38	GMOY001585-PA	XP_011213002.1	0	73	PREDICTED: probable trans-2-enoyl-CoA reductase [Bactrocera dorsalis]
39	GMOY001586-PA	XP_002018015.1	3.00E-101	44	GL16992 [Drosophila persimilis]
40	GMOY001589-PA	XP_005190041.1	0	88	PREDICTED: probable enoyl-CoA hydratase, mitochondrial [Musca domestica]
41	GMOY001677-PA	XP_005178683.1	0	88	PREDICTED: serine/threonine-protein kinase PAK 2 [Musca domestica]
42	GMOY001743-PA	XP_011203241.1	0	92	PREDICTED: protein ariadne-1 [Bactrocera dorsalis]
43	GMOY001858-PA	WP_010963123.1	0	89	DEAD/DEAH box helicase [Wolbachia endosymbiont of Drosophila melanogaster]
44	GMOY001867-PA	XP_005188959.1	1.00E-058	32	PREDICTED: LOW QUALITY PROTEIN; kelch-like protein 17 [Musca domestica]
45	GMOY001881-PA	XP_005189156.1	2.00E-022	82	PREDICTED: cytochrome b-c1 complex subunit 10 [Musca domestica]
46	GMOY001882-PA	XP_005189157.1	1.00E-124	96	PREDICTED: ras-related protein Rab-4B [Musca domestica]
47	GMOY001960-PA	XP_011292967.1	3.00E-086	67	PREDICTED: uncharacterized protein LOC105261872 [Musca

					domestica]
48	GMOY001983-PA	XP_005178934.1	9.00E-143	95	PREDICTED: ras-related protein Rab6 [Musca domestica]
49	GMOY002025-PA	XP_011188727.1	0	66	PREDICTED: tyrosine-protein kinase Fps85D isoform X3 [Bactrocera cucurbitae]
50	GMOY002040-PA	XP_005183399.1	0	69	PREDICTED: ATP-dependent RNA helicase dbp2 [Musca domestica]
51	GMOY002053-PA	XP_011188044.1	0	67	PREDICTED: dihydrolipoyllysine-residue succinyltransferase [Bactrocera cucurbitae]
52	GMOY002056-PA	XP_005189382.1	0	62	PREDICTED: chromosome-associated kinesin KIF4 isoform X1 [Musca domestica]
53	GMOY002090-PA	NP_995893.1	0	75	Par-1, isoform H [Drosophila melanogaster]
54	GMOY002098-PA	XP_005176315.1	0	97	PREDICTED: ATP-dependent RNA helicase WM6 [Musca domestica]
55	GMOY002130-PA	WP_012673123.1	0	98	MULTISPECIES: pyruvate dehydrogenase subunit beta [Wolbchia]
56	GMOY002187-PA	XP_011199820.1	9.00E-144	93	PREDICTED: ras-related protein Rab-30 [Bactrocera dorsalis]
57	GMOY002332-PA	XP_011295292.1	0	62	PREDICTED: chromosomal serine/threonine-protein kinase JIL-1 [Musca domestica]
58	GMOY002345-PA	XP_005183360.1	0	99	PREDICTED: calcium/calmodulin-dependent protein kinase [Musca domestica]
59	GMOY002370-PA	XP_005179374.1	1.00E-107	82	PREDICTED: ras-related protein Rab-18 [Musca domestica]

60	GMOY002373-PA	XP_005176928.1	0	87	PREDICTED: tyrosine-protein phosphatase non-receptor type 4 [Musca domestica]
61	GMOY002483-PA	IPR002110	0	75	PREDICTED: ankyrin repeat [Musca domestica]
62	GMOY002522-PA	XP_005177912.1	0	99	PREDICTED: eukaryotic initiation factor 4A-III [Musca domestica]
63	GMOY002529-PA	XP_005182548.2	0	47	PREDICTED: uncharacterized protein LOC101899800 [Musca domestica]
64	GMOY002598-PA	XP_005177401.1	5.00E-106	36	PREDICTED: probable cytochrome P450 4d14 [Musca domestica]
65	GMOY002745-PA	KKF20647.1	4.00E-065	99	Par-1, isoform H [Drosophila melanogaster]
66	GMOY002806-PA	XP_002049617.1	0	83	GJ21693 [Drosophila virilis]
67	GMOY002929-PA	XP_01118385.3	0	52	PREDICTED: uncharacherized protein LOC105213032 [Bactrocera cucurbitae]
68	GMOY002968-PA	XP_002049006.1	0	87	GJ1351 [Drosophila virilis]
69	GMOY002969-PA	XP_005191383.2	0	56	PREDICTED: atrial natriuretic peptide receptor 2-like [Musca domestica]
70	GMOY003090-PA	XP_005184452.1	0	91	PREDICTED: voltage-dependent anion-selective channel [Musca domestica]
71	GMOY003091-PA	XP_0051777781.1	0	75	PREDICTED: aurora kinase [Musca domestica]
72	GMOY003203-PA	XP_005190431.1	0	90	PREDICTED: ATP-dependent RNA helicase abstrakt [Musca domestica]
73	GMOY003204-PA	XP_005185265.1	0	68	PREIDICTED: probable cytochrome P450 6a18 [Musca domestica]
74	GMOY003214	XP_005184296.1	0	64	PREDICTED: apyrase-like

