Hytrosavirus in tsetse flies:
Phylogeography and molecular mode of action

Irene K. Meki

Hytrosavirus in tsetse flies:

Phylogeography and molecular mode of action

Irene Kasindi Meki

Thesis committee

Promotors
Prof. Dr M.M. van Oers
Professor of Virology
Wageningen University & Research

Prof. Dr A.A.M. Abd-Alla
Molecular Biologist and Virologist
FAO/International Atomic Energy Agency, Insect Pest Control Laboratory
Seibersdorf, Vienna, Austria

Co-promotors
Prof. Dr J.M. Vlak
Emeritus professor of Virology
Wageningen University & Research

Dr H. M. Kariithi
Molecular Virologist, Kenya Agricultural and Livestock Research Organization, Nairobi, Kenya

Other members
Prof. Dr W. Takken, Wageningen University & Research
Prof. Dr J. Van Den Abbeele, Institute of Tropical Medicine, Antwerp, Belgium
Prof. Dr E. Schnettler, Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Germany
Dr M.P. Zwart, Netherlands Institute of Ecology, Wageningen

This research was conducted under the auspices of the Graduate School of Production Ecology and Resource Conservation.

Hytrosavirus in tsetse flies:

Phylogeography and molecular mode of action

Irene Kasindi Meki

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday October 8, 2018
at 4 p.m. in the Aula.

Irene Kasindi Meki
Hytrosavirus in tsetse flies: Phylogeography and molecular mode of action,
246 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, summary in English

ISBN: 978-94-6343-314-3

DOI: 10.18174/454755

Table of contents

Chapter 1

General Introduction

Tsetse flies and African trypanosomosis

Tsetse flies (*Glossina* spp.) are the vectors of *Trypanosoma* parasites that cause two major diseases in tropical sub-Saharan Africa, the human African trypanosomosis (HAT or sleeping sickness) and the animal African trypanosomosis (AAT or nagana). There are 31 tsetse fly species and sub-species in the genus *Glossina* of the family Glossinidae, but only 8-10 of these are of medical and economic importance (Krafsur, 2009). *Glossina* species are distributed over 37 sub-Saharan African countries (**Figure 1A**) although two species (*G. fuscipes fuscipes* and *G. morsitans submorsitans*) have also been reported from the Southwest of the Arabian Peninsula (Cecchi *et al.,* 2008; Elsen *et al.,* 1991). Unlike other dipteran vectors such as mosquitoes, both male and female tsetse flies are obligatory bloodsuckers and hence serve as trypanosome vectors to uninfected mammalian host during feeding (Peacock *et al.,* 2012). The male flies of various tsetse species have been reported to have higher rates of trypanosome infection compared to the females (Peacock *et al.,* 2012; Welburn and Maudlin, 1999).

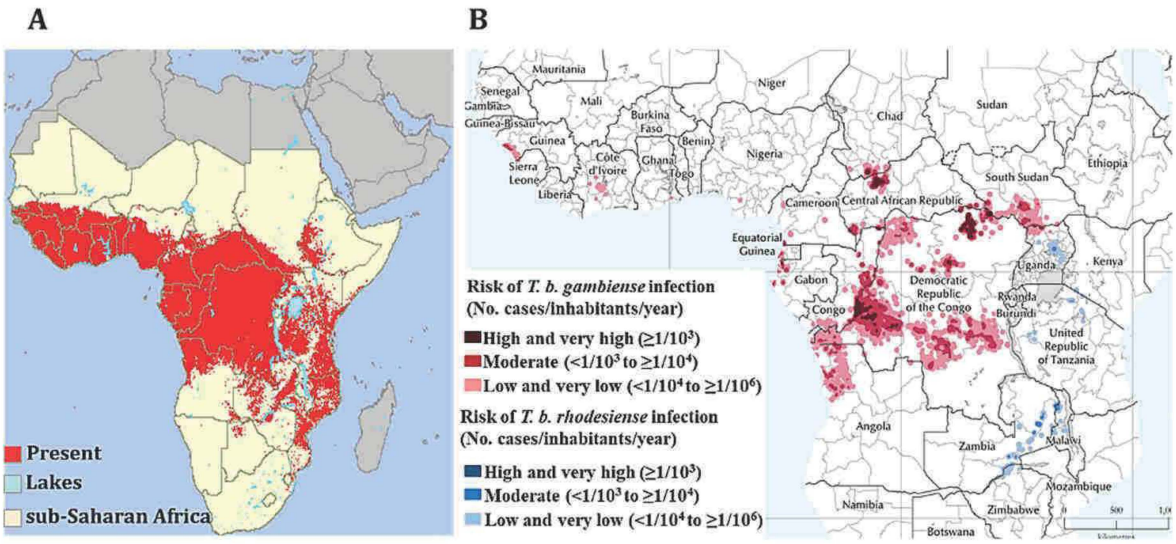


Figure 1: Distribution of tsetse flies and Trypanosomosis in Africa. (A) The predicted areas of suitability for tsetse fly species in sub-Saharan Africa. The color represents the regions where tsetse flies are found in sub-Saharan African countries in cream (Map courtesy of FAO). (B) Monitoring of human African trypanosomosis. The regions at risk of *T. b. gambiense* infections are shown in red while those at risk of *T. b. rhodesiense* infections are in blue. The color intensity of the two forms of the disease increase with the number of cases reported per inhabitant per year (Franco *et al.,* 2017).

The genus *Glossina* is divided into three distinct taxonomic sub-groups (i.e. the Palpalis, Morsitans, and Fusca group) based on morphological characteristics, such as the external genitalia of males, habitats and host preferences (Vreysen *et al.,* 2013). In addition to these morphological features, several genetic tools are under development for better distinction of the tsetse species (Chen *et al.,* 1999; Dyer *et al.,* 2008). The species belonging to the Palpalis group are the vectors of AAT in West Africa (*G. tachinoides*, *G. palpalis palpalis* and *G. palpalis gambiensis*) and HAT in central Africa (*G. fuscipes fuscipes* and *G. fuscipes quanzensis*). In the Morsitans group, *G. morsitans* spp. and *G. pallidipes* are the major vectors of AAT and HAT in eastern and southern Africa (Van den Bossche *et al.,* 2010). The species in the Fusca group are of less economic importance since they inhabit lowland rain forests and isolated forests in West Africa and they are not attracted by human. However, *G. brevipalpis* is of importance in East, Central and Southern Africa since it occurs in places where this species is in contact with livestock or man (Krafsur, 2009).

The *Trypanosoma brucei gambiens*e parasite causes the chronic form of HAT in West and Central Africa, while *T. b. rhodesiense* causes the acute form of the disease in East and Southern Africa. Uganda is the only country in sub-Saharan Africa at risk of both forms of HAT disease (**Figure 1B**) (Franco *et al.,* 2017; Picozzi *et al.,* 2005). The *T. congolense*, *T. vivax* and *T. b. brucei* cause AAT, which is mild in wild animals but severe or fatal in domestic animals (Steverding, 2008). While AAT is widespread in all tsetse-infested areas, HAT is present mainly in rural and remote foci. The latter is probably due to the complexity of the interactions between parasite-vector-host and the environment, that to a large extent remain to be elucidated (**Figure 1A and 1B**) (Büscher *et al.,* 2017). The presence of these two diseases negatively impacts the health of humans and livestock. The AAT in particular is a major obstacle to the development of a sustainable livestock production (Alsan, 2015). The FAO has estimated that an annual loss of USD 4.75 billion worth of agricultural produce are linked either directly or indirectly to a reduction of cattle production due to AAT (Bekele, 2015). Currently, according to the World health organisation (WHO), 70 million people are estimated to be at risk of contracting HAT. The number of new HAT cases is currently below 20,000

annually, but actual numbers fluctuate. In 2015, only 2,804 new cases were reported to WHO, whereby 86% of the cases were from Democratic Republic of Congo (Büscher *et al.,* 2017).

Trypanosomosis control methodology

Several strategies have been used to control or manage African trypanosomosis including the screening for infections and the curative treatment for HAT, and the prophylactic and curative AAT treatment using trypanocidal drugs (Meyer *et al.,* 2016). Some of the drugs used for HAT treatment are ineffective, toxic (Melarsoprol) and/or expensive (Eflornithine) and the use of prophylactic and trypanocidal drugs for AAT treatment has resulted in resistance to the parasite (Baker *et al.,* 2013; Geerts *et al.,* 2001; Kennedy, 2013; Matovu *et al.,* 2001). Consequently, the control of the tsetse fly vector is considered as the most feasible and sustainable way to manage these diseases (Schofield and Kabayo, 2008). In the past, vector control has involved bush clearing, wild game culling and spraying of insecticides like the DDT, dieldrin and endosulfan (Allsopp, 2001; Meyer *et al.,* 2016). Although the application of insecticides was successful in elimination of *G. m. submorsitans*, *G. p. palpalis* and *G. tachinoides* from a small region in northern Nigeria (Ormerod, 1986), there are challenges on its application especially over large areas. These include, the potential development of resistance by the tsetse flies, killing of non-target insects, outbreak or increase of other insect species due to the elimination of predators, pollution of the environment by the insecticide and health risks for staff exposed to the insecticide (Vreysen *et al.,* 2013).

As a result of the challenges facing the application of insecticides (Grant, 2001), four other vector control strategies have been developed which are environmentally and economically acceptable to control trypanosomosis. These include the use of the stationery attractive devices, the live bait technique, the sequential aerosol technique (SAT), and the sterile insect technique (SIT) (Bouyer *et al.,* 2013; Vreysen *et al.,* 2013). Although these vector control methods have demonstrated considerable success in reducing several isolated tsetse fly populations, SIT is considered the most effective among all, mainly because it is species-specific and has no effect to non-target organisms (Vreysen, 2001; Vreysen *et al.,* 2013).

The sterile insect technique (SIT)

The SIT is a potential insect birth control method, which was developed in 1950's to manage insect pests (Knipling, 1959). The SIT involves mass rearing of the target insect species, followed by sexual sterilization of the males by exposure to a precise dose of ionized radiation (Robinson, 2002; 2005). The sterile males are then released into the target wild insect populations to compete with wild type males for mating with the virgin wild type females. Mating of the sterile males with the wild females leads to embryogenesis arrest, and no viable offspring will be produced (**Figure 2**).

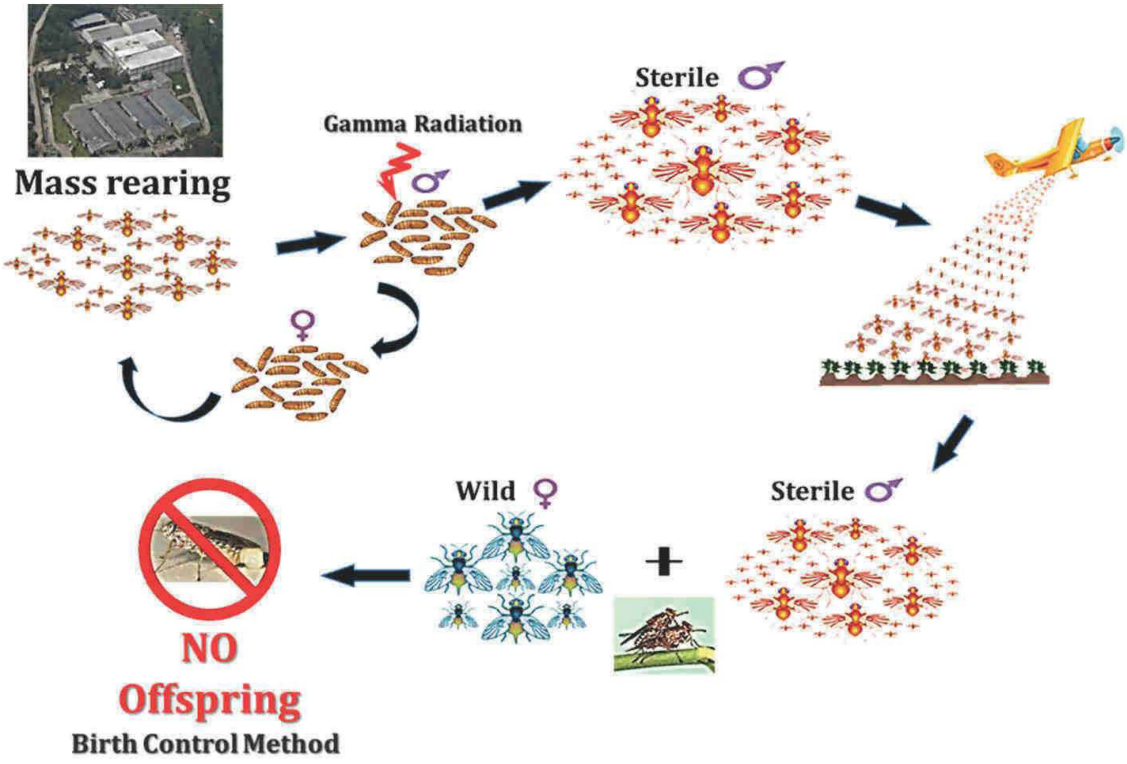


Figure 2: The principle of sterile insect technique (SIT) application for vector control. The process involves mass production of the target insect pest followed by sex separation. The females are retained in the colony in case of tsetse flies to maintain the colony while the males are sexually sterilized by irradiation. The sterile males are released to the target area to compete with wild type males for mating with virgin wild type females. The mating of a sterile male with a virgin female leads to embryogenesis arrest and no offspring is produced which can lead to the decline of the target insect (photo courtesy of IAEA).

Sustained release of sterile males over a period of time can eventually lead to a decline of the target insect population (Feldmann *et al.,* 2005). In addition to the above-mentioned advantages of SIT, there has been no evidence of development of resistance by the released sterile males. However, the method requires adequate knowledge on the biology and ecology of the target insect, and the target insect should be responsive to mass rearing. The SIT success is usually feasible when applied as part of the area-wide integrated pest management (AW-IPM) programme, which focuses on a sustained control of the target insect in a delineated area. The SIT/AW-IPM approach has been successful in control of several insect pests including the eradication of the screwworm fly in USA, Mexico, Central America and Libya, containment of the Mediterranean fruit fly in Guatemala and Mexico, and suppression of several lepidopteran pests (Bloem *et al.,* 2006; Enkerlin *et al.,* 2015; Wyss, 2000).

In the case of the tsetse vector control, SIT application led to severe suppression of *G. p. gambiensis, G. p. palpalis* and *G. tachinoides* in an agro-pastoral area in Burkina Faso (Politzar *et al.,* 1984). The most successful SIT application for tsetse control was the complete eradication of *G. austeni* from Unguja Island, United Republic of Tanzania (Vreysen *et al.,* 2000). This success inspired other African countries to apply SIT for tsetse control. For instance, one SIT programme was initiated in 1997 to eradicate *G. pallidipes* from the Southern rift valley of Ethiopia as a complement to the AW-IPM efforts that aim at a complete elimination of tsetse and trypanosomosis. Furthermore, a second programme initiated in 2009 to eradicate *G. p. gambiensis* in Niayes area in Dakar, Senegal, has not only reduced the fly population but it's towards a complete eradication (Alemu *et al.,* 2007; FAO, 2014; Vreysen *et al.,* 2013).

Challenges of SIT application for tsetse control

The SIT-mediated control for tsetse flies is advantageous because tsetse flies have a low rate of reproduction (the ovulation occurs every 9-10 days) (Gooding and Krafsur, 2005) and thus only limited numbers of sterile males are required in the field. However, at the same time, this low rate of production of tsetse flies is a challenge for SIT application since it makes the systems of mass production expensive and cumbersome (Vreysen *et al.,* 2013). Other challenges associated with

in SIT include; (i) the monitoring of fly density that can only be applied in lowly populated areas and thus prior suppression of the tsetse population is required using other control methods; (ii) the released sterile males may increase the transmission of trypanosomosis in the field (Vreysen, 2005). Most importantly, SIT requires production of high quality (sexual performance and flight ability) sterile males, which should compete with the wild type males and so proper mass rearing procedures (e.g., diets, handling, packaging and release) should be established to enable smooth implementation of the programmes (Vreysen, 2006; Vreysen *et al.,* 2011).

Virus infections: drawback to SIT implementation

During the initial implementation of SIT programme to control *G. pallidipes* in Ethiopia, the mass rearing of this tsetse species was established at the Insect Pest Control Laboratory (IPCL) of the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Seibersdorf, Austria, using field-collected pupae. Within two years of its establishment, the colony productivity was challenged by infections with the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV), which causes the salivary gland hypertrophy (SGH) syndrome (Abd-Alla *et al.,* 2011). The SGH syndrome was originally discovered in *G. pallidipes* flies in Zululand, South Africa (Whitnall, 1934) and was thereafter implicated in promoting trypanosome infections in tsetse flies (Burtt, 1945). Later, SGH-like symptoms were observed in other tsetse species including *G. m. centralis G. f. fuscipes, G. brevipalpis* and *G. swynnertoni* (Ellis and Maudlin, 1987; Jenni, 1973; Jenni and Steiger, 1974a; Shaw and Moloo, 1993). This SGH syndrome has also been implicated as the primary cause of colony collapse, initially in *G. pallidipes* colony maintained at the Biotechnology Research Institute of the Kenya Agricultural and Livestock Research Organization (BRI-KALRO; the defunct Trypanosomiasis Research Institute; KETRI). This colony declined due to poor productivity within two years after its establishment (Opiyo and Okumu, 1983). Later on, GpSGHV infections caused collapse of two *G. pallidipes* colonies maintained at the IPCL in Seibersdorf, Austria, in 1987 and 2002, with flies that originated from Kenya and Ethiopia, respectively (Abd-Alla *et al.,* 2010b). In addition to the poor performance and colony instability caused by the SGH syndrome in tsetse flies, males with SGH syndrome had reduced

mating propensity and competitiveness, which directly impacted the implementation and success of the SIT programs (Mutika *et al.,* 2012).

Hytrosavirus pathogenesis

This distinctive SGH syndrome associated with GpSGHV infections has also been reported in other non-hematophagous dipteran insects such as narcissus bulb fly, *Merodon equestris nobilis* and *Merodon equestris transversalis* (Diptera; Syphidae) (Amargier *et al.,* 1979) and in housefly, *Musca domestica* (Diptera; Muscidae) (**Figure 3**) (Coler *et al.,* 1993).

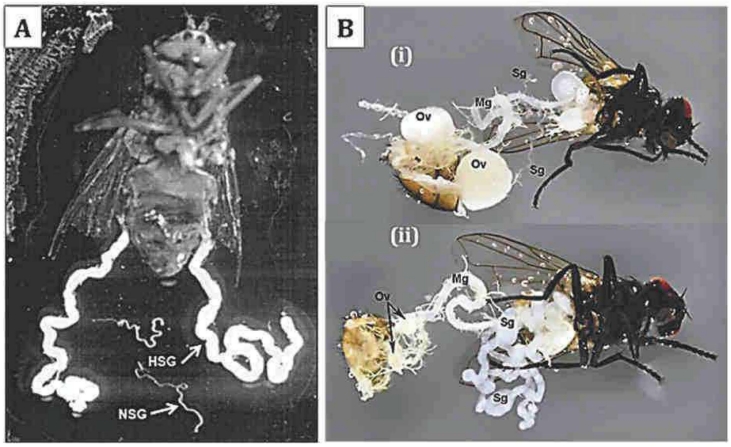


Figure 3: Phenotypes exhibited by flies infected with hytrosaviruses. (A) A GpSGHV infected *G. pallidipes* male fly with showing salivary gland hypertrophy (HSG) symptom compared to normal salivary gland (NSG) dissected from uninfected tsetse fly. (B) A comparison between uninfected female housefly with normal salivary glands (Sg) and healthy ovaries (Ov) (i), and MdSGHV infected female housefly showing hypertrophied salivary glands (Sg) and undeveloped ovaries (Ov). No morphological changes in the midgut (Md) between the non-infected and infected flies. (Figure sources; panel A (Abd-Alla *et al.,* 2007) and panel B (Lietze *et al.,* 2011a)).

In adult tsetse flies, GpSGHV can cause either asymptomatic or symptomatic infections (associated with the SGH syndrome). Sometimes, due to unknown factors the asymptomatic infections can switch to symptomatic infections, particularly in *G. pallidipes* (Abd-Alla *et al.,* 2010b). In the field incidence of SGH is much less prevalent, probably due to mortality of infected flies escaping detection. In mass rearing facilities, GpSGHV can be vertically transmitted via milk glands secretions

or transovum to the offspring and horizontally during the in vitro blood membrane feeding technique, where the virus is released into the blood meal by the infected flies via the saliva (Abd-Alla *et al.,* 2010b). The symptomatic infections in tsetse flies cause gonadal/ovarian anomalies, distorted sex ratios and reduced insemination rates, thus reducing the reproduction potential of the flies (Jaenson, 1986; Jura *et al.,* 1988; Sang *et al.,* 1998; 1999; 1996). The GpSGHV-tsetse infection drastically differs from the *Musca domestica* SGHV-housefly infection model. In houseflies the infection is invariably symptomatic, and the virus causes complete shutdown of oogenesis, which precludes vertical transmission of MdSGHV (Abd-Alla *et al.,* 2010b; Lietze *et al.,* 2007).

General features of hytrosaviruses

Although the SGH syndrome was reported in narcissus bulb fly, most of the studies have focused on hytrosaviruses (SGHVs) infecting tsetse fly and housefly, which have been classified in two genera of the *Hytrosaviridae* family, the *Glossinavirus* (GpSGHV) and *Muscavirus* (MdSGHV) (Abd-Alla *et al.,* 2009a). The SGHVs (hytrosaviruses) particles are enveloped and contain rod shaped nucleocapsids wrapping a circular double-stranded DNA (dsDNA) genome. SGHVs replicate in the nucleus of the infected cells and are specifically pathogenic to their hosts (Abd-Alla *et al.,* 2008; Garcia-Maruniak *et al.,* 2008; Jaenson, 1978; Odindo *et al.,* 1986). MdSGHV infections result in enlargement of the nuclear and cytoplasm (cellular hypertrophy), while GpSGHV infections induce rapid cell division/replication (cellular hyperplasia) of the infected SGs (**Figure 4**) (Kariithi *et al.,* 2017a).

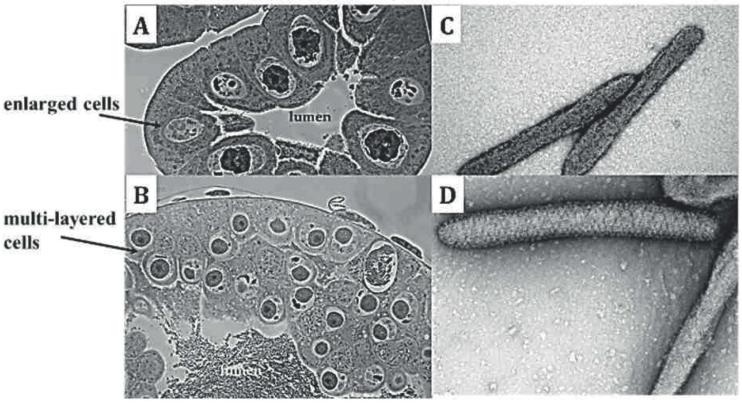


Figure 4: (A) MdSGHV-induced cellular hypertrophy (enlarged cells) in housefly salivary glands. (B) GpSGHV-induced hyperplasia (multi-layered cells) in tsetse salivary glands. Transmission electron microscopy (TEM) micrographs showing the main structural features of (C) MdSGHV and (D) GpSGHV particles: Figure source; (Kariithi *et al.,* 2017a).

The genomes of the MdSGHV and Ugandan and Ethiopian strains of GpSGHV (GpSGHV-Ug and GpSGHV-Eth) have been fully sequenced. The MdSGHV genome of about 124 kbp encodes 108 open reading frames (ORFs), while the GpSGHV-Uga and GpSGHV-Eth strains both with 190 kbp, encode 160 ORFs and 174 ORFs, respectively (Abd-Alla *et al.,* 2008; 2016; Garcia-Maruniak *et al.,* 2008). A comparative analysis of the SGHVs showed that 37 MdSGHV ORFs have homology to 42 GpSGHV ORFs, which suggest a common ancestry of the SGHVs (**Figure 5**) (Garcia-Maruniak *et al.,* 2009).

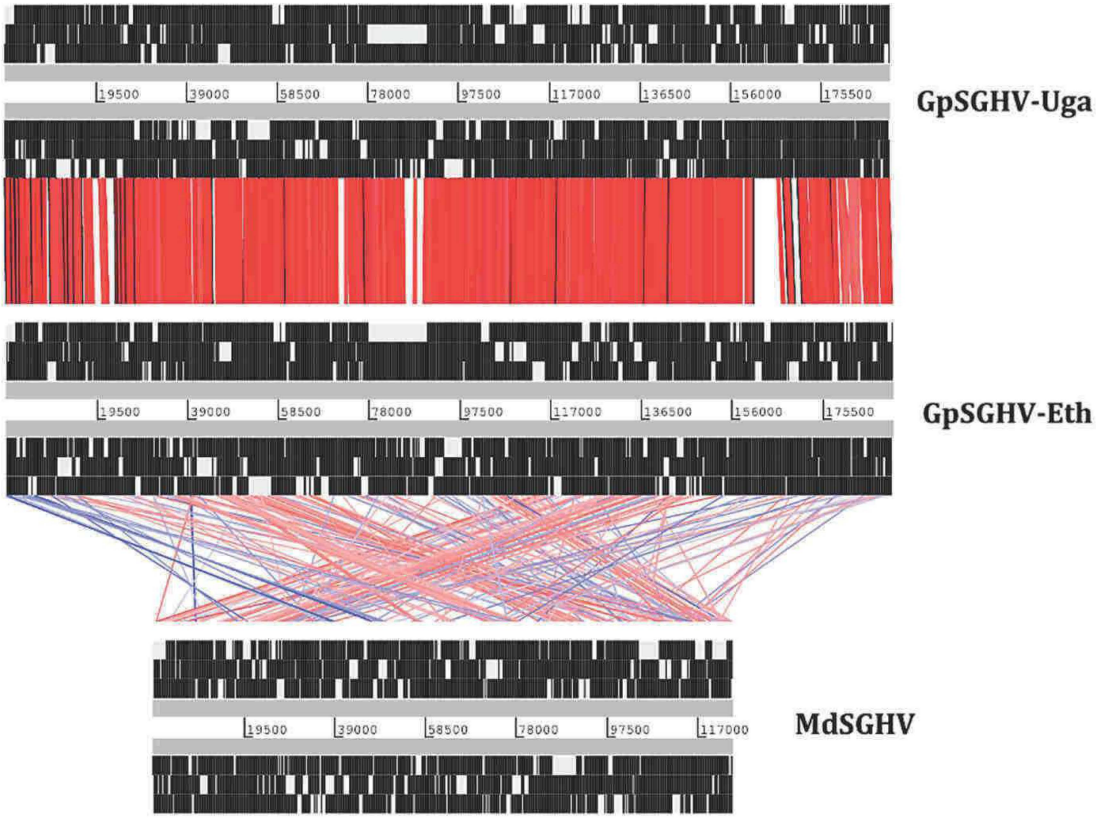


Figure 5: Synteny maps comparing the overall collinearity of GpSGHV and MdSGHV genomes. The GpSGHV-Uga (Accession Number: EF568108) and GpSGHV-Eth strains (Accession Number: KU050077) were compared then compared to the MdSGHV (Accession Number: EU522111). The red lines indicate the levels of identity between the viruses while the blue lines indicate the inversions. The genomic positions are shown between the lines of each viral genome. The black bands between the conserved regions do not necessarily indicate the ORFs, but the conserved genomic regions (Abd-Alla *et al.,* 2016).

The two GpSGHV strains induce different pathogenesis probably due to the genetic differences between the two strains and other unknown factors (Abd-Alla *et al.,* 2016). A recent study has reported that GpSGHV has diverged to a small degree in *G. pallidipes* populations in sub-Saharan African (Kariithi *et al.,* 2013b). The SGHVs share general characteristics with other nuclear-replicating insect viruses (e.g. baculoviruses and nudiviruses) also containing large, circular, dsDNA genomes and replicating in the nucleus. For example, SGHVs share with baculoviruses 12 core genes, including five *per os* infectivity factors (PIF) genes that encode proteins that are critical for oral infections (Abd-Alla *et al.,* 2008; Jehle *et al.,* 2013). Phylogenetic analysis of the conserved genes clustered the SGHVs in a separate clade from the baculoviruses and nudiviruses, which may explain the major differences in pathology, since none of the baculoviruses or nudiviruses induce SGH (Garcia-Maruniak *et al.,* 2009).

GpSGHV-tsetse-symbionts interactions

Although the circumstances under which the asymptomatic SGHV infections switch to symptomatic infections in tsetse flies are not well understood, the interactions between the virus and tsetse-symbionts have been reported to influence the viral pathogenesis (Boucias *et al.,* 2013). In general, insects harbour symbionts that contribute to their nutrition and reproduction, e.g. by modulation of the insect's immune response to offer defence against pathogens and parasites (Moreira *et al.,* 2009; Snyder *et al.,* 2010). Tsetse flies harbour one obligate (*Wigglesworthia glossinidia*) and two facultative (*Sodalis glossinidius* and *Wolbachia pipientis*) bacterial symbionts (O'Neill *et al.,* 1993). *Wigglesworthia* provides nutrients to its tsetse host, influences the host immune responses, and has an impact on vector competence of the tsetse flies to trypanosome infections (Pais *et al.,* 2008; Wang *et al.,* 2009). *Sodalis* has been reported to influence the host's ability to establish trypanosome infections (Dale and Moran, 2006; Soumana *et al.,* 2014a). *Wolbachia* is present in some tsetse species and can induce cytoplasmic incompatibility (CI), a phenomenon that leads to embryogenesis arrest when an uninfected female fly mates with a *Wolbachia*-infected male (Alam *et al.,* 2011; Doudoumis *et al.,* 2012). Many studies have reported that presence of *Wolbachia* in insects inhibits pathogen infections, probably by activating the immune response of the insects (Hussain *et al.,*

2011; Kambris *et al.,* 2010; Moreira *et al.,* 2009; Teixeira *et al.,* 2008). Recently, a fourth endosymbiont bacteria *Spiroplasma*, was found in tsetse species belonging to the Palpalis sub-group, i.e. *G. f. fuscipes* and *G. tachinoides* (Doudoumis *et al.,* 2017). Although the exact role of *Spiroplasma* is not known yet in tsetse flies, studies have shown this bacterium can play a protective role against certain microbial infections similarly to *Wolbachia* in *Drosophila* by affecting the immune signalling pathways (Xie *et al.,* 2010; Yadav *et al.,* 2018). In tsetse flies, the absence or low levels of *Wolbachia* in *G. pallidipes* may be linked to the occurrence of SGH in this species as compared to *G. m. morsitans* that presents a high *Wolbachia* prevalence (Doudoumis *et al.,* 2012). Antibiotic-mediated removal of the *Wigglesworthia-Sodalis* complex in *G. pallidipes* flies reduced transgenerational transmission of GpSGHV and reduced expression of SGH symptoms in the F_1 fly progeny (Boucias *et al.,* 2013; Wang *et al.,* 2013a). Furthermore, the absence of *Wigglesworthia* led to loss of *Sodalis* and decreased GpSGHV titers (Wang *et al.,* 2013a). Although the precise mechanisms that underpin these GpSGHV transmission and pathologies are unclear, these findings indicate that there is interdependency among the tsetse symbionts and their interaction with GpSGHV could be related to the host's immune responses. Alternatively, GpSGHV may have co-evolved with these symbionts and so the symbionts may mediate the switch from asymptomatic to symptomatic state (Boucias *et al.,* 2013).

Factors influencing the defence system in tsetse

The above-mentioned host-virus-symbiont interactions call for further investigations into the tsetse immune system and the identification of factors (biotic and abiotic) that influence or modulate the outcome of GpSGHV infections. In insects, the principle antiviral defence mechanism is the RNA interference (RNAi), which utilizes double-stranded RNAs (dsRNAs), particularly the short interfering RNAs (siRNAs) and microRNAs (miRNAs) that target and regulate host or virus gene expression and hence control virus infection (Ding, 2010). For instance, inhibition of the key RNAi elements (Argonaute-2 and Dicer-2) in some insects led to increased replication of both RNA and DNA viruses (Galiana-Arnoux *et al.,* 2006; Jayachandran *et al.,* 2012; van Rij *et al.,* 2006). The Janus kinase/signal transducers and activators of transcription (Jak-STAT) is another immune pathway that

contributes to antiviral responses usually by regulating the expression of antimicrobial peptides (AMPs) in response to viral or bacterial infections (Dostert *et al.,* 2005; Kingsolver *et al.,* 2013). Other immune pathways, which are known to mainly respond to bacteria and fungi infections, are the immune deficiency (Imd) and Toll pathways (Lemaitre and Hoffmann, 2007; Michel *et al.,* 2001), which may also directly or indirectly play a role in antiviral immunity in insects (Costa *et al.,* 2009; Xi *et al.,* 2008; Zambon *et al.,* 2005).

Among these immune pathways, the RNAi mechanism is a potent approach that can be utilized to develop strategies for insect pest management as well as to control pathogen infections in insects (Burand and Hunter, 2013). For instance, RNAi was demonstrated in honey-bees to control the Israeli Acute Paralysis Virus (IAPV) infections, which usually lead to increased mortality and decline of colonies. Injection of dsRNAs specific for IAPV led to decreased virus infections in the colonies, while the use of these dsRNAs as food additive in the field increased the honey-bee production in virus infected hives (Burand and Hunter, 2013). Although a similar RNAi approach was previously proposed in management of GpSGHV infections in mass rearing facilities by targeting specific GpSGHV genes (Abd-Alla *et al.,* 2011), further studies are required to understand the RNAi mechanisms in tsetse during asymptomatic and symptomatic infections. Currently, two control strategies are being employed to manage the virus infections in tsetse mass rearing facilities. These include the clean blood feeding system that aims to reduce the horizontal transmission and the administration of antiviral drug valacyclovir that targets the GpSGHV DNA polymerase and inhibits virus replication (Abd-Alla *et al.,* 2014).

Rationale and scope of the thesis

Although the studies described above have paved the way for further understanding of the mechanisms of GpSGHV infections and SGH outbreaks in tsetse mass rearing facilities, there are many questions that remain unanswered. For instance, how does the coevolution of SGHVs with their hosts influence the host's immunity and hence impact the outcome of the virus infection? What is the GpSGHV prevalence and diversity in tsetse species in the wild? Is there a relationship between different

GpSGHV strains and the tsetse species or subspecies they infect belonging to different taxonomic groups? What host and/or viral molecular mechanisms maintain the asymptomatic GpSGHV infection state in some tsetse species (*G. pallidipes*) and what factors prompt the switch to symptomatic infections? Which immune pathways are directly or indirectly targeted by the virus or involved in the establishment of its latency/persistence or induction of the SGH syndrome?

The major goal of this thesis was to investigate the possible mechanisms that trigger the occurrence of symptomatic GpSGHV infections associated with SGH outbreaks in tsetse mass rearing facilities. Furthermore, the thesis research aimed to understand how SGHV infection could remain asymptomatic (latency/persistence) within natural and colonized fly populations. This thesis also focuses on the diversity of GpSGHV amongst different tsetse species both from the field and in laboratory settings and how GpSGHV is able to evade the tsetse immune surveillance leading to SGH.

In **Chapter 2**, the ecologies and life-histories of tsetse fly and housefly that may influence coevolution of the hytrosaviruses GpSGHV and MdSGHV with their hosts immune responses are reviewed. The chapter discusses the host range specificity, transmission dynamics and pathogenesis of the viruses and their interaction with symbionts. The possible mechanisms of SGHV infections are explored, and immune genes and pathways that may play a role in infections are discussed. Phylogenetic relatedness of these immune genes in *G. pallidipes* and *M. domestica* is presented in reference to other insect-virus infection models for which the immune genes have been significantly annotated.

Chapter 3 describes the development of novel molecular tools that can be applied to quickly, easily, massively and cost-effectively identify different tsetse species in the field. This is an important aspect particularly in the establishment of new tsetse mass production facilities, which mainly uses materials (pupae) from the wild, or from existing colonies. This species identification tool was developed by integrating different classes of markers and by utilizing different resolution techniques, such as gel electrophoresis and sequencing.

Chapter 4 addresses the question as to whether different tsetse species are infected by GpSGHV in nature and if they are, whether they are infected with same virus strain. To address these questions, the prevalence of GpSGHV in seven tsetse species (identified using the tools developed in Chapter 3) from different geographical locations in sub-Saharan Africa was determined by polymerase chain reaction (PCR) and phylogenetic analyses.

In **Chapter 5**, the hypothesis is tested whether GpSGHV infection induces RNAi response in *G. pallidipes* or not to prevent the development of overt SGH and instead induces a covert infection state. This hypothesis was investigated by a comparative analysis of the expression and modulation of key genes of the RNAi machinery, particularly the short interfering RNA (siRNA) pathway.

Chapter 6 investigates the possible involvement of host and GpSGHV-encoded miRNAs in virus-host interaction during GpSGHV infections in *G. pallidipes*. This was done using a variety of methods, including deep sequencing of small RNAs (sRNA) molecules, miRNA target prediction analysis and miRNA inhibition assays.

Finally, in **Chapter 7** the findings of the thesis research are discussed and placed in a wide perspective on their applications in the control/management of GpSGHV infections in tsetse mass production facilities and the impacts on the implementation of the SIT applications in the control of the tsetse vector.

Chapter 2

Coevolution of hytrosaviruses and host immune responses

This chapter has been modified from:

Kariithi, H.M., Boucias, D.G., Murungi, E.K., **Meki, I.K.,** Demirbas-Uzel, D.G., van Oers, M.M., Vreysen, M.J.B., Abd-Alla, A.M.M. and Vlak, J.M. Coevolution of hytrosaviruses and host immune responses. BMC Microbiology, (*in press*).