	-PA				[Musca domestica]
75	GMOY003219-PA	XP_005184676.1	0	77	PREDICTED: dnaJ protein homolog 1 [Musca domestica]
76	GMOY003287-PA	XP_005179420.1	0	98	PREDICTED: eukaryotic initiation factor 4A isoform X1 [Musca domestica]
77	GMOY003369-PA	XP_005175511.1	0	81	PREDICTED: lipoyltransferase 1, mitochondrial [Musca domestica]
78	GMOY003370-PA	XP_005190947.1	0	61	PREDICTED: tyrosine-protein kinase shark-like [Musca domestica]
79	GMOY003455-PA	XP_0051798117.2	0	69	PREDICTED: probable ATP-dependent RNA helicase DDX52 [Musca domestica]
80	GMOY003456-PA	XP_005179810.1	0	69	PREDICTED: protein halfway [Musca domestica]
81	GMOY003468-PA	XP_011290569.1	0	71	PREDICTED: protein notum homolog [Musca domestica]
82	GMOY003563-PA	XP_005183099.1	0	53	PREDICTED: kinesin-related protein 4 isoform X1 [Musca domestica]
83	GMOY003584-PA	XP_011294148.1	0	78	PREDICTED: serine/threonine-protein kinase N-like isoform X3 [Musca domestica]
84	GMOY003585-PA	XP_005184032.1	2.00E-132	92	PREDICTED: ras-related protein Rab-35 [Musca domestica]
85	GMOY003730-PA	XP_011296566.1	0	61	PREDICTED: probable cytochrome Ps3 [Musca domestica]
86	GMOY003850-PA	NP_476955.1	3.00E-130	100	ADP ribosylation factor at 79F, isoform B [Drosophila melanogaster]
87	GMOY003984-PA	XP_005191574.1	0	81	PREDICTED: tyrosine-protein kinase Src64B

					[Musca domestica]
88	GMOY003990-PA	XP_005183948.1	3.00E-158	47	PREDICTED: ADP-ribosylation factor-like protein 1 [Musca domestica]
89	GMOY004008-PA	XP_005177659.2	0	63	PREDICTED: phospholipase B1, membrane-associated-like [Musca domestica]
90	GMOY004136-PA	XP_011290238.1	0	77	PREDICTED: protein kinase C. eye isozyme [Musca domestica]
91	GMOY004141-PA	XP_005182735.1	3.00E-119	63	PREDICTED: RNA polymerase II degradation factor 1 [Musca domestica]
92	GMOY004171-PA	XP_005178339.1	4.00E-119	79	PREDICTED: ras-related Rab-11A [Musca domestica]
93	GMOY004213-PA	XP_005175397.2	0	85	PREDICTED: protein FAM188A [Musca domestica]
94	GMOY004220-PA	XP_005188409.1	0	76	PREDICTED: probable cation-transporting ATPase [Musca domestica]
95	GMOY004272-PA	XP_011290227.1	0	77	PREDICTED: LOW QUALITY PROTEIN: kinesin-like protein KIF13A [Musca domestica]
96	GMOY004318-PA	XP_005176971.1	0	81	PREDICTED: ATP-dependent RNA helicase p62 [Musca domestica]
97	GMOY004331-PA	XP_005176034.1	0	69	PREDICTED: probable cytochrome P450 4aa1 [Musca domestica]
98	GMOY004370-PA	XP_005186441.1	0	56	PREDICTED: kinesin-like protein subito [Musca domestica]
99	GMOY004375-PA	XP_005176932.1	0	95	PREDICTED: heat shock protein 83 [Musca domestica]
100	GMOY004382	XP_005189134.2	0	84	PREDICTED: sodium-dependent noradrenaline