Abstract

Hytrosaviruses (*Hytrosaviridae* family) are double-stranded DNA (dsDNA) viruses that cause salivary gland hypertrophy (SGH) syndrome in adult dipterans. Two structurally and functionally distinct hytrosaviruses are currently known; *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) and *Musca domestica* salivary gland hypertrophy virus (MdSGHV), that infect the hematophagous tsetse fly and the housefly, respectively. Genome sizes and gene contents of GpSGHV (~190 kb; 160-174 genes) and MdSGHV (~124 kb; 108 genes) may reflect an evolution of these viruses with their hosts resulting in differences in pathobiology. Whereas GpSGHV can switch from asymptomatic to symptomatic infections in response to certain unknown circumstances, MdSGHV only infects symptomatically, which is characterised by overt SGH symptoms. Whereas MdSGHV induces extensive nuclear and cellular hypertrophy of the SGs (enlarged non-replicative cells) thereby causing a non-lytic increase in individual cell sizes, but without an increase in cell numbers, GpSGHV induces cellular hyperplasia (enlarged replicative cells). GpSGHV infects specifically *Glossina* species, while MdSGHV can infect also other sympatric muscids. The MdSGHV infections induce total shutdown of oogenesis in the housefly and hence inhibit the vertical transmission of the virus to the offspring, while the GpSGHV asymptomatic infections promote vertical transmission. This chapter discusses the coevolution of the SGHVs and their hosts' immune responses, the influence of the evolution of the transmission modes of these viruses on their pathogenesis, and the contribution of bacterial symbionts to expression of overt SGH symptoms. The chapter also discusses the hypothesis that, by recruiting cellular genes from ancestral host(s) into its genome, GpSGHV seems to have evolved to avoid its host's immune responses in contrast to MdSGHV, which appears to have evolved strategies to counteract both the housefly's RNAi and apoptotic responses. With respect to the hosts' measures to counter SGHVs invasion, the housefly appears to have expanded its repertoire of immune effector, modulator and melanization genes compared to the tsetse fly. Therefore, the ecologies and life-histories of the housefly and tsetse fly may significantly have influenced the coevolution of MdSGHV and GpSGHV with their hosts.

Introduction

The salivary gland hypertrophy viruses (SGHVs) belong to the *Hytrosaviridae* family (Abd-Alla *et al.,* 2009a), a relative newly described family of insect double-stranded DNA (dsDNA) viruses that infect dipteran insects with distinct ecologies and evolutionary histories. Known SGHV hosts are the hematophagous *Glossina* species (tsetse fly), *Musca domestica* (common housefly), and most probably the *Merodon equestris* Fabricius (narcissus bulb fly) (Kariithi *et al.,* 2013a). SGHV infections result in the swelling of the host's salivary glands (SGs) thereby producing diagnostic SG hypertrophy (SGH) syndrome (Kariithi *et al.,* 2017a). The SGHVs primarily replicate in adult flies, and cause a chronic infection that leads to reproductive dysfunctions (Lietze *et al.,* 2011a). In tsetse fly mass rearing facilities, asymptomatic *Glossina pallidipes* SGHV (GpSGHV) infections can switch to symptomatic infections and reduce the flies' productivity which can consequently cause colony collapse (Abd-Alla *et al.,* 2011). Although the SGHVs replicate also in non-SG tissues, these viruses have not been found to replicate in cell lines established from homologous or heterologous insect hosts, which has hindered detailed genetic studies of the SGHVs (Arif and Pavlik, 2013).

SGHVs are enveloped and rod-shaped viruses with a circular dsDNA genome (Abd-Alla *et al.,* 2008; 2016; Garcia-Maruniak *et al.,* 2008). Structurally, SGHVs resemble the well-studied baculoviruses (Jehle *et al.,* 2013), which are phylogenetically clustered together with other nuclear-replicating large dsDNA viruses such as the nudiviruses and more distantly also the nimavirus (Wang and Jehle, 2009). Functionally however, the SGHVs are distinguished from baculoviruses since they lack the occlusion bodies and they rarely kill their host (i.e. lower lethality) (Ishimwe *et al.,* 2015). The GpSGHV 190 kbp genome encodes 160 open reading frames (ORFs), of which 42 are homology to 37 ORFs of the 108 ORFs encoded by the 124 kbp *Musca domestica* SGHV (MdSGHV) genome (Abd-Alla *et al.,* 2008; 2016; Garcia-Maruniak *et al.,* 2008). The relatively low number of genes shared between the two sequenced SGHV genomes as well as the remarkable differences in genome organisation has led to their placement into two distinct genera (GpSGHV: *Glossinavirus* and MdSGHV: *Muscavirus*) within the *Hytrosaviridae* family.

Phylogenetically, neither of these SGHVs could be placed within any of the other established dsDNA virus families (Garcia-Maruniak *et al.,* 2009; Jehle *et al.,* 2013).

The only current member of the genus *Muscavirus* (MdSGHV) infects houseflies and then causes symptomatic infections (Lietze *et al.,* 2007; 2012). In addition, injection of MdSGHV into healthy adult houseflies results in overt SGH and a total shutdown of oogenesis, thus inhibiting vertical transmission to the progeny (Kariithi *et al.,* 2017b; Lietze *et al.,* 2007). Asymptomatic infections of MdSGHV do not occur. The MdSGHV is widely distributed within populations of the housefly (Prompiboon *et al.,* 2010), a mobile insect that moves several kilometers in search for feeding and oviposition sites associated with livestock keeping (Nayduch and Burrus, 2017). Sequence analysis of MdSGHV isolates from different geographical regions using selected genes, revealed low polymorphism of the isolates associated with the host's geographical origin (Prompiboon *et al.,* 2010). This low viral polymorphism may be due to the close associations of the domestic housefly with human movements, which may influence the frequency of MdSGHV-housefly interactions. Within housefly populations, MdSGHV induces variable rates of SGH prevalence, for example 0-40% prevalence was found in North Florida dairy farms (Geden *et al.,* 2008; Lietze *et al.,* 2013). This variation is probably due to the fly's seasonal densities at the various sampling sites since high housefly densities may favor horizontal transmission of MdSGHV via contaminated shared feeding sites (Antonovics *et al.,* 2017).

The GpSGHV exists mainly in an asymptomatic infection state in tsetse flies, but certain unknown (biotic and abiotic) factors can trigger development of overt SGH symptoms (Abd-Alla *et al.,* 2010b; Kariithi *et al.,* 2013a). Injection of GpSGHV into healthy *G. pallidipes* flies does not induce overt SGH symptoms in the injected generation, but in some of the F_1 progenies (Boucias *et al.,* 2013). Unlike the widely distributed MdSGHV, GpSGHV is specific to *Glossina* species, which are restricted to sub-Saharan Africa (see **Chapter 1** of this thesis). Tsetse fly distribution is mainly determined by habitat, environmental conditions and host animal dynamics. Compared to the highly mobile houseflies, tsetse flies are fairly inactive, and make random movements of only 150-550 meters per day (Leak, 1998). Based on selected conserved viral genes, GpSGHV has a lower level of polymorphism than MdSGHV

(Kariithi *et al.,* 2013b; Prompiboon *et al.,* 2010). Of all the *Glossina* species (in both laboratory and wild populations), it is only *G. pallidipes* that often exhibits overt SGH symptoms. Given that an asymptomatic (persistent) infection signifies the best adapted or most successful virus-host coevolution (Goic and Saleh, 2012), *G. pallidipes* could be the most recent GpSGHV host, which may partially explains the absence of overt SGH symptoms in other *Glossina* species (see **Chapter 4** of this thesis).

The current chapter evaluates the coevolution of SGHVs (Abd-Alla *et al.,* 2008; 2016; Garcia-Maruniak *et al.,* 2008) and their hosts (International Glossina Genome Initiative, 2014; Scott *et al.,* 2014). The possible mechanisms of SGHV infections are explored, and immune genes and pathways that may play a role during the SGHVs infections are discussed. Phylogenetic relatedness of the immune genes in *G. pallidipes* and *M. domestica* was determined in relation to orthologs from the model insect, *Drosophila melanogaster,* and the African malaria mosquito, *Anopheles gambiae,* the virus-host systems for which the immune genes have been identified and characterized (Kingsolver *et al.,* 2013; Merkling and van Rij, 2013).

Methods

The annotated immune genes of *D. melanogaster* and *An. gambiae* were retrieved from ImmunoDB (Waterhouse *et al.,* 2007) and used to query (BLASTp; e-value ≤10−4) the predicted proteomes of *G. pallidipes*, and *M. domestica* that were retrieved from VectorBase (Giraldo-Calderón *et al.,* 2015). Canonical domains in the identified immune genes were ascertained using Pfam (Finn *et al.,* 2016), and pathways were confirmed by BLASTp searches in the Insect innate Immunity Database (IIID) (Brucker *et al.,* 2012). To decipher phylogenetic relatedness of orthologous immune genes in the genomes of these four insect species, the retrieved sequences were aligned using MAFFT v7 (Katoh and Standley, 2013) and used for phylogenetic reconstructions using PhyML 3.0 (Guindon *et al.,* 2010) and MrBayes v3.2 (Ronquist *et al.,* 2012). The robustness of internal branches was evaluated using 100 bootstraps.

SGHV host-range specificity, transmission dynamics and pathogenesis

Infection dynamics of MdSGHV in houseflies

Studies on the MdSGHV-infected houseflies showed that SGH is caused by the hypertrophy of the nucleus and cytoplasm of the infected SG cells (Lietze *et al.,* 2011b). In addition to the MdSGHV replication in the SGs, evidence suggests that this virus also replicates in the housefly's corpora-allata/cardiaca (CA/CC), which are the glands that produce neurohormones and juvenile hormones to regulate reproduction and metamorphosis (Kariithi *et al.,* 2017b). In addition, reverse transcription quantitative PCR (RT-qPCR) analysis of female specific proteins in MdSGHV-infected houseflies demonstrated that the virus replication blocked the transcription of hexamerin and yolk proteins involved in egg development, which may explain how the virus-induced reproductive dysfunctions (Lietze *et al.,* 2007).

Maintaining uninfected houseflies together with SGH-positive houseflies in cages, at various densities with a shared food source for 10 filial generations, resulted in low percentages (~10%) of SGH-positive houseflies that persisted throughout the generations (~12 weeks) (Lietze *et al.,* 2012). This result demonstrates that although MdSGHV can be transmitted *per os* amongst houseflies that feed on shared food source (Geden *et al.,* 2008), the virus is maintained at a low prevalence in the housefly populations. In addition, feeding newly-eclosed (< 2 h-old) houseflies with MdSGHV suspensions in the laboratory induced overt SGH symptoms in ~53% of the individuals (Prompiboon *et al.,* 2010). In this scenario, during feeding, an infected housefly can release onto the food salivary secretions containing copious amounts of infectious viral particles which can be ingested by the non-infected flies (Lietze *et al.,* 2009). However, in nature, newly eclosed flies do not ingest food until after several hours (12-24 h) post-eclosion, after which the synthesis of the peritrophic matrix (PM) that protects the gut has occurred and thus rendering theflies highly resistant to orally-ingested virus (Lietze *et al.,* 2013; Prompiboon *et al.,* 2010). Together, these findings suggest that MdSGHV can be transmitted horizontally to healthy conspecifics through either feeding or even through cuticle wounds (high densities) when flies feed together at virus-contaminated sites (Vallejo *et al.,* 2013), a scenario that has been reported in laboratory-reared houseflies which

contract MdSGHV infections when introduced into virus-contaminated fly cages (Geden *et al.,* 2008).

The absence of an asymptomatic infection state, and the lack of vertical transmission of MdSGHV in house flies raise the question of how MdSGHV has evolved to maintain its infection within natural fly populations especially during low seasonal fly densities (reduced horizontal transmission possibilities). Moreover, the SGHVs lack the occlusion bodies which in the case of baculoviruses facilitate their long-term survival outside of their hosts (Slack and Arif, 2006). In addition, although the housefly's saliva and its released contents may stabilize the MdSGHV particles released during the feeding events, cage studies have shown that the MdSGHV maintains a low frequency of infection over multiple fly generations (Lietze *et al.,* 2012). However, in nature, in addition to the virus transmission at high densities via cuticle wounds as mentioned above, MdSGHV may reside asymptomatically in reservoir hosts, which may contribute to the maintenance of the virus within fly populations. For example, under laboratory conditions, other muscids such as the obligate hematophagous stable fly (*Stomoxys calcitrans*), and the larval predator of the housefly, the black dump fly (*Hydrotaea aenescens*) supported MdSGHV replication without expression of overt SGH symptoms (Geden *et al.,* 2011a; 2011b). However, it is not known whether these or other muscids can transmit infectious MdSGHV particles to healthy houseflies if they share the same food sources.

Infection dynamics of GpSGHV in tsetse flies

Investigations on GpSGHV symptomatically infected tsetse SGs have showed that unlike the MdSGHV that induces both nuclear and cellular hypertrophy (enlarged cells incapable of replication), GpSGHV induces cellular hyperplasia (enlarged cells capable of replication), which result in the overt SGH symptoms (Kariithi *et al.,* 2013a; 2017b). Different *Glossina* species show wide variations in their susceptibilities to GpSGHV infection. For example, intra-haemocoelic GpSGHV injection in six *Glossina* spp. derived from laboratory colonies, showed that *G. pallidipes and G. morsitans morsitans* were the most susceptible to the virus (Demirbas-Uzel *et al.,* 2018). Additionally, as discussed in Chapter 4 of this thesis, GpSGHV prevalence is higher in the wild-caught *G. pallidipes* (0-100%) and *G. m. morsitans* (0-60%) than in other tsetse species. Further analysis by sequencing

some selected viral genes showed that, whereas only one GpSGHV haplotype (viral sequences with similar genetic variants and mutations events) infected the other tsetse species in this study, 14 and four GpSGHV haplotypes infected the *G. pallidipes* and *G. m. morsitans* flies respectively (see **Chapter 4**).

Under laboratory conditions, GpSGHV can be transmitted horizontally by the infected flies via saliva during the *in vitro* membrane feeding technique used in mass-rearing facilities, where flies can ingest infectious viral particles released via saliva (Abd-Alla *et al.,* 2013; 2010b). In the wild populations, it is hypothesized that GpSGHV can also be horizontally transmitted since tsetse flies aggregate on specific parts of the host to feed (Spath, 2000; Van Den Abbeele *et al.,* 2010). The flies then produce active saliva components that are deposited at the feeding site to interfere with host responses such as vasoconstriction and thrombocyte aggregation. This helps create a blood pool at the bite site and maintain blood fluidity as well as reducing the blood-diffusion rate (Caljon *et al.,* 2010). This may reduce the dilution of any infectious viral particles released via the saliva of infected flies at the bite site and hence increase the chances of horizontal virus transmission to other flies feeding at the bite site. However, it is unclear whether different tsetse species in the wild populations feed together on the same animal (Leak, 1998), as well as the fact that majority of tsetse species display asymptomatic infections, and the number of virus particles deposited during feeding is lower (~10^2 viral copies) as compared with the levels deposited by symptomatic flies (~10^6 viral copies) (Abd-Alla *et al.,* 2010b). The dynamics of GpSGHV transmission probably depend on the feeding behavior of specific *Glossina* species, their feeding preferences, feeding time, proximity of uninfected flies to infected flies, and the susceptibility of a particular tsetse population to virus infection. For instance, more than 1000 *G. pallidipes* flies living in the same habitat can feed on the same individual animal host daily (Pollock, 1982). In addition, compared to the short-lived (15-30 days) gregarious housefly, the solitary tsetse fly lives longer (120-150 days), which may influence virus transmission.

Notably, tsetse flies reproduce by adenotrophic viviparity since female's egg contains sufficient yolk, which allows the development of the embryo and the larvae are further fed with intrauterine produced by the milk glands in the uterus of the

female fly (Benoit *et al.,* 2015). In the case of vertical transmission by *G. pallidipes* females, the virus undergoes a certain degree of replication in the milk gland cells providing a channel into developing larvae/pupae. Virus-injected mothers therefore produce an F1generation that displays a high prevalence of SGH and reproductive dysfunctions (Boucias *et al.,* 2013). In addition, experimental data suggest that adults that emerge from GpSGHV injected *G. pallidipes* third-instars (instantly deposited larvae) have a high SGH prevalence (Demirbas-Uzel *et al.,* 2018).

It is unclear whether the asymptomatic infections observed within laboratory-bred and wild tsetse populations represents a persistent infection state, or a latent infection state (Kariithi *et al.,* 2013b) (see also **Chapter 4**). In persistent infections, the virus remains in specific cells of infected individuals, and is accompanied by a low-level production of virions, but without cellular damages (Boldogh *et al.,* 1996). During latent infections, viral genomes and proteins are present in infected cells for a certain period, but without formation of infectious viral particles (Goic and Saleh, 2012). Notably, a virus can cause both persistent and latent infections in the same host at the same time, but in different cells or tissues, which may or may not be dictated by the tissue tropism of the virus (Abd-Alla *et al.,* 2010b; Boldogh *et al.,* 1996). It is speculated that during asymptomatic infections, GpSGHV exists in both persistent and latent infection states at the same time. For instance, the release of low amounts of virus (~10^2 viral genome copies) via saliva during feeding by an asymptomatic fly (Abd-Alla *et al.,* 2010b) supports the idea that GpSGHV exists in a persistent infection state in the SG cells, i.e. the virus replicates at such low levels that small amounts of viral particles are produced in the SGs. At the same time, the virus may latently infect other tissues in which viral DNA is detectable but no transcripts (Kariithi *et al.,* 2017a) and in both cases, the virus does not induce SGH symptoms or reproductive dysfunctions (Abd-Alla *et al.,* 2010b). It is unlikely that GpSGHV integrates into the host genome in the form of a provirus, since using GpSGHV genes as probes did not indicate such integration (Unpublished data). Altogether, it appears that GpSGHV has evolutionarily selected asymptomatic infection as a survival strategy.

The reciprocal tripartite SGHV-host-symbiont interactions

Symbiont-mediated host immune system

Insects with restricted diets such as the hematophagous tsetse flies harbour specific symbionts that help synthesize essential nutrients or digest the ingested food. Tsetse flies harbor a unique community of three bacterial endosymbionts, i.e. *Wigglesworthia glossinidia, Sodalis glossinidius* and *Wolbachia pipientis* (Wang *et al.,* 2013b). In addition, a fourth endosymbiont bacterium known as *Spiroplasma*, was discovered in *G. f. fuscipes* and *G. tachinoides* (tsetse species belonging to the Palpalis group). Similar to GpSGHV, these three symbionts are maternally transmitted to the developing larvae transovarially or via the milk gland secretions (Boucias *et al.,* 2013; De Vooght *et al.,* 2015; Wang *et al.,* 2013b). On the other hand, houseflies lack or harbor only low densities of these symbionts. For example Bahrndorff *et al.* (2017) reported *Wolbachia* infections in less than 4% of *M. domestica* females collected from 10 widely distributed farms in Denmark.

The presence or variable densities of these symbionts influences the immune competence of the insects host (Weiss *et al.,* 2011). For example, presence of *Wolbachia* in mosquitoes, *Drosophila*, silkworms, and some parasitoid wasps upregulates to various levels, the transcription of genes involved in several immune pathways (Eleftherianos *et al.,* 2013; Thomas *et al.,* 2011). *Wolbachia* is presumed to facilitate its own persistence and maintain its close association with the host insect by modulating the immune responses of that host (Kremer *et al.,* 2012). The *Wolbachia*-modulated immune effectors include cecropins, defensins, thioester-containing proteins (TEPs), C-type lectins (CTLs), reactive oxygen species (ROS), relish 1 (REL1), Spätzle 1A (Spz1A), and attacins (Bian *et al.,* 2010; Pan *et al.,* 2012). Introduction of some *Drosophila*-derived *Wolbachia* strains into *Aedes aegypti* induced expression of immune effector genes in the mosquito, and allowed *Wolbachia* replication (Rancès *et al.,* 2012). The authors reported that the the presence of *Wolbachia* interfered with the pathogenesis and/or dissemination of dengue virus (DENV) in these transfected *Ae. aegypti*. Another study demonstrated that resident *Wolbachia* improved refractoriness of *Culex quinquefasciatus* and *D. melanogaster* to West Nile virus (WNV) infections by significantly reducing the viral titers and transmission during feeding compared to their *Wolbachia*-free

counterparts (Glaser and Meola, 2010). In addition to its direct influence on the insect's immunity, *Wolbachia* could also indirectly impact the viral titers in the host by for example, competing for cellular space and resources (Moreira *et al.,* 2009), which could assist the host's immune system to suppress replication of progeny viruses. On the other hand, *Wolbachia* can also lead to increased virus infections as reported in the African armyworm, *Spodoptera exempta,* infected by *S. exempta* nucleopolyhedrovirus (SpexNPV) (Graham *et al.,* 2012).

In tsetse flies the absence (or low densities) of *Wolbachia* is thought to contribute to the expression of SGH symptoms in laboratory-bred *G. pallidipes* colonies as compared with other *Glossina* species that usually do not show overt SGH symptoms (Boucias *et al.,* 2013). Although up to 100% of the different tsetse species in laboratory colonies can be *Wolbachia*-infected, the prevalence of infection significantly varies amongst wild tsetse populations. Doudoumis *et al.* (2012) observed that *Wolbachia* prevalence varied among different tsetse species; *G. m. morsitans* in the laboratory (100%) and in wild (90-100%), 100% in laboratory *G. m. centralis,* 52-100% in wild *G. austeni*, 2-41% in wild *G. brevipalpis*, 0.3% and 0-8.5% in laboratory and wild *G. pallidipes* respectively, and 0% and 0-8.3% in laboratory and wild *G. p. gambiensis* respectively. *Wolbachia* was not detected in wild and laboratory populations of *G. p. palpalis*, *G. f. fuscipes* and *G. tachinoides*.

Although the correlation between *Wolbachia* densities and the SGH incidence is yet to be experimentally demonstrated, data obtained from various studies on laboratory-bred and field-collected *Glossina* spp. implicate a species-specific relationship between the occurrence of SGH and *Wolbachia* infections. For instance, the high *Wolbachia* prevalence in various laboratory-bred colonies of *G. m. morsitans* and *G. m. centralis* (100%) and *G. brevipalpis* (41.2%) (Doudoumis *et al.,* 2012) may be linked to the absence of overt SGH symptoms in these three *Glossina* species. However, this relationship differs from some laboratory-bred *Glossina* species (e.g. *G. palpalis*, *G. p. gambiensis* and *G. f. fuscipes*) that have low or undetectable *Wolbachia* infections (Doudoumis *et al.,* 2013; 2012) but no evidence for the occurrence of overt SGH. However, despite the absence of diagnostic SGH symptoms, the above-mentioned species are susceptible to various degrees of GpSGHV infections either in nature or by artificial virus injections (Demirbas-Uzel

et al., 2018; Kariithi *et al.,* 2013b). It is only in laboratory-bred and wild *G. pallidipes* that appear to be either *Wolbachia*-free or harbor low densities of this symbiont, and are known to exhibit high prevalence of GpSGHV (up to 100%) and overt SGH symptoms and (Boucias *et al.,* 2013; Doudoumis *et al.,* 2012).

Wigglesworthia does not directly provide pathogen resistance to its tsetse host, but it is required for larval maturation and development and for proper functioning of the immune system in adult tsetse flies (Weiss *et al.,* 2012; 2011). On the other hand, the role of *Sodalis* in tsetse is largely unclear, but certain *Sodalis* genotypes are postulated to enhance both the tsetse's susceptibility to trypanosome infections and its innate vectorial competency for transmission of the parasites (Cirimotich *et al.,* 2011). Moreover, *Sodalis* produces many enzymes that impact various host metabolic and biosynthetic processes such as nutrient uptake, and cellular transport (Soumana *et al.,* 2014a). This information implies that *Sodalis* could indirectly influence the outcome of virus infection. Notably, antibiotic-mediated suppression of the *Wigglesworthia* and *Sodalis* in *G. pallidipes* reduced vertical transmission of GpSGHV and inhibited development of overt SGH symptoms in the F_1 progeny. However in that study, the laboratory-bred *G. pallidipes* flies used did not have detectable *Wolbachia*, thus excluding its effects on GpSGHV pathogenesis and transmission (Boucias *et al.,* 2013). The role of the recently detected *Spiroplasma* in tsetse fly during GpSGHV infections requires further investigations.

Impacts of symbiont deficiency in housefly susceptibility to MdSGHV infections

Despite the absence (or presence of low titres) of symbionts in the housefly, this insect harbors highly diverse non-symbiotic microbiota that are environmentally acquired and that vary significantly between individuals (Bahrndorff *et al.,* 2017). Due to the symbiont-mediated reduction of pathogen proliferation in dipterans, it is tempting to hypothesize that the absence or presence of only low densities of symbionts in housefly may result in the occurrence of only symptomatic MdSGHV infections and may explain the high virus prevalence amongst housefly populations.

The 'arms-race' between SGHVs and their dipteran hosts

It is well known that evolution favors hosts that develop strategies to avoid or limit pathogen infections, as well as pathogens that develop effective mechanisms to

modulate or evade the host's immune defenses. This 'arms-race' may result in for instance a stable but dynamic equilibrium (homeostasis) between a virus and its host (Ghazal *et al.,* 2000a), whereby the virus does not significantly compromise the host's reproductive capacities, nor does the host's immune system completely block production of virus progeny. In addition to the host immune pathways, non-immune responses such as vitamin and hormonal signalling pathways, may also be involved in the establishment of the virus-host homeostasis at the cellular level by influencing the transcription and replication cycle of the virus (Ghazal *et al.,* 2000b; Kotzamanis *et al.,* 2015).

The first line of insect's defense is the physical barrier provided by the external cuticle as well as maintaining a low pH and secretion of various antimicrobial peptides (AMPs) in the gut that may block the pathogen's ingress/replication (Lemaitre and Hoffmann, 2007; Tzou *et al.,* 2002; 2000). When breaking these defense barriers, pathogens induce both cellular and humoral innate defenses within the insect's body (Lemaitre and Hoffmann, 2007). Many of the immune genes in these pathways were found to be significantly upregulated during SGHVs infections (Kariithi *et al.,* 2011; 2016; 2017b).

The outcome of virus infection, whether acute or chronic, depends on the balance between the viral replication and the efficiency of the host's immune responses. As a counter-defense, large dsDNA viruses appear to possess many genes that can manipulate host immune responses (Chaston and Lidbury, 2001). Unlike MdSGHV, GpSGHV appears to have recruited into its genome several genes from ancestral hosts during its evolution. The inheritance of host genes by the GpSGHV implies a long evolutionary relationship with the tsetse flies, allowing multiple changes for horizontal gene transfer, leading to a large viral genome (190kpb). The uptake of these genes potentially accounts for its dimorphic life style. On the other hand, the relatively smaller MdSGHV genome (120 kb) probably indicates that the housefly virus has acquired less cellular genes than GpSGHV. The GpSGHV host-derived genes encode proteins that may assist in evading the host's immune system (Kariithi *et al.,* 2013c) to ensure a conducive cellular environment for virus replication (Raftery *et al.,* 2000; Tidona and Darai, 2000). These "acquired" viral proteins are homologs to host/cellular immunity-related protein and could function by mimicking

or interfering with the immune functions of their cellular homologs, a phenomenon that may have been adopted especially by DNA viruses (Chaston and Lidbury, 2001). Notably, the homologies between the viral and cellular proteins may either be throughout the entire amino acid sequences, or only in the functional domains. **Table 1** provides details of the 14 GpSGHV-encoded proteins that are homologous to known cellular genes, of which only two have limited similarity (~20%) with MdSGHV, i.e. lecithine cholesterol acetyltransferase and glutathione S-transferase (Abd-Alla *et al.,* 2008; 2016; Garcia-Maruniak *et al.,* 2008). Nine of the 14 genes have been confirmed to be both transcribed (by RNASeq) and translated (mass spectrometry), implying that they are likely functional in GpSGHV infections (Abd-Alla *et al.,* 2016). Whether these cell-derived genes mentioned have any roles in the evolution of GpSGHV is yet to be elucidated.

Insect immunity consists of three main pillars, including the humoral immune response composed of the Toll and immune deficiency (IMD) pathways (Tanji *et al.,* 2007). The second pillar consists of cellular responses (e.g. phagocytosis) that result in pathogen phagocytosis and melanization (Strand, 2008), as well as other pathways such as c-Jun N-terminal kinase (JNK) and the Janus kinase/signal transducers and activators of transcription (JAK/STAT), and the third pillar consists of the RNA interference (RNAi). The immune genes under these three pillars can be broadly grouped in six functional categories, i.e. recognition, signaling, effectors, modulators, melanization, and RNAi.

Table 1: SGHV-encoded orthologs of cellular genes: The protein families shown in this table were restricted to those that showed significant domain structural conservations. The proteins listed here have been described during the reporting of the genome sequences of the SGHVs (Abd-Alla *et al.,* 2008; 2016; Garcia-Maruniak *et al.,* 2008).

Protein Name	GpSGHV (ORF No.)		Location in virus particle	MdSGHV (ORF No.)	Homology or description
	GpSGHV-Eth	GpSGHV-Uga			
Lecithine-cholesterol acyltransferase* (T,P)	5	5	ICSVP £	46	*Pseudomonas* sp.
D-3-phosphoglycerate dehydrogenase* (T,P)	6	7	Tegument	-	*Clostridium ultunense*
MAL7P1.132* (T,P)	8	9	ICSVP	-	*Plasmodium falciparum* Str. 3D7
UDP-glucose-6 dehydrogenase (T)	13	16	Unknown	-	*Pseudobutyrivibrio ruminis*
NADH ubiquinone oxidoreductase (T)	30	29	Virion protein	-	Styphylococcal AgrD protein
Maltodextrin glycosyltransferase* (T, P)	39	38	Tegument	-	RGD-domain containing protein
Glutathione S-transferase* (T, P)	48	46	Tegument	84	Pre-mRNA splicing factor
Cellular protein CBG22662* (T, P)	49	47	Tegument	-	*Coenorhabditis briggsae*
Rhoptry protein kinase (T)	58	57	-	-	*Plasmodium yolei* Str. 17XNL
Signaling mucin HKR1	64	-	-	-	*Xenopus* (*Silurana*) *tropicalis*
RpoD protein (T)	66	59	-	-	*Plasmodium falciparum*
ECF transporter* (T, P)	75	68	Envelop	-	-
Cellular protein PY00593* (T, P)	124	113	Nucleocapsid	NaN	*Plasmodium yolei* Str. 17XNL
Tail length tape-measure* (T, P)	149	134	ICSVP	-	Oenococcus phage phi9805

*Expression confirmed by transcriptomics (T) and proteomics (P). Unmarked genes have no detectable transcripts or peptides; £ These proteins do not have specific localization and were designated as 'infected cell-specific viral proteins (ICSVP) (Abd-Alla *et al.,* 2016).

Using *D. melanogaster* and *An. gambiae* as the references, analysis of these immune genes in *G. pallidipes* and *M. domestica* showed species-specific and extensive expansion of the pathogen recognition genes (**Table 2**). For example, in *M. domestica*, calcium-dependent lectins (CTLs; *n*=37) and thioester-containing proteins (TEPs; *n*=22) have expanded when compared to *Drosophila* (34 CTLs; 10 TEPs), *G. pallidipes* (17 CTLs; 7 TEPs) and *An. gambiae* (25 CTLs; 13 TEPs) (**Table 2**). The expansion of CTLs and TEPs in *M. domestica* implicates gene duplication driven by selective evolutionary pressures. The TEPs are characterized are phagocytic opsonins in many species, from insects to mammals (Janeway, 1989; Kim *et al.,* 2010). The significant expansion of TEPs in *M. domestica* is probably an evolutionary necessity to enable this insect to deal with the large number of diets or habitat-associated microbes. Compared to the other three insects, *M. domestica* seems to have acquired two additional homologs of the Down-syndrome adhesion molecule-1 (Dscam1), an insect opsonin fitted to cope with a broad range of pathogens (Kim *et al.,* 2010).

The core immune signaling genes (Toll, Imd, JAK/STAT and JNK pathways) shows single-copy orthologs with similar divergence levels across the four dipterans (*D. melanogaster*, *An. gambiae*, *M. domestica* and *G. pallidipes*; **Table 2**). Although these pathways are traditionally thought to protect insects against infections by bacteria, fungi and parasites, evidence suggest that these pathways play a significant role in the defense against many viruses in both mosquitoes and *Drosophila* (Lemaitre and Hoffmann, 2007; Prasad *et al.,* 2013). Despite the overall conservation of the signaling immune genes in the four insects analyzed here, there are a few cases of gene losses and gains. For example, compared to the four Spätzle and one Dorsal protein homologs in *M. domestica*, *G. pallidipes* has expanded the original two genes to seven *spätzle* and eight *dorsal* genes. Spätzle is an insect hemolymph cytokine, which in the moth, *Manduca sexta*, functions as a ligand that stimulates immune response to kill invading pathogens (An *et al.,* 2010). In *Drosophila*, Spätzle initiates a signaling cascade that terminate upon the release of Dorsal from the protein Cactus to activate genes that are important for dorsal-ventral patterning in early embryonic development (An *et al.,* 2010).

Table 2: Major immunity genes in *M. domestica* and *G. pallidipes*: The immune genes described for the model insect, *D. melanogaster* and *An. gambiae* were obtained from the ImmunoDB (Waterhouse *et al.,* 2007). The pathway for the putative immune-related proteins in *G. pallidipes* and *M. domestica* were verified by BLASTp searches at the Insect Innate Immunity Database (IIID) (≤10-6; bit score >75) (Brucker *et al.,* 2012). The pathways shown in this table have been reviewed by Kingsolver *et al.* (2013). Abbreviations: D. mel; *D. melanogaster,* An. gam; *An. gambiae*, M. dom; *M. domestica*, G. pal; *G. pallidipes.*

Description of the functions and pathways of immune-related proteins in *D. mel*			**Numbers of homologs**			
Immune function	**Key pathway**	**Protein name/sub-family**	***D. mel***	***An. gam***	***M. dom***	***G. pal***
Pathogen recognition	Lectin	Calcium-dependent (C-type) lectins (CTLs)	34	25	37	17
	Phagocytic	Down Syndrome cell adhesion molecule-1 (Dscam1)	1	1	3	1
		Pathogen pattern-recognition receptor Eater (Eater)	1	1	-	1
		Thioester-containing proteins (TEPs)	10	13	22	7
Signaling	Toll	Spätzle-like proteins (Spätzle)	6	6	4	7
		Toll receptors (Tolls)	9	10	7	6
		MyD88	1	1	1	1
		Tube	1	1	1	1
		Pelle	1	1	1	1
		TNF-receptor-associated factor-like (TRAF)	1	1	1	2
		Cactus	1	1	1	2

Description of the functions and pathways of immune-related proteins in *D. mel*				Numbers of homologs		
Immune function	**Key pathway**	**Protein name/sub-family**	***D. mel***	***An. gam***	***M. dom***	***G. pal***
		Dorsal	2	-	1	8
Signaling	Imd	Immune deficiency (Imd)	1	1	1	-
		TGF-beta activated kinase 1 (Tak1)	1	1	1	1
		Kenny	1	1	-	1
		Inhibitor of nuclear factor kB kinase β (IKKb/ird5)	1	1	1	1
		Fas-associated death domain (FADD)	1	1	1	1
		Poor Imd response upon knock-in (PIRK)	1	-	1	-
		Caspar (Casp)	1	1	1	1
		TAK1-associated binding protein 2 (Tab2)	1	1	1	1
		Relish (Rel)	1	2	1	2
Signaling	JAK/STAT	Domeless	1	1	-	1
		Janus kinase (Hopscotch)	1	1	1	1
		Signal transducer and activator of transcription (Stat92E)	1	2	1	2
Signaling	JNK	Jun kinase (JNK)/basket	1	1	1	1
		Dual-specificity MAPK hemipterous (hep)	1	1	1	1
		Jun-related antigen (Jra/Jun)	1	1	-	2

Description of the functions and pathways of immune-related proteins in *D. mel*			Numbers of homologs			
Immune function	**Key pathway**	**Protein name/sub-family**	***D. mel***	***An. gam***	***M. dom***	***G. pal***
Effectors	AMP	Antimicrobial peptides (AMPs)	21	11	21	4
		Lysozyme	17	8	29	5
		Nitric oxide synthase (NOS)	1	1	1	1
Modulators	Exocytic	CLIP-Domain Serine Proteases (CLIPs)	47	55	132	72
	Proteolytic	Serine protease inhibitors (serpins)	29	21	35	14
Melanization/ Encapsulation	Humoral	Prophenoloxidase (PPO)	3	9	25	8
RNAi response	Small RNA Regulatory Pathways (SRRPs)	Argonaute (Ago)	3	3	2	3
		Armitage (Armi)	1	1	1	1
		Aubergine (Aub)	1	1	1	1
		Dicer (Dcr)	2	2	2	1
		Drosha	1	1	1	1
		Loquacious (Loqs)	1	1	2	1
		Partner of Drosha (Pasha)	1	1	-	1
		P-element induced wimpy testis (Piwi)	1	1	1	1
		R2D2	1	1	1	2
		Spindle-E (Spn-E) or Homeless	1	1	2	1
		Tudor staphylococcal nuclease (Tudor-SN)	1	1	1	1

The effector and modulator gene categories seemed significantly diverged across the analysed species, except single copies of nitric oxide synthase (NOS) in each of the species. There seems to be species-specific and extensive expansion of the modulator genes, mainly in the CLIP-domain serine proteases (CLIPs) and serpins in *M. domestica* compared to *G. pallidipes* (**Table 2**). In insect hemolymph, CLIPs proteolytically activate Spätzle (involved in signalling) and other proteins (Kanost and Jiang, 2015), thereby serving as a mediator of insect immunity against invading pathogens. Perhaps the nutrition and ecology of the housefly has evolutionarily driven the selection of a large number of effector and modulator genes to counter potential pathogens that are likely to be acquired from the environment. In terms of the humoral responses, the most notable expansion in the housefly was the prophenoloxidase (PPO) gene family (n=25) compared to the significantly lower numbers of PPOs in *Drosophila* (n=3), *An. gambiae* (n=9), and *G. pallidipes* (n=8) (**Table 2**). In many arthropods, the PPO cascade is not only evolutionary conserved, but it is the primary extracellular pathway for wound healing and melanization of infecting pathogens (Christensen *et al.,* 2005), which may be important for the ecology of the housefly. Further, the by-products of the PPO pathway have been reported to have antiviral effects in some viruses such as the baculoviruses (Popham *et al.,* 2004; Shelby and Popham, 2006; Zhao *et al.,* 2011), sindbis virus (SINV) (Tamang *et al.,* 2004) and Semliki Forest virus (SFV) (Rodriguez-Andres *et al.,* 2012). The enrichment of the PPO pathway in *M. domestica* requires further investigations of the extent to which the pathway is engaged in the pathogenesis of SGHVs.