	-PA				transport [Musca domestica]
111	GMOY004398-PA	XP_005177464.1	0	50	PREDICTED: IQ and AAA domain-containing protein 1-like [Musca domestica]
112	GMOY004429-PA	XP_005182853.1	0	84	PREDICTED: ATP-dependent RNA helicase p62-like [Musca domestica]
113	GMOY004495-PA	XP_005188319.1	5.00E-157	98	PREDICTED: ras-related protein Rab-14 [Musca domestica]
114	GMOY004498-PA	XP_005183608.1	0	87	PREDICTED: sodium-dependent neutral amin acid transporter B [Musca domestica]
115	GMOY004546-PA	XP_005187523.2	0	76	PREDICTED: ATP-dependent RNA helicase vasa, isoform A [Musca domestica]
116	GMOY004562-PA	XP_005177333.1	7.00E-110	64	PREDICTED: smad nuclear-interacting protein 1 [Musca domestica]
117	GMOY004571-PA	XP_005191858.1	5.00E-041	33	PREDICTED: tudor domain-containing protein 1 [Musca domestica
118	GMOY004574-PA	XP_005187238.1	0	76	PREDICTED: organic cation transporter protein [Musca domestica]
119	GMOY004621-PA	XP_011293663.1	6.00E-150	65	PREDICTED: 3'(2'),5'-bisphosphate nucleotidase 1 [Musca domestica]
120	GMOY004622-PA	XP_005186278.2	0	61	PREDICTED: protein claret segregational [Musca domestica]
121	GMOY004675-PA	XP_00517928.2	1.00E-138	75	PREDICTED: pyrimidodiazepine synthase-like [Musca domestica]
122	GMOY004705-PA	XP_005175110.1	0	71	PREDICTED: prostatic acid phosphatase [Musca domestica]

123	GMOY004740-PA	XP_005183707.1	4.00E-118	68	PREDICTED: dihydrolipoyllysine-residue succinyltransferase [Musca domestica]
124	GMOY004749-PA	XP_005185071.1	4.00E-123	96	PREDICTED: ADP-ribosylation factor-like protein 5B [Musca domestica]
125	GMOY004919-PA	XP_005185153.1	4.00E-075	70	PREDICTED: tyrosine-protein kinase Abl isoform X4 [Musca domestica]
126	**GMOY004962-PA**	**CBZ06032.1**	**8.00E-046**	**44**	**Maverick capsid-like p31.10 [Cotesia congregata bracovirus]**
127	**GMOY004963-PA**	**ACE75264.1**	**3.00E-071**	**47**	**DNA Pol B2 domain-containing protein [Glyptapanteles flavicoxis]**
128	GMOY005001-PA	XP_005180410.1	0	79	PREDICTED: ATP-dependent RNA helicase bel [Musca domestica]
129	GMOY005038-PA	XP_005190292.1	0	58	PREDICTED: serine/threonine-protein kinase PLK4 [Musca domestica]
130	GMOY005106-PA	XP_011293863.1	0	69	PREDICTED: organic cation transporter protein isoform X1 [Musca domestica]
131	GMOY005109-PA	XP_011293860.1	0	78	PREDICTED: organic cation transporter protein isoform X1 [Musca domestica]
132	GMOY005119-PA	XP_005188392.1	6.00E-178	50	PREDICTED: kinesin heavy chain [Musca domestica]
133	GMOY005120-PA	XP_005188392.1	0	51	PREDICTED: kinesin heavy chain [Musca domestica]
134	GMOY005121-PA	XP_005181676.1	0	79	PREDICTED: TBC1 domain family member 23 [Musca domestica]

135	GMOY005309-PA	XP_011295954.1	4.00E-116	69	PREDICTED: serine protease SP24D [Musca domestica]
136	GMOY005418-PA	XP_011206339.1	0	70	PREDICTED: tyrosine-protein phosphatase non-receptor type 9 [Bactrocera dorsalis]
137	GMOY005601-PA	XP_011291039.1	0		PREDICTED: serine/threonine-protein kinase par-1 [Musca domestica]
138	GMOY005644-PA	XP_005176364.1	0	75	PREDICTED: probable 2-oxoglutarate dehydrogenase E1 [Musca domestica]
139	GMOY005651-PA	XP_005187238.1	2.00E-055	32	PREDICTED: organic cation transporter protein [Musca domestica]
140	GMOY005704-PA	XP_011292927.1	0	57	PREDICTED: uncharacterized protein LOC101890374 [Musca domestica]
141	GMOY005802-PA	XP_005186507.1	0	87	PREDICTED: RAC serine/threonine-protein kinase [Musca domestica]
142	GMOY005831-PA	XP_005184902.1	0	59	PREDICTED: ubiquitin-associated domain-containing protein 1 [Musca domestica]
143	GMOY005834-PA	XP_005184676.1	1.00E-180	73	PREDICTED: dnaJ protein homolog 1 [Musca domestica]
144	GMOY005873-PA	XP_005177707.1	0	94	PREDICTED: protein ariadne-2 [Musca domestica]
145	GMOY005957-PA	XP_002002880.1	0	57	GI10628 [Drosophila mojavensis]
146	GMOY005961-PA	XP_005191779.2	0	93	PREDICTED:putative ATP-dependent RNA helicase me31b [Musca domestica]
147	GMOY005984-PA	XP_005185980.1	1.00E-104	57	PREDICTED: uncharacterized protein LOC101895261 [Musca