The host RNAi machinery and the SGHV's evolutionary mechanisms

RNAi is recognized as a conserved anti-viral defense mechanism in insects that is not only active against RNA viruses but also against several groups of large dsDNA insect viruses (ascovirus, baculovirus, iridovirus and nudivirus) (Bronkhorst *et al.,* 2012; Burand and Hunter, 2013). RNAi involves short interfering RNAs (siRNAs; derived from exogenous dsRNA) and microRNAs (miRNAs; encoded by the host or viral genome), which interfere with gene expression by targeting specific mRNAs (Li and Ding, 2006). Three of the RNAi key genes (*Ago-2*, *Dcr-2* and *R2D2)* are amongst the fastest evolving and in *Drosophila*, these genes are subject to a great

positive selection and selective sweeps (Obbard and Dudas, 2014; Obbard *et al.,* 2011). Compared to *Drosophila*, *An. gambiae* and *G. pallidipes*, *M. domestica* appears to have lost *Ago-1* and *Pasha*, but has acquired extra copies of *Loquacious* (*Loqs*) and *Spindle-E/Homeless* (*Spn-E/hls*), and expanded *Ago-2* (**Table 2**). On the other hand, *G. pallidipes* appears to have duplicated *R2D2* (**Table 2**; **Figure 1A**). In *Drosophila*, *Spn-E* (together with Piwi and Aubergine) is involved in the RNAi-mediated (via the piRNA pathway) silencing of heterochromatin (Pal-Bhadra *et al.,* 2004). Specifically, *Spn-E* is required for activation of RNAi-mediated regulation of maternal mRNAs during oogenesis in *Drosophila* (Kennerdell *et al.,* 2002), and in defense against transposable elements (Obbard *et al.,* 2009), but its anti-viral roles are yet to be defined. One of the *M. domestica Spn-E* duplicates contained all the three signature domains found in the *Drosophila Spn-E*; the *Spn-E* homologs from mosquito and tsetse lacked the catalytic tetrad DExH box/Tudor domain (**Figure 1C**). However, absence of the catalytic tetrad is not unique since only a subset of the family members possesses cleavage activity (Meister, 2013). The three cofactors of Dcr and Drosha (Pasha, R2D2 and Loqs) that are required in the first step of the RNAi pathway (i.e. generation of dsRNAs) (Haac *et al.,* 2015; Hammond, 2005) contained the functional domains. Further, of the four AGO proteins (AGO 1-4), AGO-2 is singly capable of executing the ultimate aim of the RNAi pathway (Hammond, 2005), which partially makes up for the loss of *Ago-1* gene in *M. domestica* trivial. Notably, the *M. domestica* Dcr-2 homolog lacks the dsRNA-binding domain (**Figure 1B**), thus raising the question of what effects this has in the functionality of the protein. This species-specific expansions and/or losses of the RNAi genes may have significant implications on the functionality of the pathway in tsetse and the housefly.

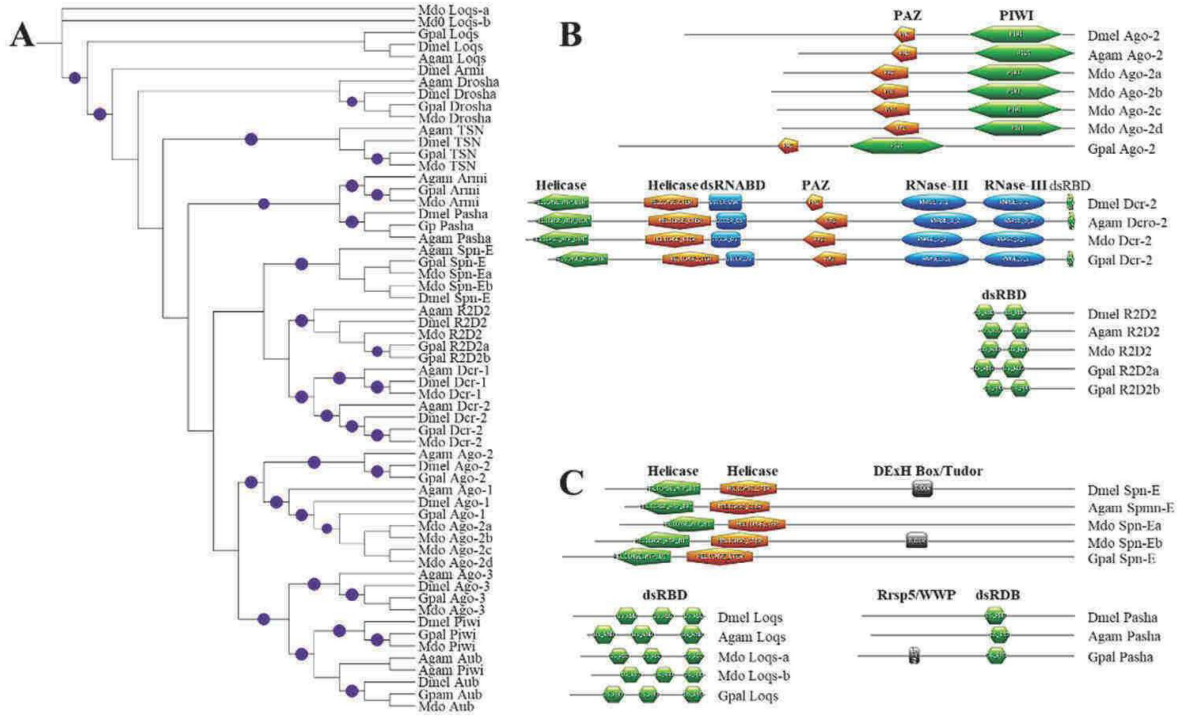


Figure 1: Phylogenetic analysis of the key RNAi pathway proteins in dipterans: (A) Clustering of the housefly (Mdo) and tsetse fly (Gpal) RNAi proteins with their homologs in the fruit fly (Dmel) and the malaria mosquito (Agam). Purple circles indicate bootstrap support of >80%. (B) Domains of the three key RNAi pathway proteins, Ago-2, Dcr-2, R2D2, and (C) three of their main cofactors Spn-E, Pasha and Loqs.

The RNAi machinery has been demonstrated and proven to be functional against dsDNA viruses in flies, not only due to the presence of the key genes of RNAi pathway, but also because of the fact that flies with loss-of-function for some of these genes (e.g. in *Drosophila*) are reported to be highly susceptible to viral infections (Bronkhorst *et al.,* 2012; Galiana-Arnoux *et al.,* 2006; van Rij *et al.,* 2006; Wang *et al.,* 2006). In *G. pallidipes,* both *Dcr-2* and *Ago-2* genes were found to be significantly up-regulated in asymptomatically infected flies but down-regulated in symptomatically infected flies (see **Chapter 5** of this thesis). In the case of houseflies, two isoforms of *Ago-2* and *Dcr-1* genes were up-regulated in MdSGHV infected females compared to their uninfected conspecifics (Kariithi *et al.,* 2017b). Together, these data imply that the RNAi pathway may actively be involved in the dynamics of the SGHV-*Musca*/*Glossina* system.

To favor their own replication some viruses manipulate the host's transcriptome via specific downregulation or upregulation of specific host miRNAs (Asgari, 2014; Lucas and Raikhel, 2013). In the case of tsetse flies, recent data suggests that GpSGHV alters the host miRNA profile in *G. pallidipes*, possibly indicating a functional importance of host miRNAs during GpSGHV infections (see **Chapter 6** of this thesis). DNA virus-encoded miRNAs are thought to act in a similar manner to host miRNAs by inhibiting (e.g. in herpesvirus saimiri; HVS), boosting (e.g. in Epstein-Barr virus; EBV), or hijacking (e.g. in human cytomegalovirus; HCV) host miRNAs to reshape the cellular environment to the benefit of virus replication (Guo and Steitz, 2014). More importantly, due to their ability to weaken the host's immune responses, the virus-encoded miRNAs have been implicated in the reactivation of some viruses from latency (Cullen, 2009), whereby only minimal numbers of genes are expressed to evade the host immune system (Gottwein *et al.,* 2007; Pfeffer *et al.,* 2005). Examples of viral-encoded miRNAs involved in the latent/persistent virus infections include miR-H2-3p and miR-H6 encoded by herpes simplex virus 1; HSV-1, miR-UL112-1 encoded by HCV and miR-K5 encoded by Kaposi's sarcoma-associated herpesvirus; KSV (Boss and Renne, 2010). In some cases, certain virus-encoded miRNAs may repress expression of viral genes to control the latent and lytic infection stages (Asgari, 2015; Lieberman, 2016; Lu *et al.,* 2010).

In a genome-wide screen on the MdSGHV and GpSGHV, Garcia-Maruniak *et al.* (2009) identified six and seven miRNAs in these genomes respectively. Although the presence of miRNA encoding sequences in the SGHVs were predicted *in silico*, it is likely that these miRNAs, are functional especially in the case of GpSGHV which can switch from asymptomatic to symptomatic infections (Boucias *et al.,* 2013). Besides, the same GpSGHV-encoded miRNAs were identified by next generation sequencing (NGS) of small RNAs in symptomatically infected *G. pallidipes* (see **Chapter 6**). In this respect, GpSGHV is more likely to utilize miRNAs than MdSGHV, since most of the known virus-encoded miRNAs are involved in facilitating viral latency/persistence by prolonging lifespan of the infected cells, regulating virus and/or host genes expression and limiting symptomatic infections (Grundhoff and Sullivan, 2011).

Role of apoptosis and SGHVs escape strategies

Viruses from different families can induce apoptosis, a biochemically and genetically-regulated cell death process (Everett and McFadden, 1999). Apoptosis is important for host's innate immune response, which aim to limit the time for virus replication and dissemination (Nainu *et al.,* 2015). Viruses can disrupt the balance between the synthesis and degradation of apoptosis inhibitors and activators/initiators in order to facilitate virus replication and dissemination (e.g. influenza and SINV) (Teodoro and Branton, 1997). In such cases, the virus-induced apoptotic response may result in phagocytosis of infected cells by the neighboring (uninfected) cells thus facilitating virus dissemination within the host without eliciting an immune response (Teodoro and Branton, 1997).

Apoptosis is initiated and executed by *c*ysteine-dependent *as*partate-specific *p*rote*ases* (caspases) upon activation by apoptotic stimuli. Activated caspases cleave their target substrates such as protein kinases, signal transduction proteins, chromatin modifiers, DNA repair enzymes, inhibitory endonucleases, etc.) (Cooper *et al.,* 2009). The model organism, *Drosophila*, encodes seven caspases (**Table 3**) (Hay and Guo, 2006; Lamkanfi *et al.,* 2002). These include three apoptosis initiators – [death regulator Nedd2-like caspase (DRONC/caspase-9), death-related ced-3/Nedd2 (DREDD/Caspase-8), and serine/Threonine-rich caspase-A (STRICA)], in addition to four effectors – [death associated molecule related to Mch2 (DAMM), *Drosophila* interleukin-1β-converting enzyme (DrICE), death executioner caspase-related to Apopain/Yama (DECAY/Caspase-3/7), and death caspase protein 1 (DCP-1)] (Li and Yuan, 2008). In addition, several caspases have also been reported in other insects such as mosquitoes, *Ae. aegypti* and *An. gambiae* (Cooper *et al.,* 2009). Apoptosis in *Drosophila* is activated by DRONC, and can be inhibited by *Drosophila i*nhibitor of *a*poptosis *p*rotein 1 (DIAP1). DIAP1 binds directly to DRONC using its domain to promote ubiquitination and hence inhibition of DRONC (Hay and Guo, 2006). DREDD is essential for activation of the innate immune responses via cleavage of Relish, a NF-κB family member of the Imd pathway (Foley and O'Farrell, 2004). In *Drosophila*, DAMM was upregulated upon *Drosophila* C virus (DCV) infections (Dostert *et al.,* 2005) suggesting the involvement of apoptosis in DCV infection, perhaps via the JAK/STAT pathway.

Table 3: Apoptotic and/or immunity-related roles of *Drosophila* caspases.

***Drosophila* caspase**	**GenBank Acc. No.**	**Apoptotic and, (or anti-viral roles in insects**	**Refs.**
Apoptosis Initiators			
Death regulator Nedd2-like caspase (DRONC); Caspase-9 homolog	NP_524017.1	Ecdysone-induced (developmental and stress-induced apoptosis);	(Lee *et al.,* 1998; Lietze *et al.,* 2009)
Death-related ced-3/Nedd2 (DREDD) or DCP-2; Caspase-8 homolog	NP_477249.3	Essential for activation of innate immune signaling (activates Relish of the Imd pathway)	(Everett and McFadden, 1999; Manji *et al.,* 1997)
Serine/Threonine-rich caspase-A (STRICA) or Downstream regulatory element-antagonist modulator (DREAM)	NP_610193.1	Together with DRONC, STRICA activates DCP-1 and DRICE	(Huang *et al.,* 2000)
Apoptosis Effectors			
Death associated molecule related to Mch2 (DAMM)	AAF82437.1	Upregulated in DCV-infected *Drosophila*	(Dostert *et al.,* 2005)
Death caspase protein 1 (DCP-1)	NP_476974.1	Essential for germ-line apoptosis in mid-oogenesis; cleaves P35	(Huang *et al.,* 2000)
Death executioner caspase-related to Apopain/Yama (DECAY); Caspase-3, (7 homolog	NP_477462.1	Involved in developmental apoptosis and immunity; (upregulated in DENV-infected mosquito)	(Everett and McFadden, 1999; Manji *et al.,* 1997; Wang *et al.,* 1999)
Drosophila interleukin-1β-converting enzyme (DrICE)	NP_524551.2	Cysteine protease that inhibits baculovirus P35 and *Drosophila* lamin DmO	(Best, 2008)

To determine the presence and potential roles of the caspases in *M. domestica* and *Glossina* spp, the seven *Drosophila* caspase sequences were used as references for homolog searches in *M. domestica* and *Glossina* spp found in the database, followed by phylogenetic analyses using the conserved caspase domains (**Figure 2**). The three initiator caspases (DREDD, DRONC and STRICA) clustered into distinct clades in the three dipterans (**Figure 2A**).

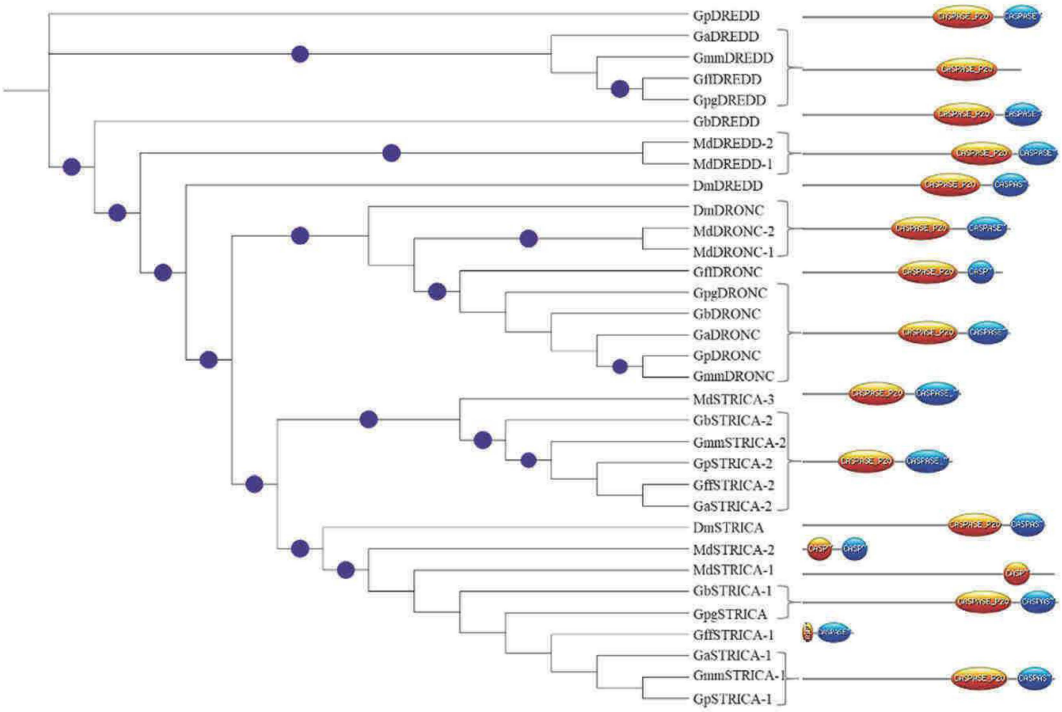


Figure 2A. Phylogenetic analysis of the initiator caspases (DRONC, DREDD and STRICA): The three initiator caspases showed clear clustering across *Drosophila* (Dm), housefly (Md) and the *Glossina* species (Gp, Gmm, Ga, Gff, Gpg and Gb). Shown are the caspase prodomains of variable lengths, followed by p20 (orange) and p10 (blue) units that contain essential amino acid residues required for substrate recognition and catalysis. The prodomains were excluded during the phylogenetic reconstructions. Purple circles indicate bootstrap support of >80%.

Glossina species had single copies of DREDD and DRONC, but the two initiator caspases were duplicated in *M. domestica*. STRICA was duplicated in both *Glossina* spp., and *M. domestica*. For the effector caspases, *Glossina* spp and *M. domestica* DAMM delineated into distinct orthologous clusters (**Figure 2B**). *Glossina* DECAY caspases formed a distinct cluster closely related to the DCP-1 cluster in which *M. domestica* DECAY segregated. The *M. domestica* DECAY was apparently

duplicated. All the *M. domestica* DCP-1 segregated together with DrICE caspases (**Figure 2B**).

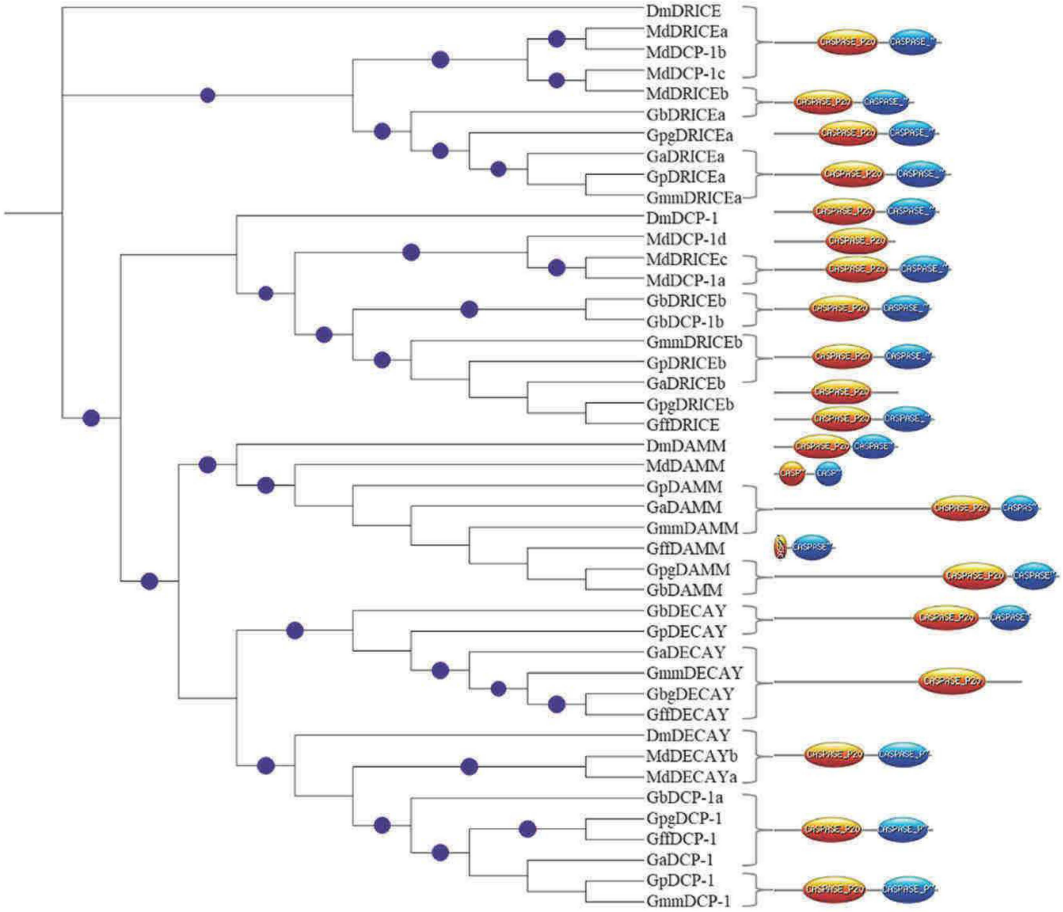


Figure 2B. Phylogenetic analysis of the effector caspases (DAMM, DrICE, DECAY and DCP-1): The effector caspase DAMM, and to a large extend DECAY, segregated clearly, but not for DRICE and DCP-1. Shown are the caspase prodomains of *Drosophila* (Dm), housefly (Md) and the *Glossina* species (Gp, Gmm, Ga, Gff, Gpg and Gb) of variable lengths, followed by p20 (orange) and p10 (blue) units that contain essential amino acid residues required for substrate recognition and catalysis. The prodomains were excluded during the phylogenetic reconstructions. Purple circles indicate bootstrap support of >80%.