					domestica]
148	GMOY006012-PA	XP_005177036.1	0	75	PREDICTED: maestro heat-like repeat-containing protein family [Musca domestica]
149	GMOY006210-PA	XP_005177669.1	3.00E-156	54	PREDICTED: phosphalipase B1, membrane-associated-like [Musca domestica]
150	GMOY006211-PA	XP_011296455.1	1.00E-178	53	PREDICTED: probable cytochrome P450 4d14 [Musca domestica]
151	GMOY006212-PA	XP_005177308.1	0	56	PREDICTED: cytochrome P450 4d2-like [Musca domestica]
152	GMOY006213-PA	XP_011296455.1	3.00E-167	54	PREDICTED: probable cytochrome P450 4d14 [Musca domestica]
153	GMOY006214-PA	XP_005177309.1	0	71	PREDICTED: cytochrome P450 4d2-like isoform X1 [Musca domestica]
154	GMOY006272-PA	XP_005184382.2	7.00E-010	53	PREDICTED: uncharacterized SDCCAG3 family protein [Musca domestica]
155	GMOY006273-PA	XP_005184372.1	7.00E-086	85	PREDICTED: mediator of RNA polymerase II transcription [Musca domestica]
156	GMOY006293-PA	XP_005190871.1	0	95	PREDICTED: tyrosine-protein phosphatase 99A isoform X1 [Musca domestica]
157	GMOY006372-PA	XP_005180198.1	0	78	PREDICTED: endoplasmin [Musca domestica]
158	GMOY006407-PA	XP_005190739.1	2.00E-093	73	PREDICTED: cAMP-dependent protein kinase catalytic [Musca domestica]
159	GMOY006494-PA	XP_005191902.2	4.00E-113	51	PREDICTED: kinesin -like protein KIF14 [Musca domestica]

160	GMOY006501-PA	XP_005192107.2	0	96	PREDICTED: cGMP-dependent protein kinase [Bactrocera cucurbitae]
161	GMOY006554-PA	XP_005188211.1	0	77	PREDICTED: probable ATP-dependent RNA helicase DDX55 [Musca domestica]
162	GMOY006555-PA	XP_005176317.1	0	63	PREDICTED: ovarian-specific serine/threonine-protein kinase Lok [Musca domestica]
163	GMOY006609-PA	XP_005183208.1	6.00E-150	90	PREDICTED: GYF domain-containing protein C18H9.3 [Musca domestica]
164	GMOY006624-PA	XP_005174859.1	0	73	PREDICTED: tyrosone-protein kinase transmembrane [Musca domestica]
165	GMOY006625-PA	XP_00517486.3	3.00E-075	76	PREDICTED: apolipoprotein D isoform X2 [Musca domestica]
166	GMOY006686-PA	XP_005187702.1	1.00E-081	52	PREDICTED: trypsin alpha-4-like [Musca domestica]
167	GMOY006724-PA	XP_005184536.1	0	82	PREDICTED: tyrosine-protein kinase Drl [Musca domestica]
168	GMOY006725-PA	XP_005178583.1	0	55	PREDICTED: insulin-like receptor isoform [Musca domestica]
169	GMOY006735-PA	XP_011290279.1	0	95	PREDICTED: tyrosine-protein phosphatase Lar [Musca domestica]
170	GMOY006831-PA	XP_005174763.1	0	83	PREDICTED: epidermal growth factor receptor isoform X1 [Musca domestica]
171	GMOY006880-PA	XP_005183236.1	1.00E-151	65	PREDICTED: nicotinamide riboside kinase 1 [Musca domestica]
172	GMOY006952	XP_005184740.1	0	73	PREDICTED: organic cation transporter-like

	-PA				protein [Musca domestica]
173	GMOY006960-PA	XP_005180743.1	0	83	PREDICTED: venom acid phosphatase Acph-1 [Musca domestica]
174	GMOY007054-PA	XP_005190125.2	0	54	PREDICTED: kinesin-like protein KIF18A [Musca domestica]
175	GMOY007081-PA	XP_005175083.2	0	84	PREDICTED: kinesin-like protein Klp10A [Musca domestica]
176	GMOY007115-PA	XP_005177457.1	0	96	PREDICTED: uncharacterized protein LOC101891623 isoform X2 [Musca domestica]
177	GMOY007161-PA	XP_005176519.1	0	70	PREDICTED: SET and MYND domain-containing protein 4 [Musca domestica]
178	GMOY007294-PA	XP_011296323.1	0	87	PREDICTED: 2-oxoglutarate dehydrogenase, mitochondrial [Musca domestica]
179	GMOY007296-PA	XP_005189081.1	0	72	PREDICTED: 2-oxoglutarate dehydrogenase, mitochondrial [Musca domestica]
180	GMOY007343-PA	XP_005179484.1	2.00E-107	62	PREDICTED: chymotrypsin-2-like [Musca domestica]
181	GMOY007350-PA	XP_005186282.1	2.00E-056	46	PREDICTED: protein Tube [Musca domestica]
182	GMOY007366-PA	XP_005182011.1	0	83	PREDICTED: ribosoma protein S6 kinase 2 beta [Musca domestica]
183	GMOY007447-PA	XP_005181003.1	0	74	PREDICTED: sodium-dependent nutrient amino acid [Musca domestica]
184	GMOY007473-PA	XP_007023326.1	0.69	27	Tetratricopeptide repeat-like superfamily protein, putative [Theobroma cacao]

185	GMOY007474-PA	XP_005189985.2	0	63	PREIDICTED: cGMP-dependent protein kinase, isozyme 1 [Musca domestica]
186	GMOY007493-PA	KKF20647.1	4.00E-065	99	Histone H4 [Larimichthys crocea]
187	GMOY007518-PA	XP_005185716.1	0	64	PREDICTED: probable ATP-dependent RNA helicase DDX43 [Musca domestica]
188	GMOY007587-PA	XP_005176403.1	0	77	PREDICTED: probable cytochrome P450 6v1 [Musca domestica]
189	GMOY007678-PA	XP_005186030.1	0	79	PREDICTED: cytochrome P450 4c3 [Musca domestica]
190	GMOY007684-PA	XP_005186030.1	0	79	PREDICTED: cytochrome P450 4c3 [Musca domestica]
191	GMOY007696-PA	XP_005183701.1	1.00E-153	97	PREDICTED: ras-related protein Rab-13 [Musca domestica]
192	GMOY007775-PA	XP_005188525.2	0	57	PREDICTED: bipolar kinesin KRP-130 [Musca domestica]
193	GMOY007810-PA	XP_005190125.2	0	58	PREDICTED: kinesin-like protein KIF18A [Musca domestica]
194	GMOY007838-PA	XP_005179585.1	7.00E-168	44	PREDICTED: mucin-5AC [Musca domestica]
195	GMOY007841-PA	XP_005186298.1	0	63	PREDICTED: protein claret segregational [Musca domestica]
196	GMOY007843-PA	XP_005191966.1	0	53	PREDICTED: cystein-rich protein 2-binding protein [Musca domestica]
197	GMOY007844-PA	XP_005191965.1	0	82	PREDICTED: probable ATP-dependent RNA helicase DDX23 [Musca domestica]
198	GMOY007851	XP_011291891.1	0	90	PREDICTED: dnaJ protein homolog 1 isoform [Musca

	-PA				domestica]
199	GMOY007863-PA	XP_005182217.1	1.00E-097	48	PREDICTED: probable cytochrome P450 6u1 [Musca domestica]
200	GMOY007864-PA	XP_005182216.1	6.00E-046	96	PREDICTED: probable protein BRICK1-B [Musca domestica]
201	GMOY007894-PA	XP_005189076.1	0	72	PREDICTED: tyrosine-protein phosphatase 69D [Musca domestica]
202	GMOY007904-PA	XP_011294203.1	0	42	PREDICTED: LOW QUALITY PROTEIN: protein sevenless [Musca domestica]
203	GMOY007954-PA	XP_005176097.1	6.00E-149	79	PREDICTED: phosphoglycerate mutase 2 [Musca domestica]
204	GMOY007955-PA	XP_005176095.1	0	62	PREDICTED: eukaryotic initiation factor 2D [Musca domestica]
205	GMOY007978-PA	XP_012285884.1	1.00E-080	45	PREDICTED: uncharacherized protein LOC105702699 [Orussus abietinus]
206	GMOY007991-PA	XP_005188894.1	0	84	PREDICTED: protein ariadne-1 [Musca domestica]
207	GMOY008073-PA	XP_005184312.1	1.00E-167	50	PREDICTED: probable cytochrome P450 28a5 [Musca domestica]
208	GMOY008074-PA	XP_011291227.1	0	65	PREDICTED: probable cytochrome P450 28c1 [Musca domestica]
209	GMOY008774-PA	XP_005182432.1	0	77	PREDICTED: probable ATP-dependent RNA helicase Dbp45A [Musca domestica]
210	GMOY008348-PA	XP_0051178029.1	0	80	PREDICTED: kinesin-like protein KIF19 [Musca domestica]
211	GMOY008404	XP_005186952.1	0	66	PREDICTED: EGF domain-specific O-inked N-