Together, the analyses showed widespread duplication of *M. domestica* caspases as compared to their *Glossina* spp. homologs. It has been reported that the caspase-8 (DREDD) homolog which isessential for activation of innate immune response, the caspase-3 (DECAY) homolog, an apoptosis effector, and Relish, were upregulated

in MdSGHV-infected females houseflies compared to control houseflies (Kariithi *et al.,* 2017b).

To prolong infected cell viability and facilitate virus replication, viruses have evolutionary devised multiple mechanisms to inhibit apoptosis by mimicking key regulators of apoptosis (Benedict *et al.,* 2002). There are four main protein families of viral inhibitors of apoptosis, i.e. serpins, baculovirus P35 (and P49; P33), inhibitors of apoptosis (IAPs), and viral FLICE-inhibitory proteins. Three of these are well-known in large dsDNA insect viruses, i.e. P35 of the baculovirus *Autographa californica* MNPV (AcMNPV), its P49 homolog in *Cydia pomonella* granulovirus (CpGV), and the IAPs present in *Orgia pseudotsugata* MNPV, and CpGV (Clem, 2001). P35 is known to inhibit the evolutionary conserved interleukin-1β–converting enzyme (ICE)/ICE-like proteases (Clarke and Clem, 2003) which lead to increased AcMNPV titers to allow successful virus infection in the host (Mehrabadi *et al.,* 2015). Besides, AcMNPV with p35 deletion was reported to fail in inhibiting apoptosis (Clarke and Clem, 2003). Although neither MdSGHV nor GpSGHV encodes P35 or P49 homologs, MdSGHV encodes a single copy of IAP (MdSGHV078) (Scott *et al.,* 2014) whose transcripts are moderately enriched in MdSGHV infected flies (Kariithi *et al.,* 2017b). The IAPs prevent apoptosis by blocking caspase activation (Yang and Li, 2000) via ubiquitination of host's pro-apoptotic proteins or via direct interactions with caspases (Shi, 2004), similarly to the p35 that inhibits the downstream process of apoptosis.

The phylogenetic analysis of the MdSGHV IAP compared to the reported homologs in OpMNPV, *Spodoptera exigua* MNPV (SeMNPV), *Epiphyas postvittana* NPV (EppoNPV), *Bombyx mori* NPV (BmNPV), *Buzura suppressaria* NPV (BusuNPV), CpGV, *Chilo iridescent virus* (CIV) and African swine fever virus (AsFV) showed that MdSGHV IAP clustered with IAPs from AsFV and CIV (**Figure 3A**). Compared to the domains of the IAPs from the other viruses, the MdSGHV, AsFV and CIV IAPs contained one baculovirus inhibitor repeat (BIR) domain and an additional zinc binding fold RING domain (**Table 4**; and **Figure 3B**). Although cellular IAPs contain up to three tandem copies of the BIR domain, viral IAP proteins contain one or two BIRs; a single BIR domain is sufficient for suppression of apoptosis (Best, 2008).

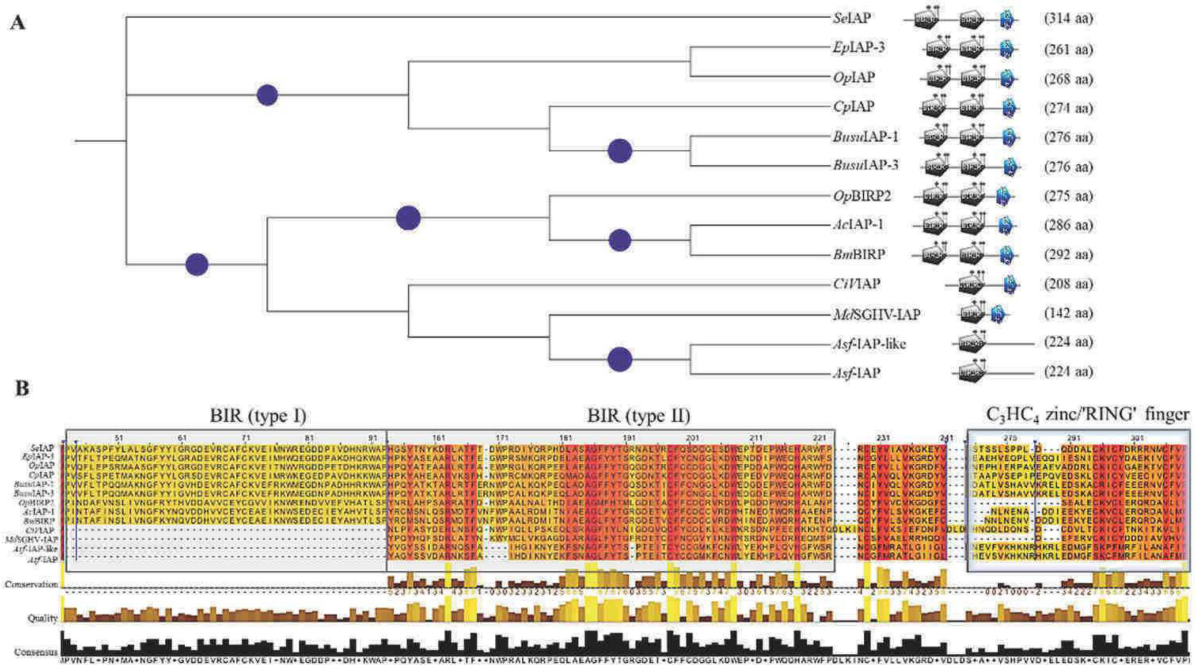


Figure 3. Phylogenetic analysis of IAP homologs from several viruses: (A) Phylogenetic clustering of MdSGHV IAP with homologs from *Spodoptera exigua* MNPV (seIAP), *Epiphyas postvittana* MNPV (EpIAP), *Orgia pseudotsugata* NPV (OpIAP/BIRP2, *Cydia pomonella* granulovirus (CpCpIAP), *Buzura suppressaria* NPV (BusuIAP-1/3), *Autographa californica* MNPV (AcIAP), *Bombyx mori* NPV (BmBIRP), *Chilo iridescent* virus (CiVIAP), MdSGHV (MdSGHV-IAP) and African swine fever virus (Asf-IAP/IAP-like). (B) Alignment of the IAP showing the functional baculovirus IAP repeat domains (BIR-1/2), and zinc binding fold. The IAPs from MdSGHV, AsfV and CIV contained a single BIR domain.

In addition to the presence of an *iap* gene in MdSGHV, the housefly genome encodes three pro-apoptotic proteins (Grim, Reaper, and HID) (Scott *et al.,* 2014). These findings imply that during MdSGHV symptomatic infection, apoptosis may occur in the housefly, and that MdSGHV controls progression of apoptosis to its benefit and ensure that the infected cells do not only survive, but also grow and produce progeny virus and perhaps contribute to development of SGH symptoms. In addition to possessing the *iap* gene, it is known that unlike GpSGHV, MdSGHV rapidly multiplies and induces detectable SGH symptoms in 100% of injected flies within 2-3 days post injection (Kariithi *et al.,* 2017a). This rapid replication and possession of *iap* gene by MdSGHV may provide this virus with an additional strategy of protecting itself from the housefly's apoptotic response as compared to GpSGHV in tsetse.

Table 4: Comparison of MdSGHV IAP with homologs reported in other viruses: The analysis revealed clustering of the IAPs from MdSGHV, ASFV and CIV. Compared to the other viruses analyzed, the domain architecture of MdSGHV, ASFV and CIV IAPs contained one BIR domain and an additional RING domain.

Virus	Sequence Name	GenBank Acc. No.	Length (aa)	BIR domain coordinates	RING domain coordinates
*Cydia pomonell*a granulovirus (CpGV)	ORF17 iap-3	AIU36666.1	275	10-74; 111-176	224-269
Orgyia pseudotsugata multiple nucleopolyhedrovirus (OpMNPV)	IAP-3	NP_046191.1	268	21-85; 114-179	217-262
	IAP-1	NP_046197.1	275	27-92; 129-194	223-269
Spodoptera exigua multiple nucleopolyhedrovirus (SeMNPV)	IAP-3	CDG72862.1	314	34-99; 157-222	263-308
Musca domestica SGHV (MdSGHV)	IAP	YP_001883406.1	142	12-77	92-136
Chilo/invertebrate iridescent virus (CIV/IIV-6)	193R (BIRP)	NP_149656.1	208	40-109	159-203
Autographa californica multiple nucleopolyhedrovirus (AcMNPV)	IAP	NP_054056.1	286	32-97; 134-200	234-280
African swine fever virus (ASFV)	IAP	P0C9X4.1	224	32-93	94-224
	IAP-homolog	NP_042727.1	224	32-93	94-224
Buzura suppressaria nucleopolyhedrovirus (BusuNPV)	IAP-1	AAC34373.1	276	15-79; 111-177	225-270
	IAP-3	YP_009001870.1	276	15-79; 111-177	225-270
Epiphyas postvittana nucleopolyhedrovirus (EppoNPV)	IAP-3	NP_203195.1	261	14-78; 104-169	210-255
Bombyx mori nucleopolyhedrovirus (BmNPV)	IAP-1	NP_047432.1	292	32-97; 134-200	240-286

The lack of anti-apoptotic gene homologs in GpSGHV suggests that this virus may have alternative strategies to counteract their host's apoptotic responses. For instance, it is possible that GpSGHV has adopted the asymptomatic infection so that the virus infection is undetectable by the host's immune surveillance and only a few viral genes are expressed.

Conclusions

The findings presented in this chapter suggest that the ecologies and life-history traits of the housefly and tsetse fly (for various tsetse species as discussed in **Chapter 3**) have influenced the coevolution between the host and the particular SGHV and their persistence and transmission strategies. In the case of the GpSGHV, the existence of mixed modes of vertical and horizontal transmission may be evolutionary beneficial to the GpSGHV which could contribute to the generation and maintenance of the virus diversity (haplotypes) as demonstrated in Chapter 4. The possibility of existence of muscid hosts for MdSGHV other than the housefly as reservoirs or alternative hosts may have favoured the selection for horizontal transmission of this virus. Additionally, the highly virulent nature of the MdSGHV (i.e. infects only symptomatically) potentially hints to this virus as a regulating factor for housefly populations in a density-dependent manner. However, more studies are required to investigate the SGHVs infection and transmission dynamics and the roles of these viruses on regulating the housefly and tsetse fly populations. The data in this chapter however left several questions remain unanswered with regard to SGHVs dynamics. For instance, why are GpSGHV-induced epizootics (SGH outbreaks) a rare occurrence in the field, and what are the genetic elements accounting for the differences in the pathogenesis of the two GpSGHV isolates (i.e. GpSGHV-Uga causes <10% SGH prevalence compared to the 85% SGH prevalence caused by the GpSGHV-Eth)? Some of the factors that may influence the outcome of GpSGHV infections such the RNAi and miRNAs in particular have been further discussed in **Chapter 5 and 6,** respectively. In the case of the MdSGHV, how is this virus maintained when the host populations fluctuate to low densities, and does the MdSGHV virulence (symptomatic infections) modulate the community structures of

the houseflies in nature? Lastly, to what extent the tripartite host-SGHV-microbiota interactions influence SGH epizootics?

Acknowledgements

The Joint FAO, (IAEA Division of Nuclear Techniques in Food and Agriculture, Vienna, Austria, and Wageningen University funded this work (Coordinate Research Project No. D42015) and Wageningen University and Research, The Netherlands.

Chapter 3

Nuclear and *Wolbachia*-based multi-marker approach for the rapid and accurate identification of tsetse species

This chapter has been modified from:

Augustinos, A.A., **Meki, I.K.**, Demirbas-Uzel, G., Ouédraogo, G.M.S., Saridaki, A., Tsiamis, G., Parker, A., Abd-Alla, A., Bourtzis, K. Nuclear and *Wolbachia*-based multi-marker approach for the rapid and accurate identification of tsetse species. BMC Microbiology, (*in press*).

Abstract

Tsetse flies (Diptera: Glossinidae) are solely responsible for the transmission of African trypanosomes, the causative agents of sleeping sickness in humans and nagana in livestock. Due to a lack of efficient vaccines and emergence of drug resistance, vector control approaches, such as the sterile insect technique (SIT) as a component of integrated pest management strategies, remain the most effective method to control the disease. SIT is a species-specific approach that requires accurate species identification, which is usually challenging especially for tsetse species that occur in overlapping ecogeographical areas and are nearly morphologically indistinguishable (species complexes and sub-species). Correct species identification is crucial for the initial establishment of a new colony as well as management of pathogen (e.g. viruses) since specific host's genetics may predispose the host to different level of pathogen susceptibility. For instance, the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) which is specifically pathogenic to tsetse flies (discussed in **Chapter 2**), leads to reduced fecundity and infertility in *G. pallidipes* species thereby hindering SIT application by preventing production of sufficient numbers of male flies required for the mass release. In this chapter, different molecular tools that can be applied for the correct identification of different *Glossina* species were evaluated using tsetse samples derived from laboratory colonies, natural populations and museum specimens. The combined use of mitochondrial markers, nuclear markers (including internal transcribed spacer 1 (ITS1) and different microsatellites), and bacterial symbiotic markers (*Wolbachia* infection status), as well as relatively inexpensive techniques such as PCR, agarose gel electrophoresis, and, to some extent, sequencing, provided a rapid, cost effective, and accurate identification of several tsetse species. These molecular tools complement the conventional tools thereby enhancing proper identification and pathogen management for specific or multiple *Glossina* species in laboratory colonies. The tools directly benefit the SIT from the fine resolution of species by use of universalized protocols, which can be easily applied by countries/laboratories with limited resources and expertise.

Introduction

Tsetse flies are responsible for the cyclic transmission of trypanosomes, the causative agents of sleeping sickness or human African trypanosomosis (HAT) in humans and nagana or animal African trypanosomosis (AAT) in livestock (Aksoy, 2011; Krafsur, 2009). There are about 31 tsetse fly species and sub-species within the *Glossina* genus (Diptera: Glossinidae), which are distributed in 37 sub-Saharan African countries. However, only 8-10 of these species are of economic importance (Cecchi *et al.,* 2008). Due to a lack of vaccines against trypanosomes and increasing resistance of the AAT parasites to available drugs (Allsopp, 2009; Geerts *et al.,* 2001), vector control remains the most effective way of managing African trypanosomosis (Schofield and Kabayo, 2008). Some of the vector control strategies that have been applied for the control of trypanosomosis include the use of stationery attractive devices, live bait technique, sequential aerosol technique (SAT), and sterile insect technique (SIT) (Green, 1994; Kgori *et al.,* 2006; Knipling, 1959; Vreysen *et al.,* 2013). The SIT involves production of large numbers of the target insect species in specialized mass rearing facilities, followed by sexual sterilization of the males by irradiation (Robinson, 2005). The sustained and systematic mass release of the sterile males over the target area out-competes the wild male population for mating with wild females. Mating of sterile males with wild females leads to no offspring and subsequent decrease of the targeted population (Abila *et al.,* 2003). SIT is a species-specific and environmentally friendly control method that has been successfully applied for the eradication of a population of *Glossina austeni* from Unguja Island in Zanzibar (Vreysen *et al.,* 2000).

For a successful SIT application, the correct species identification is of critical importance as it enables target insect compatibility with the released sterile males. Not only this, but the initial fly materials required to start up new mass rearing facilities need to be healthy (free of pathogen infections), or if they are infected, the pathogens should be identified and catalogued. Of importance, is the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) that naturally infects tsetse flies and causes salivary gland hypertrophy (SGH) syndrome that has been linked to the reduced fecundity and infertility of tsetse colonies (Abd-Alla *et al.,* 2009a; Ellis and Maudlin, 1987; Odindo, 1982; Otieno *et al.,* 1980). GpSGHV

(*Hytrosaviridae* family) is a large rod-shaped dsDNA virus that replicates in the nucleus of the infected cell (Abd-Alla *et al.,* 2010a). Although this virus has been reported to infect several other wild tsetse species, studies have focused on GpSGHV infections in *G. pallidipes*, which is an ideal model species for studies as it presents both asymptomatic and symptomatic infections (Abd-Alla *et al.,* 2010b; Burtt, 1945; Ellis and Maudlin, 1987; Kariithi *et al.,* 2013b; Otieno *et al.,* 1980). Two GpSGHV strains (infecting Ugandan and Ethiopian *G. pallidipes* populations) that are responsible for differential pathogenesis have been sequenced (Abd-Alla *et al.,* 2008; 2016). However, whether various tsetse species in the wild are infected with the same or different GpSGHV strains has so far not been elucidated.

To identify tsetse species in the field, several methods have been applied, including morphological characteristics such as external genitalia of males, the fly's habitat requirements and host preference (Vreysen *et al.,* 2013). Based on these characteristics, *Glossina* species are divided into three distinct taxonomic groups i.e. Morsitans, Palpalis and Fusca (Pollock, 1982). However, identification of closely related species and/or subspecies is challenging especially for the species occurring in overlapping ecogeographical regions and are nearly morphologically similar. This is because the experience and expertise of the taxonomists may be limited, and certain external features used in morphological identifications may be damaged during handling of the species. In addition to morphological taxonomic identification of *Glossina* species, molecular and genetic markers have also been used for decades. Nuclear markers, such as internal transcribed space 1 (ITS1) and ITS2, were reported to distinguish some of the species based on the size and/or specificity of the amplicons, as revealed by both agarose gel electrophoresis and sequencing of these markers (Chen *et al.,* 1999; Dyer *et al.,* 2009; 2008; 2011). Microsatellite markers have also been developed for different *Glossina* species and have provided encouraging results regarding their potential in phylogenetic analyses and species identification (Baker and Krafsur, 2001; Krafsur and Endsley, 2002; Luna *et al.,* 2001; Ouma *et al.,* 2003; 2006). Mitochondrial markers, including cytochrome oxidase 1 (COI), cytochrome oxidase 2 (COII), cytochrome b (CYTB), 16S rRNA, and NADH dehydrogenase 2 (ND2), have also been implemented for the phylogenetic analysis of *Glossina* species, based on DNA sequencing (Cordon-

Obras *et al.,* 2014; Dyer *et al.,* 2009; 2008; 2011; Echodu *et al.,* 2013; Marquez *et al.,* 2004; Solano, 2010). The availability of polytene chromosomes in *Glossina* and the development of polytene chromosome maps provide additional genetic tools that can shed light on specific chromosomal banding pattern changes and/or rearrangements that could as well provide diagnostic characters for species identification (Gariou-Papalexiou *et al.,* 2007; Pell and Southern, 1976; Southern and Pell, 1974).