	-PA				acetyglucosamine [Musca domestica]
212	GMOY008434-PA	XP_005174963.1	0	78	PREDICTED: serine/threionine-protein kinase ULK3-like [Musca domestica]
213	GMOY008560-PA	XP_005188392.1	7.00E-099	35	PREDICTED: kinesin heavy chain [Musca domestica]
214	GMOY008619-PA	XP_005182835.1	2.00E-144	98	PREDICTED: ras-related protein Rab-10 [Musca domestica]
215	GMOY008630-PA	XP_005188709.1	0	87	PREDICTED: uncharacterized protein LOC101895796 isoform X1 [Musca domestica]
216	GMOY008680-PA	XP_005176508.1	3.00E-046	47	PREDICTED: WASH complex subunit FAM21 homolog isoform X2 [Musca domestica]
217	GMOY008681-PA	XP_005177950.1	5.00E-058	57	PREDICTED: PIN2/TERF1-interacting telomerase inhibitor 1 [Musca domestica]
218	GMOY008746-PA	XP_011858001.1	3.00E-084	45	PREDICTED: uncharacterized protein LOC105556512 [Vollenhovia emeryi]
219	GMOY008760-PA	XP_005185353.2	0	59	PREDICTED: tyrosine kinase receptor Cad96Ca [Musca domestica]
220	GMOY008774-PA	XP_005176748.1	8.00E-173	76	PREDICTED: serine/threonine-protein kinase polo [Musca domestica]
221	GMOY008795-PA	XP_005181558.1	5.00E-147	98	PREDICTED: ras-related protein Rab-1A [Musca domestica]
222	GMOY008814-PA	XP_005186842.1	0	82	PREDICTED: manganese-transporting ATPase 13A1 [Musca domestica]
223	GMOY008941	XP_005190740.1	1.00E	78	PREDICTED: cAMP-dependent protein kinase

	-PA		-092		catalytic subunit [Musca domestica]
224	GMOY008985-PA	XP_005179314.1	2.00E-159	99	PREDICTED: ras-related protein Rab-3 [Musca domestica]
225	**GMOY009081-PA**	**YP_184883.1**	**2.00E-026**	**52**	**unnamed protein product [Cotesia congregata bracovirus]**
226	**GMOY009082-PA**	**CBZ0632.1**	**2.00E-045**	**54**	**Maverick capsid-like p31.10 [Cotesia congregata bracovirus]**
227	GMOY009088-PA	XP_005188891.1	0	86	PREDICTED: receptor-type tyrosine-protein phosphatase kappa [Musca domestica]
228	GMOY009095-PA	XP_011292382.1	0	78	PREDICTED: protein 4.1 homolog [Musca domestica]
229	GMOY009192-PA	XP_011295546.1	0	64	PREDICTED: PI-PLC X domain-containing protein 1 isoform X2 [Musca domestica]
230	GMOY009213-PA	XP_005183225.1	0	86	PREDICTED: tyrosine-protein phosphatase [Musca domestica]
231	GMOY009326-PA	XP_005187098.1	0	71	PREDICTED: sodium-dependent nutrient amino acid transporter [Musca domestica]
232	GMOY009343-PA	XP_011291203.1	0	83	PREDICTED: uncharacterized protein PB18E9.04c [Musca domestica]
233	GMOY009377-PA	XP_011291997.1	0	70	PREDICTED: uncharacterized protein LOC101888476 [Musca domestica]
234	GMOY009378-PA	XP_005180107.1	0	54	PREDICTED: probable cytochrome P450 9f2 [Musca domestica]
235	GMOY009386-PA	XP_011294575.1	0	69	PREDICTED: sodium-dependent nutrient amino acid transporter 1 [Musca

					domestica]
236	GMOY009388-PA	XP_011294575.1	0	76	PREDICTED: sodium-dependent nutrient amino acid transporter 1 [Musca domestica]
237	GMOY009612-PA	XP_005180450.1	3.00E-178	67	PREDICTED: ras-related protein Rab-26-like [Musca domestica]
238	GMOY009698-PA	XP_005189172.1	0	73	PREDICTED: ATP-dependent RNA helicase DBP2 [Musca domestica]
239	GMOY009756-PA	XP_005191789.2	0	90	PREDICTED: uncharacterized protein LOC101899724 [Musca domestica]
240	GMOY009767-PA	XP_011296124.1	0	52	PREDICTED: probable cytochrome P450 28a5 [Musca domestica]
241	GMOY009835-PA	XP_011293343.1	0	62	PREDICTED: uncharacterized protein LOC101890546 isoform X1 [Musca domestica]
242	GMOY009836-PA	XP_011291067.1	0	94	PREDICTED: 5'-AMP-activated protein kinase catalytic subunit [Musca domestica]
243	GMOY009864-PA	XP_011295758.1	0	64	PREDICTED: mucin-12-like [Musca domestica]
244	GMOY009933-PA	XP_011202828.1	0	77	PREDICTED: probable ATP-dependent RNA helicase pitchoune [Bactrocera cucurbitae]
245	GMOY009965-PA	XP_0051849291.	6.00E-137	88	PREDICTED: trafficking protein particle complex subunit 4 [Musca domestica]
246	GMOY009966-PA	XP_005184928.1	0	82	PREDICTED: kinesin-like protein Klp68D [Musca domestica]
247	GMOY009985-PA	XP_001632579.1	4.00E-116	86	predicted protein [Nematostella vectensis]

248	GMOY009988-PA	XP_005178336.1	5.00E-153	99	PREDICTED: ras-related protein Rab-aaA [Musca domestica]
249	GMOY010012-PA	XP_005182162.2	0	81	PREDICTED: sodium-and chloride-dependent GABA transporter [Musca domestica]
250	GMOY010028-PA	XP_005175113.1	3.00E-172	59	PREDICTED: phospholipase B1, membrane-associated-like [Musca domestica]
251	**GMOY010042-PA**	**ACE75264.1**	**1.00E-033**	**46**	**DNA Pol B2 domain-containing protein [Glyptapanteles flavicoxis]**
252	**GMOY010043-PA**	**CBZ06032.1**	**2.00E-043**	**53**	**Maverick capsid-like p31.10 [Cotesia congregata bracovirus]**
253	GMOY010066-PA	XP_005181629.1	0	71	PREDICTED: sodium-dependent nutrient amino acid transporter 1 [Musca domestica]
254	GMOY010116-PA	XP_002047905.1	0	73	GJ13697 [Drosophila virilis]
255	GMOY010132-PA	NP_001273811.1	0	63	cytochrome P450 6g1-like [Musca domestica]
256	GMOY010133-PA	XP_005187941.1	0	72	PREDICTED: probable cytochrome P450 6g2 [Musca domestica]
257	GMOY010134-PA	XP_005188729.1	1.00E-169	50	PREDICTED: probable cytochrome P450 6t3 [Musca domestica]
258	GMOY010135-PA	XP_005187941.1	0	65	PREDICTED: probable cytochrome P450 6g2 [Musca domestica]
260	GMOY010137-PA	XP_005187941.1	0	66	PREDICTED: probable cytochrome P450 6g2 [Musca domestica]
261	GMOY010221-PA	XP_005185133.1	0	59	PREDICTED: cGMP-dependent protein kinase 1 isoform X2 [Musca domestica]

262	GMOY010239-PA	XP_005188458.1	0	67	PREDICTED: uncharacterized protein LOC101897593 [Musca domestica]
263	GMOY010311-PA	XP_011296455.1	6.00E-177	52	PREDICTED: probable cytochrome P450 4d14 [Musca domestica]
264	GMOY010312-PA	XP_005177305.1	1.00E-066	71	PREDICTED: ADP-ribosylation factor 1-like [Musca domestica]
265	GMOY010324-PA	XP_005176197.1	0	51	PREDICTED: pre-mRNA-processing ATP-dependent RNA [Musca domestica]
266	GMOY010335-PA	XP_011291860.1	2.00E-118	60	PREDICTED: kinesin-like protein KIF14 [Musca domestica]
267	GMOY010399-PA	XP_011290537.1	0	60	PREDICTED: citron Rho-interacting kinase [Musca domestica]
268	GMOY010400-PA	XP_005183363.2	7.00E-092	76	PREDICTED: kxDL motif-containing protein CG10681 [Musca domestica]
269	GMOY010462-PA	XP_005189100.1	0	81	PREDICTED: ATP-dependent RNA helicase DDX42 [Musca domestica]
270	GMOY010635-PA	ABV48810.1	0	61	cytochrome P450 CYP6A5v2 [Musca domestica]
271	GMOY010636-PA	XP_005184401.1	0	62	PREDICTED: probable cytochrome P450 6a13 [Musca domestica]
272	GMOY010807-PA	XP_005188392.1	0	92	PREDICTED: kinesin heavy chain [Musca domestica]
273	GMOY010869-PA	XP_0052004446.1	2.00E-115	99	GI19608 [Drosophila mojavensis]
274	GMOY010893-PA	XP_0051818173.1	2.00E-174	51	PREDICTED: glomulin [Musca domestica]
275	GMOY010928-PA	XP_005188894.1	8.00E-176	57	PREDICTED: protein ariadne-1 [Musca