A previously neglected parameter regarding speciation of species is the development of intimate relationships of the tsetse fly with bacterial symbionts, such as *Wigglesworthia glossinidia*, *Sodalis glossinidius,* and *Wolbachia*, that may alter the host's behavior (Doudoumis *et al.,* 2017; Soumana *et al.,* 2014b; Wamwiri *et al.,* 2013). *Wolbachia* is obligatory intracellular and maternally transmitted and is known to cause reproductive alterations and cytoplasmic incompatibility (CI) (Saridaki and Bourtzis, 2010). CI is mainly expressed as embryonic mortality when an infected male mate with an uninfected female (unidirectional CI) (Alam *et al.,* 2011), or when the crossed male and female harbor different and mutually incompatible *Wolbachia* strains (bidirectional CI) (O'Neill and Karr, 1990). Such incompatibilities lead to restriction of gene flow among natural populations and can be both 'accelerators' and diagnostic markers of speciation (Shropshire and Bordenstein, 2016). Another aspect of symbiosis that could be exploited is the presence of ancient, species-specific, horizontal gene transfer events in the host's chromosomal DNA. Such events have been demonstrated in *Glossina*, through the presence of fixed chromosomal introgressions of *Wolbachia* (only in *Glossina morsitans morsitans* up to now), and can provide additional diagnostic markers (Brelsfoard *et al.,* 2014; Doudoumis *et al.,* 2013).

Regarding the constraint of identification of closely related species and given that speciation can be driven through different or combined forces, integrative taxonomy suggests the utilization of multidisciplinary approaches for making robust conclusions regarding species identities and phylogenetic relationships (Dayrat, 2005; Fujita *et al.,* 2012; Padial *et al.,* 2010; Schlick-Steiner *et al.,* 2010; Schutze *et al.,* 2015). The utilization of a single marker, or a single class of tightly linked markers (e.g. mitochondrial genes), although easy to universally apply, they hardly

guarantee correct species identification (Meier *et al.,* 2006; Will *et al.,* 2005). The fact that the phylogenetic signal of mitochondrial markers can be masked or altered by the presence of reproductive symbionts, such as *Wolbachia* (through, for example, mitochondrial sweeps) and the limitation that mitochondrial markers are unable to identify hybrids among closely related species also points to the need for 'the more, the better' approaches in species identification (Kodandaramaiah *et al.,* 2013). Previous studies have documented that different classes of markers for tsetse flies, may provide either a differential depth of analysis or even contradicting results (Dyer *et al.,* 2009; 2011; Ouma *et al.,* 2007).

Besides robustness, it is critical to develop diagnostic tools that can be applied quickly, easily, massively and cost effectively. This can be done by integrating different classes of markers and by utilizing different resolution techniques, such as gel electrophoresis and sequencing. Such integrated approaches allow the screening of many individuals with reduced cost in a relatively short time and without the need of highly specialized equipment/skills. This chapter reports the evaluation of different classes of molecular markers (nuclear ITS1, nuclear microsatellites, mitochondrial genes, and the *Wolbachia* infection status) for the identification of tsetse species. These tools were evaluated against tsetse laboratory colonies that were used as reference material. At the same time, the data was correlated with previously published sequences as well as data from museum-derived tsetse specimens. Finally, the discriminative power of ITS1 amplicon and electrophoresis was evaluated through the genotyping of an extended collection of samples derived from nature. The chapter recommends a set of markers and analytical approaches that can quickly, and cost effectively support the morphometric taxonomy or even stand alone to identify *Glossina* species.

Materials and methods

Laboratory colonies

Glossina species maintained at the Insect Pest Control Laboratory (IPCL) of the Joint FAO/IAEA Programme of Nuclear Applications in Food and Agriculture (NAFA) were used in the analyses described in this chapter. The species were *G. pallidipes, G. morsitans morsitans, G. morsitans centralis, G. palpalis gambiensis, G. fuscipes*

fuscipes, and *G. brevipalpis*. Two more *Glossina* species (*G. morsitans submorsitans* and *G. tachinoides*) obtained from the Centre International de Recherche-Développement sur l'Elevage en zone Subhumide (CIRDES) laboratory in Burkina Faso were also included. The initial species identification of these samples was based on previously described standard morphological characters such as colour, size and male genitalia variations (Pollock, 1982). Details of the *Glossina* species and colonies used in this study are provided in **Table 1**. All the tsetse colonies were fed on heated, defibrinated bovine blood for 10-15 min, three days per week using an *in vitro* membrane feeding technique (Langley and Maly, 1969).

Museum specimens

Seven *Glossina* specimens were obtained from Mr Nigel P. Wyatt, Department of Entomology, Natural History Museum, London, UK (loan no. 2011-159), and comprised of representatives from *G. m. morsitans* (n=4), *G. m. centralis* (n=2), and *G. p. gambiensis* (n=1). These specimens were collected between 1915 and 1952 and were assigned to the respective taxa based on morphological characters mentioned in the previous section (**Table 1**).

Natural populations

A total of 2695 individual tsetse flies, representing 32 species/geographical location combinations from sub-Saharan Africa, were included in the analyses. These samples were collected during different periods between 1994 and 2015 (**Table 1**) and were used as a 'blind test' to verify their species status using the tools developed in this study.

Table 1: *Glossina* samples used in this study

***Glossina* species**	**Place of origin**	**Original collection date**	**Description of the tsetse species/colonies**	**No.**
		A. Laboratory colony		
G. pallidipes	Uganda (Tororo)	1975	1978 IPCL (from Institute of Experimental Entomology, Amsterdam, The Netherlands)	8
	Ethiopia (Arba Minch)	1997-2001	2005 IPCL (Arba Minch colony)	8
G. m. morsitans	Zimbabwe	1968	1972 IPCL (from Bristol laboratory colony)	8
G. m. centralis	Tanzania	N/A	1999 IPCL	8
G. p. gambiensis	Burkina Faso	1972	2005 IPCL (from CIRDES laboratory colony)	8
	Senegal (Pout)	2009	2009 IPCL	8
G. f. fuscipes	Central Africa Republic	1986	2009 IPCL	8
G. brevipalpis	Kenya (Shimba hills)	1987	2002 IPCL	8
G. tachinoides	Burkina Faso	N/A	CIRDES	12
G. m. submorsitans	Burkina Faso	N/A	CIRDES	12
Total				**88**
		B. Museum specimen		
G. m. morsitans	Tanganyika Terr (Morogoro, Uluguru)	1915	Dr. A. G. Wilkins	1
	Tanganyika (Korogwed Handeni)	1952	16-IX-52 Brit. Mus. 1959-638 Dr. E. Burtt	1
	Tanganyika Terr: (Morogoro, Uluguru)	1921	Dr. A.G. Wilkins Pres. by Imp. Bur. Ent. Brit. Mus. 1921-152.	2
G. m. centralis	Tanganyika Terr.	1923	Brit. Mus. 1923-269	1
	Sedamara (Mbulu)	1950	26.9.50 London School of Hygiene & Tropical Medicine coll. BMNH	1
G. p. gambiensis	Sierra Leone (Scarcies, Kambia)	1946	Nash & Walton, 26/1/46	1
Total				**7**

Glossina species	Place of origin	Original collection date	Description of the tsetse species/colonies	No.
		C. Field collection		
G. pallidipes	Ethiopia (Arba Minch)	2014		30
	Uganda (Lukoma – Bavuma)	2013		27
	Kenya (BioRI-KALRO)	2008		3
	Zambia (Mfuwe)	2007		1
	Zimbabwe (Ruckomechi)	2006		1
	Zimbabwe (Makuti)	2006		1
	Tanzania (Tanga)	2005		2
G. m. morsitans	Zambia (Mfuwe)	2007		3
	Zimbabwe (Ruckomechi)	2006		2
	Zimbabwe (Makuti)	2006		1
	Tanzania (Usinge)	2013		9
	Kenya (BioRI-KALRO)	2008		1
G. m. centralis	Angola (Guissakina)	2013		25
	Tanzania (Ugalla)	2013		60
*G. m. submorsitans**	Burkina Faso (Comoe)	2009		277
*G. p. gambiensis**	Senegal (Sebikotane)	2009		3
	Senegal (Sebikotane)	2013		9
	Senegal (Kayar)	2010		3
	Senegal (Kayar)	2013		17
	Senegal (Niokolo-Koba)	2012		3
	Senegal (Niokolo-Koba)	2013		30
	Senegal (Pout)	2009		11
	Senegal (Pout)	2013		30
	Burkina Faso (Comoe)	2008		1152
	Mali	2010		8
	Guinea	2010		1

Glossina species	Place of origin	Original collection date	Description of the tsetse species/colonies	No.
G. f. quanzensis	Angola (Guissakina)	2013		3
	Uganda	2013		52
G. brevipalpis	Mozambique (Maputo GR)	2013		6
*G. swynnertoni**	Tanzania (Ikorongo GR)	2015		24
G. medicorum	Burkina Faso (Comoe)	2009		86
*G. tachinoides**	Burkina Faso (Comoe)	2009		792
	Ghana	2009		7
G. austeni	Mozambique (Maputo G)	2013		7
	Tanzania (Jozani)	1994		1
	Zanzibar (Unguja island)	1995		5
	South Africa (Zululand)	1999		1
Total				**2695**

No.: Number of individuals tested, N/A = not available; CIRDES = Centre International de Recherche-Développement sur l'Elevage en zone Subhumide, Bobo Dioulasso, Burkina Faso; IPCL = Insect Pest Control Laboratory, *these collections included false assigned individuals (see **Table 5**).

DNA extraction, PCR, and sequencing of flies derived from laboratory colonies and natural populations

DNA from teneral (unfed; within 24 h post eclosion) adult flies of each laboratory colony was isolated using the Qiagen DNeasy kit (Qiagen, Valencia, CA), following the manufacturer's instructions. DNA samples were stored at 4°C until their use and at -20°C for long-term storage. Natural population tsetse flies were initially morphologically sorted and labelled by species during filed collection, preserved in 95% ethanol (or propylene-1,2-diol), and then shipped to the IPCL for downstream analyses. DNA extraction was performed as described for the laboratory colonies. For all PCR amplifications (in a 25 µl reaction volumes), 1.1X pre-aliquoted PCR master mix was used (ABgene, UK). In 22.5 µl of the mix, 1.5 µl (~50 ng) of DNA template and 1µl of forward and reverse primer were added (10µM each). Nuclear (ITS1 and microsatellite), mitochondrial (*COI*, 16S *rRNA,* and 12S *rRNA*), and symbiotic markers (*Wolbachia* 16S *rRNA* gene) that were used in the present study are shown in **Table 2**.

PCR conditions to amplify *COI*, *16S rRNA* and *ITS1* genes were as described previously (Dyer *et al.,* 2008). Primers 12SCFR and 12SCRR were used to amplify a 377 bp fragment of the *12S rRNA* mitochondrial gene, as previously described by Doudoumis *et al*. (2012). PCR conditions to detect the presence of cytoplasmic or nuclear *Wolbachia 16S rRNA* gene were as described previously using the *Wolbachia* specific primers wspecF and wspecR (Doudoumis *et al.,* 2012). PCR conditions used for the different sets of microsatellite markers have been previously described by various researchers (Baker and Krafsur, 2001; Brown *et al.,* 2008; Dyer *et al.,* 2008; Luna *et al.,* 2001; Ouma *et al.,* 2006; Solano *et al.,* 1997). PCR products were analysed on 1.5 % or 2.5 % agarose gel electrophoresis and visualized using ethidium bromide. Amplicons of the mitochondrial genes were purified using QIAquick PCR kit (Qiagen Valencia, CA) and sequenced by MWG (MWG-Biotech AG, Germany). Forward and reverse sequences with good quality read were assembled and aligned using SeqMan Pro software (Lasergene 7.0, Dnastar Inc). The consensus sequences for each gene were aligned and trimmed using the ClustalW algorithm in MEGA version 6.0.

Table 2: A list of the molecular markers and primers used in this study

Molecular marker	Marker		Primer name	Primer sequence 5'-3'	Reference	Method of analysis
Nuclear markers	ITS1		*Glossina*ITS1_for	GTGATCCACCGCTTAGAGTGA	(Dyer *et al.*, 2008)	Gel electrophoresis
			*Glossina*ITS1_rev	GCAAAAGTTGACCGAACTTGA		
	Microsatellite markers	A10	A10 F	GCAACGCCAAGTGAAATAAAG		
			A10 R	TACTGGGCTCGCGTACATAAT		
		Gmm14	Gmm14 F	CACACCCTGGATTACAAA	(Baker and Krafsur, 2001)	
			Gmm14 R	TGAAATGCAACCCTTCTT		
Mitochondrial markers	COI		COI	TTGATTTTTTGGTCATCCAGAAGT	(Simon *et al.*, 1994)	DNA sequencing
			CULR	TGAAGCTTAAATTCATTGCACTAATC		
	16S rRNA		NI-J-12585	GGTCCCTTACGAATTTGAATATATCCT		
			LR-N-12866	ACATGATCTGAGTTCAAACCGG		
	12S rRNA		12SCFR	GAGAGTGACGGGCGATATGT	(Doudoumis *et al.*, 2012)	
			12SCRR	AAACCAGGATTAGATACCCTATTAT		
Symbiotic markers	*Wolbachia*	16S rRNA	WspecF	YATACCTATTCGAAGGGATAG		Gel electrophoresis
			WspecR	AGCTTCGAGTGAAACCAATTC		

DNA extraction, PCR, and sequencing of museum specimens

Prior to DNA extraction, *Glossina* specimens from the museum were surface-sterilized by immersing in 80% ethanol and then rinsed twice with sterile phosphate buffered saline (PBS). DNA was extracted using Nucleospin Tissue Kit (Macheray-Nagel) following the manufacturer's instructions. DNA samples were stored at 4°C until further analyses and at -20°C for long-term storage. Only the mitochondrial *12S rRNA* gene was analyzed for these samples as described above. PCR amplifications were performed in reactions containing 10 ng DNA, 10 pmol of each primer, 0.5 units KAPA Taq (KAPA Biosystems), 1x KAPA buffer A (KAPA Biosystems), 0.25 mM deoxynucleotide triphosphate mixture (dNTPs) and water to a final volume of 20 µl. Amplification was performed in a PTC-200 Thermal Cycler (MJ Research), using the following cycling conditions: 95°C for 5 min, followed by 40 cycles of 30 s at 95 °C, 30 s at 54 °C, 1 min at 72 °C and a final extension of 10 min at 72 °C. PCR reactions were electrophoresed on a 1.5% agarose gel. To ensure that the negative samples were in deed negative, these samples were reamplified by PCR using 2 µl of the respective first PCR reaction as template and the same set of primers and conditions for 35 cycles. Positive samples of the first or the second PCR reaction products were further analyzed by double-stranded sequencing with both forward and reverse primers. A dye terminator-labelled cycle sequencing reaction was conducted with the BigDye Terminator v3.1 Cycle Sequencing Kit (PE Applied Biosystems). Reaction products were analyzed using an ABI PRISM 310 Genetic Analyzer (PE Applied Biosystems). The gene sequences generated were assembled and manually edited with SeqManII by DNAStar (Lasergene). For each sample, a majority-rule consensus sequence was created.

Phylogenetic analysis

Phylogenetic analysis was performed using MEGA 6.0 software (Tamura *et al.,* 2013), using Maximum-Likelihood (ML) based on the General Time Reversible model with gamma distributed rates with 1000 bootstrap replications. *Musca domestica* sequences, which are closely related to *Glossina* genus according to the BLAST search were used as outgroup for each of the analysed genes (*COI*; gi|514058521, *12S rRNA*; GI: 51039400).

Results

Evaluation of the discriminating power of different molecular tools

For the initial evaluation of the available molecular tools, 10 laboratory tsetse colonies were used, and 8-12 individuals were genotyped per colony (**Table 1**).

Mitochondrial markers: COI and 16S rRNA genes

Sequence datasets generated for each of the mitochondrial genes (600 bp for *COI* and 207 bp for *16S rRNA*) were aligned for all ten *Glossina* laboratory colonies. The phylogenetic reconstruction for each of the mitochondrial markers clearly clustered into the three *Glossina* taxonomic groups; Palpalis, Morsitans and Fusca (**Figure 1**).

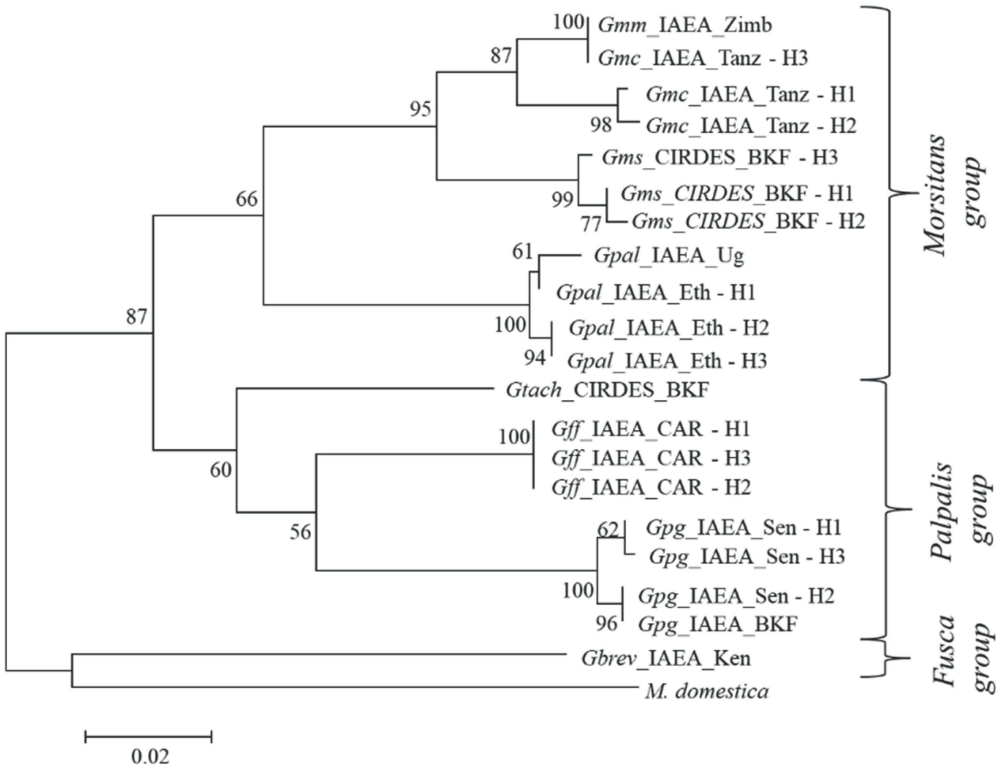


Figure 1: Molecular phylogenetic analysis of laboratory populations by Maximum Likelihood (ML) method, using a *COI* gene fragment. The evolutionary history was inferred by using the ML method based on the Tamura-Nei model. *Musca domestica COI* was used as outgroup. The numbers at each node represent bootstrap proportions based on 1000 replications. Abbreviations (Species): *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides,* (Countries): BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ug; Uganda, Ken; Kenya, Sen; Senegal, Tanz; Tanzania, Zimb; Zimbabwe, H= Haplotype.