					domestica]
276	GMOY010979-PA	XP_005185070.1	1.00E-133	95	PREDICTED: ras-related protein Rab-5B [Musca domestica]
277	GMOY011001-PA	XP_005185960.1	8.00E-165	76	PREDICTED: sodium- and chloride-dependent glycine [Musca domestica]
278	GMOY011073-PA	XP_005177246.1	0	64	PREDICTED: peroxisome biogenesis protein 1 [Musca domestica]
279	GMOY011112-PA	XP_005186810.1	0	81	PREDICTED: organic cation transporter protein [Musca domestica]
280	GMOY011147-PA	XP_005186186.1	8.00E-092	43	PREDICTED: myosin-G heavy chain isoform X1 [Musca domestica]
281	GMOY011148-PA	XP_005186190.1	0	88	PREDICTED: serine/threonine-protein kinase Aurora-2 [Musca domestica]
282	GMOY011163-PA	XP_005178111.1	0	79	PREDICTED: IQ and AAA domain-containing protein 1 [Musca domestica]
283	GMOY011214-PA	XP_005191193.1	0	69	PREDICTED: prostatic acid phosphatase isoform X1 [Musca domestica]
284	GMOY011222-PA	XP_0051989699.1	2.00E-176	93	PREDICTED: 40S ribosomal protein S4 [Musca domestica]
285	GMOY011233-PA	XP_005182477.2	4.00E-147	98	PREDICTED: ras-related protein Rab-8A [Musca domestica]
286	**GMOY011252-PA**	**YP_184883.1**	**2.00E-073**	**66**	**unnamed protein product [Cotesia congregata bracovirus]**
287	GMOY011294-PA	XP_005176748.1	0	89	PREDICTED: serine/threonine-protein kinase polo [Musca domestica]
288	GMOY011352-PA	XP_005184170.2	0	84	PREDICTED: ephrin type-B receptor 1-B [Musca

					domestica]
289	GMOY011363-PA	XP_005182182.1	2.00E-061	56	PREDICTED: protein RRNAD1 [Musca domestica]
290	GMOY011364-PA	XP_005182179.1	0	74	PREDICTED: zinc carboxypeptidase [Musca domestica]
291	GMOY011396-PA	XP_005177233.1	7.00E-061	31	PREDICTED: charged multivesicular body protein 7 [Musca domestica]
292	GMOY011512-PA	XP_005176317.1	0	67	PREDICTED: ovarian-specific serine/threonine-protein kinase Lok [Musca domestica]
293	GMOY011532-PA	XP_011293196.1	8.00E-132	68	PREDICTED: phosphoglycerate mutase 2 [Musca domestica]
294	GMOY011690-PA	XP_005183854.1	0	73	PREDICTED: uncharacterized protein LOC101888768 isoform X1 [Musca domestica]
295	GMOY011723-PA	XP_012234650.1	1.00E-066	41	PREDICTED: uncharacterized protein LOC105679297 [Linepithema humile]
296	GMOY011744-PA	XP_001966647.1	9.00E-119	91	GF23416 [Drosophila ananassae]
297	GMOY011774-PA	XP_008549827.1	2.00E-036	47	PREDICTED: uncharacterized protein LOC103572821 [Microplitis demolitor]
298	GMOY011786-PA	XP_005190543.1	3.00E-140	93	PREDICTED: ras-related protein Rab6-like [Musca domestica]
299	GMOY011961-PA	XP_005177701.1	0	68	PREDICTED: ATP-dependent DNA helicase Q4 [Musca domestica]
300	GMOY011962-PA	XP_005177685.1	0	83	PREDICTED: glycylpeptide N-tetadecanoyltransferase [Musca domestica]

301	GMOY012004-PA	XP_011555639.1	0	35	PREDICTED: uncharacterized protein LOC105386719 [Plutella xylostella]
302	GMOY012057-PA	XP_005187723.1	0	73	PREDICTED: organic cation transporter protein [Musca domestica]
303	GMOY012098-PA	XP_005181928.1	0	90	PREDICTED: G protein-coupled receptor kinase 1 [Musca domestica]
304	GMOY012121-PA	XP_005184742.1	0	85	PREDICTED : organic cation transporter protein [Musca domestica]
305	GMOY012151-PA	XP_011293368.1	0	80	PREDICTED: G protein-coupled receptor kinase 2 [Musca domestica]
306	GMOY012181-PA	XP_005185596.1	0	91	PREDICTED: protein kinase C isoform X2 [Musca domestica]
307	GMOY012312-PA	XP_011292815.1	2.00E-125	58	PREDICTED: apyrase-like isoform X1 [Musca domestica]
308	GMOY012313-PA	XP_011292817.1	0	50	PREDICTED: apyrase-like isoform X3 [Musca domestica]
309	GMOY012343-PA	XP_005188432.1	0	76	PREDICTED: solute carrier family 22 member 13-like [Musca domestica]
310	GMOY012367-PA	XP_005187391.1	0	71	PREDICTED: organic cation transporter protein isoform X1 [Musca domestica]