COI was more informative since it identified various haplotypes found in different tsetse species than *16S rRNA* and was selected as a representative gene of the mitochondrial DNA (**Figure 1**). However, clustering of sub-species and closely-related species was not always accurate, as in the case of *G. m. morsitans* and *G. m. centralis*. Within some species, distinct haplotypes (a set of alleles at linked loci) were observed using either the *COI* gene (**Figure 1**) or the 16S *rRNA* gene (data not shown). For instance, *G. m. centralis, G. pallidipes* from Ethiopia, *G. f. fuscipes,* and *G. p. gambiensis* from Senegal were found to have three haplotypes each (H1, H2, H3) for the *COI* dataset.

Nuclear markers: ITS1 and microsatellite markers

Variation in the length of the ITS1 amplicon was observed across the different *Glossina* laboratory colonies, consistent with the species identification (**Figure 2**). Based on size and/or number of the amplicons on the agarose gels, most of the species were successfully separated. Among eight screened species, only *G. m. centralis/G. m. submorsitans* and *G. m. morsitans/G. brevipalpis* could not be distinguished from each other. To further evaluate the discriminative power of ITS1, field-collection representing *G. swynnertoni* from Tanzania, which is closely related to the *G. m. morsitans* was added in this analysis. This sample shared the ITS1 pattern of the *G. m. morsitans/G. brevipalpis* group (~775 bp) (**Figure 2**). However, sequencing analysis showed a three bp difference between the amplicons of *G. brevipalpis* (778 bp) and those of *G. m. morsitans* and *G. swynnertoni* (775 bp). This difference can be used to identify *G. brevipalpis*, using higher resolution fragment analysis approaches, such as polyacrylamide gel, low melting agarose or capillary electrophoresis.

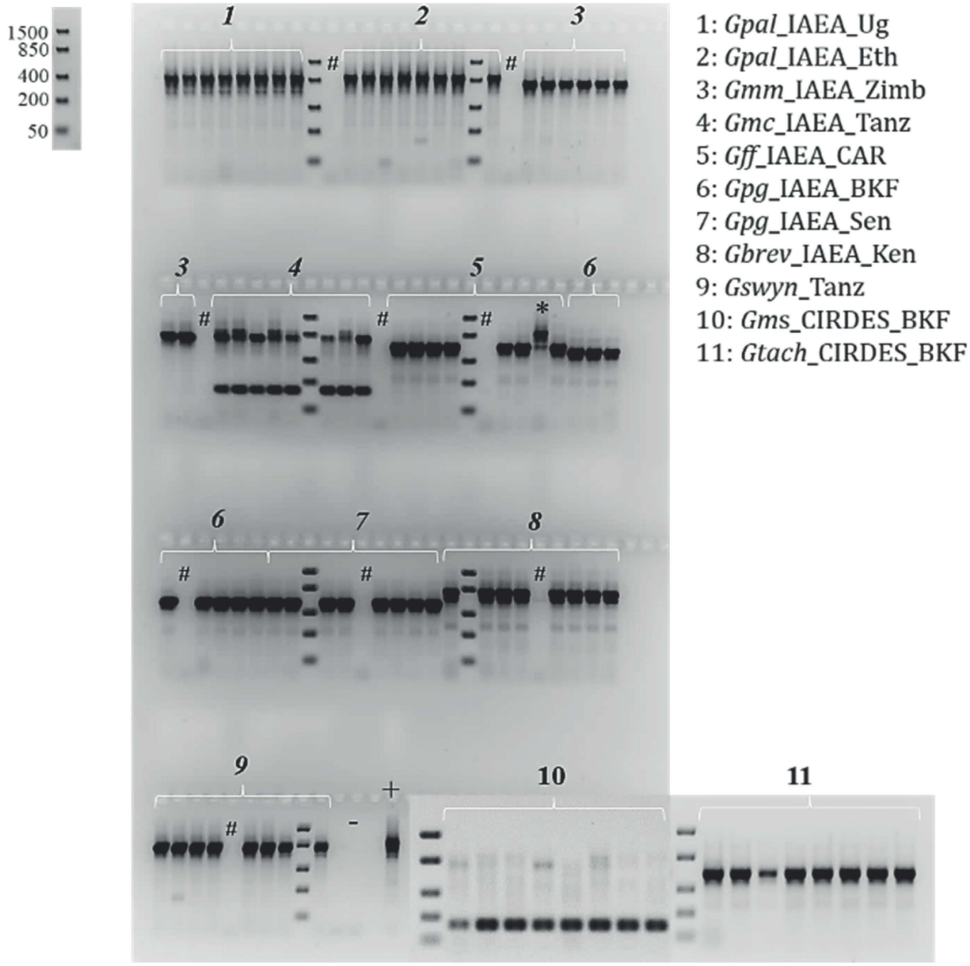


Figure 2: Agarose gel electrophoresis (2.5% agarose) showing the ITS1 gene amplicons for the different tsetse laboratory populations. Eight flies per laboratory population are presented. The DNA ladder used to determine the size of the analyzed PCR products is also shown (top left corner of the figure). #: Negative control during DNA extraction; *: misplaced sample; -: negative PCR control; +: positive PCR control (*G. pallidipes* DNA). Abbreviations (Species): *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides, Gswyn; G. swynnertoni,* (Countries): BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ug; Uganda, Ken; Kenya, Sen; Senegal, Tanz; Tanzania, Zimb; Zimbabwe.

A set of 36 previously published microsatellite markers were also tested against 1-3 individuals of each of the ten laboratory populations (**Table 3**) (Baker and Krafsur, 2001; Brown *et al.,* 2008; Dyer *et al.,* 2008; Luna *et al.,* 2001; Ouma *et al.,* 2006; Solano *et al.,* 1997). The analysis was carried out only with agarose gel electrophoresis, which showed that there are microsatellite markers producing species-specific amplicons in the expected size range.

Table 3: The set of microsatellites markers tested for the identification of *Glossina* species. These markers were evaluated against different laboratory populations, considering the amplification of the expected PCR product.

Microsatellite marker	*Glossina* Species								
	Gpal	***Gmm***	***Gmc***	***Gms***	***Gswyn***	***Gtach***	***Gbrev***	***Gff***	***Gpg***
GffA3				X	X		X	X	X
GffA9					X	X		X	X
GffB101								X	X
GffA10 (or 'A10')							X	X	X
69.22Gpg		X	X		X		X	X	X
GffB8								X	X
GffA19a		X	X	X	X	X	X	X	X
GffA23b									
GpB6b	X								
GffA6		X	X	X	X	X	X	X	X
Gpc107	X	X	X	X	X	X	X	X	X
55.3Gpg						X	X	X	X
19.62Gpg						X	X	X	X
Gmm8	X	X	X	X	X	X	X	X	X
Gmm14	X	X	X	X	X	X		X	X
Gmm15		X	X	X	X				
Gmm22	X	X	X	X	X	X	X	X	X
Gmm5	X	X	X	X	X				
GpB115	X	X	X	X	X				
GpB20b	X	X	X	X	X		X	X	X
GpC5b	X	X	X		X				
Gmm9B		X			X				

Microsatellite marker	*Glossina* Species								
	Gpal	***Gmm***	***Gmc***	***Gms***	***Gswyn***	***Gtach***	***Gbrev***	***Gff***	***Gpg***
GmsCAG6	X	X	X	X	X	X	X	X	X
GmcCA16c	X	X	X	X	X		X	X	X
GmsCAG2	X	X	X	X	X	X	X	X	X
GmsCAG29B	X	X	X	X	X	X	X	X	X
GpCAG133	X	X	X	X	X			X	X
Gff112								X	
Gpc101		X	X	X	X	X	X	X	X
GpD18b		X	X	X	X				
GpC10b	X	X	X	X	X	X		X	X
GpC26b	X	X	X	X					
Gmm127	X	X	X	X	X				
GffC107	X	X	X	X	X	X	X	X	X
GffD6	X	X	X	X	X	X		X	X
GffD109	X	X	X	X	X	X	X	X	X

X: presence of the expected amplicon. In bold: markers selected for downstream genotyping purposes. Abbreviations: *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides, Gswyn; G. swynnertoni.*

As an example, microsatellite marker A10, which had been designed for *G. f. fuscipes* and was reported to be specific for *G. p. gambiensis* (Dyer *et al.*, 2008), produced the expected amplicon in all *G. p. gambiensis* individuals plus some of the *G. f. fuscipes* and *G. brevipalpis* samples but gave no amplicons in all other species (**Figure 3A**). In addition, microsatellite marker Gmm14 was amplified in all the species analyzed except *G. brevipalpis* (**Figure 3B**).

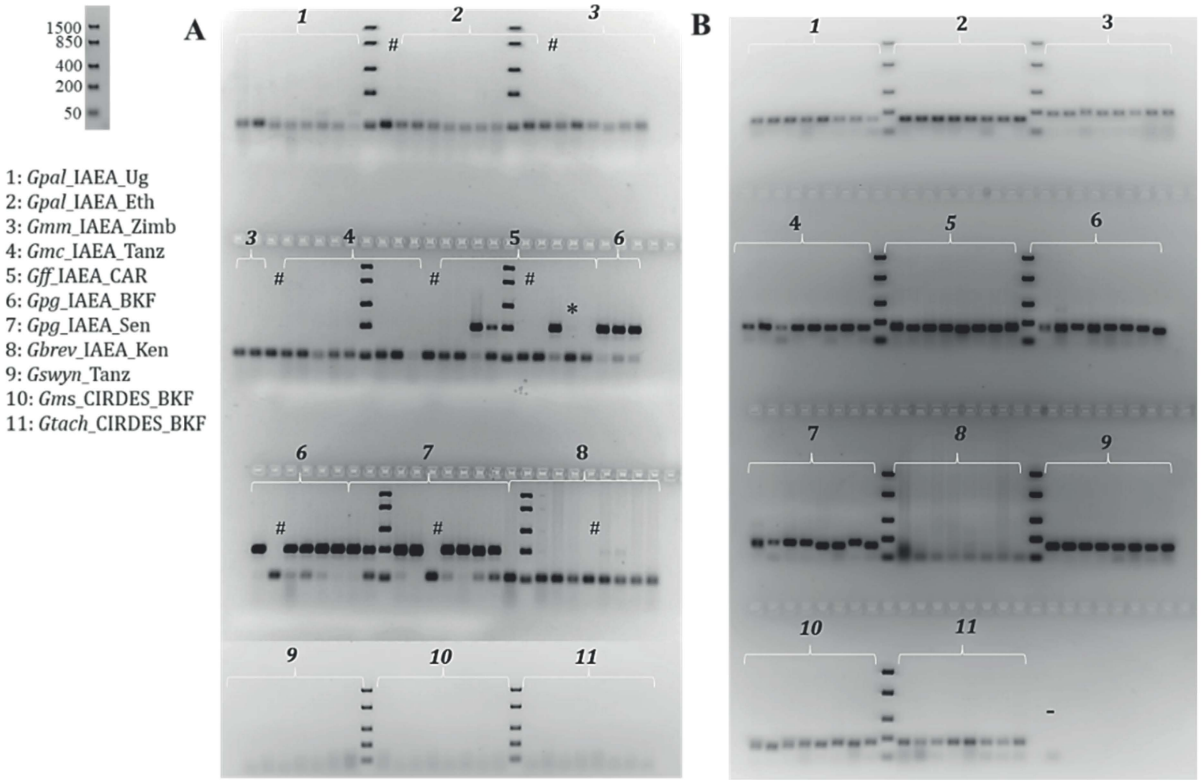


Figure 3: Agarose gel electrophoresis (2.5 % agarose) presenting microsatellite markers A10 (A) and Gmm14 (B) amplicons for the different laboratory populations. Eight flies per laboratory population are presented. The DNA ladder used to determine the size of the analyzed PCR products is also shown. #: Negative control during DNA extraction; *: misplaced sample; -: negative PCR control. Abbreviations (Species): *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides, Gswyn; G. swynnertoni,* (Countries): BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ug; Uganda, Ken; Kenya, Sen; Senegal, Tanz; Tanzania, Zimb; Zimbabwe.

Wolbachia 16S rRNA

The presence of *Wolbachia* was analyzed with a *Wolbachia* specific *16S rRNA*-based PCR. The prevalence of *Wolbachia* infections differed significantly between the different laboratory colonies. A fixed cytoplasmic *Wolbachia* infection (with strong PCR amplicons) was detected only in *G. m. centralis*. High infection prevalence (with strong PCR amplicons) was also observed in *G. brevipalpis* and *G. m. morsitans*. Sporadic infections (with weak PCR amplicons) were observed in *G. pallidipes* and *G. f. fuscipes*. However, *G. m. morsitans* presented the fixed chromosomal insertion (296 bp amplicon), which was previously reported by Doudoumis *et al.* (2012) to be absent in the other *G. m. morsitans* laboratory colonies. The remaining colonized species (*G. m. sub-morsitans, G. p. gambiensis,* and *G. tachinoides*) did not give any amplicon indicative of either active cytoplasmic infection or chromosomal insertion of *Wolbachia*. In addition, *G. swynnertoni* from Tanzania, which shared the ITS1 pattern with the *G. m. morsitans/G. brevipalpis* group, was negative of both active cytoplasmic infection and chromosomal insertion of *Wolbachia* (**Figure 4**).

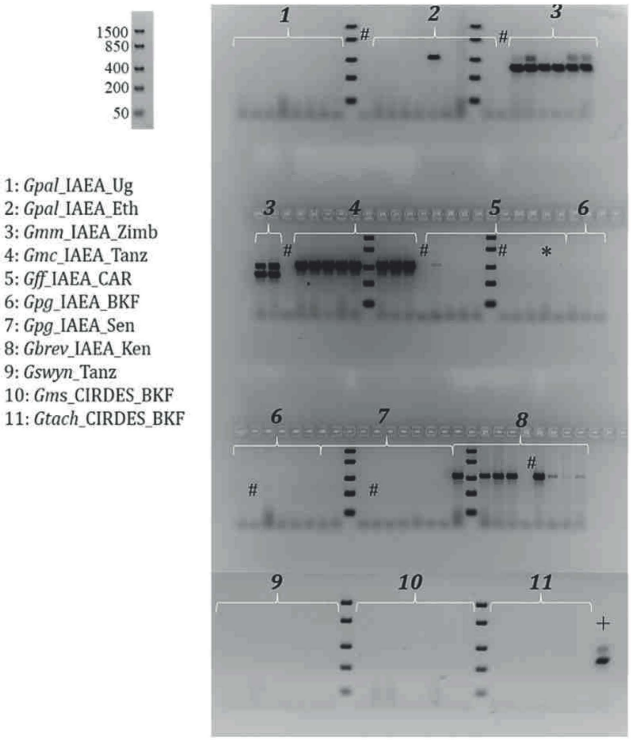


Figure 4: Agarose gel electrophoresis (1.5% agarose) showing the *Wolbachia* amplicons for the different laboratory populations. The presence of the 438 bp amplicon is indicative of an active (cytoplasmic) *Wolbachia* infection, while the 296 bp amplicon is indicative of the presence of the partial sequence of the *Wolbachia* 16S rRNA gene that is integrated into the tsetse genome. Eight flies per laboratory population are presented. The DNA ladder used to determine the size of the analyzed PCR products is also shown. #: Negative control during DNA extraction; *: misplaced sample; +: positive PCR control (*G. m. morsitans* DNA). Abbreviations (Species): *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides, Gswyn; G. swynnertoni,* (Countries): BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ug; Uganda, Ken; Kenya, Sen; Senegal, Tanz; Tanzania, Zimb; Zimbabwe.

Correlation of laboratory Glossina species with museum specimens

Due to low DNA quality, only few amplicons were obtained from the museum specimens and only for the 12S *rRNA* gene. Therefore, representative samples from all laboratory colonies were also sequenced for the 12S *rRNA* gene. Despite the limited resolution provided, the laboratory colony sequences aligned (180 bp) and correlated well with the museum specimens (**Figure 5**).

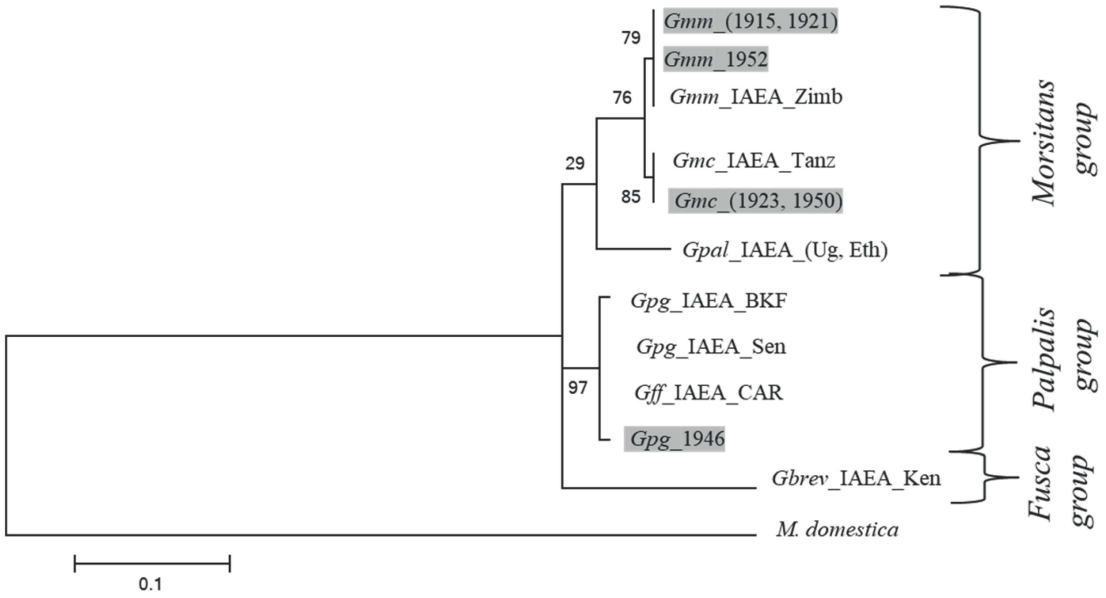


Figure 5: Molecular phylogenetic analysis of laboratory populations (not highlighted) and museum specimens (highlighted in grey) by Maximum Likelihood (ML) analyses, using the *12S rRNA* gene sequence. The evolutionary history was inferred by using the ML method based on the Tamura-Nei model. *Musca domestica* was used as outgroup. The numbers at each node represent bootstrap proportions based on 1000 replications. Abbreviations (Species): *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides,* (Countries): BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ug; Uganda, Ken; Kenya, Sen; Senegal, Tanz; Tanzania, Zimb; Zimbabwe.

Evaluation of COI as a 'stand-alone' marker for species identification

COI gene sequence was used to first correlate the reference laboratory colonies with published sequences of different species and secondly to identify selected samples from the field that were available in IPCL DNA base. In general, laboratory colonies correlated well to both the previously published sequences (**Figure 6A**) and to the

field-collected samples available from the IPCL DNA base (**Figure 6B**). In addition, these phylogenetic analyses clearly clustered all the analysed samples into the three taxonomic groups of *Glossina* (Palpalis, Morsitans and Fusca).

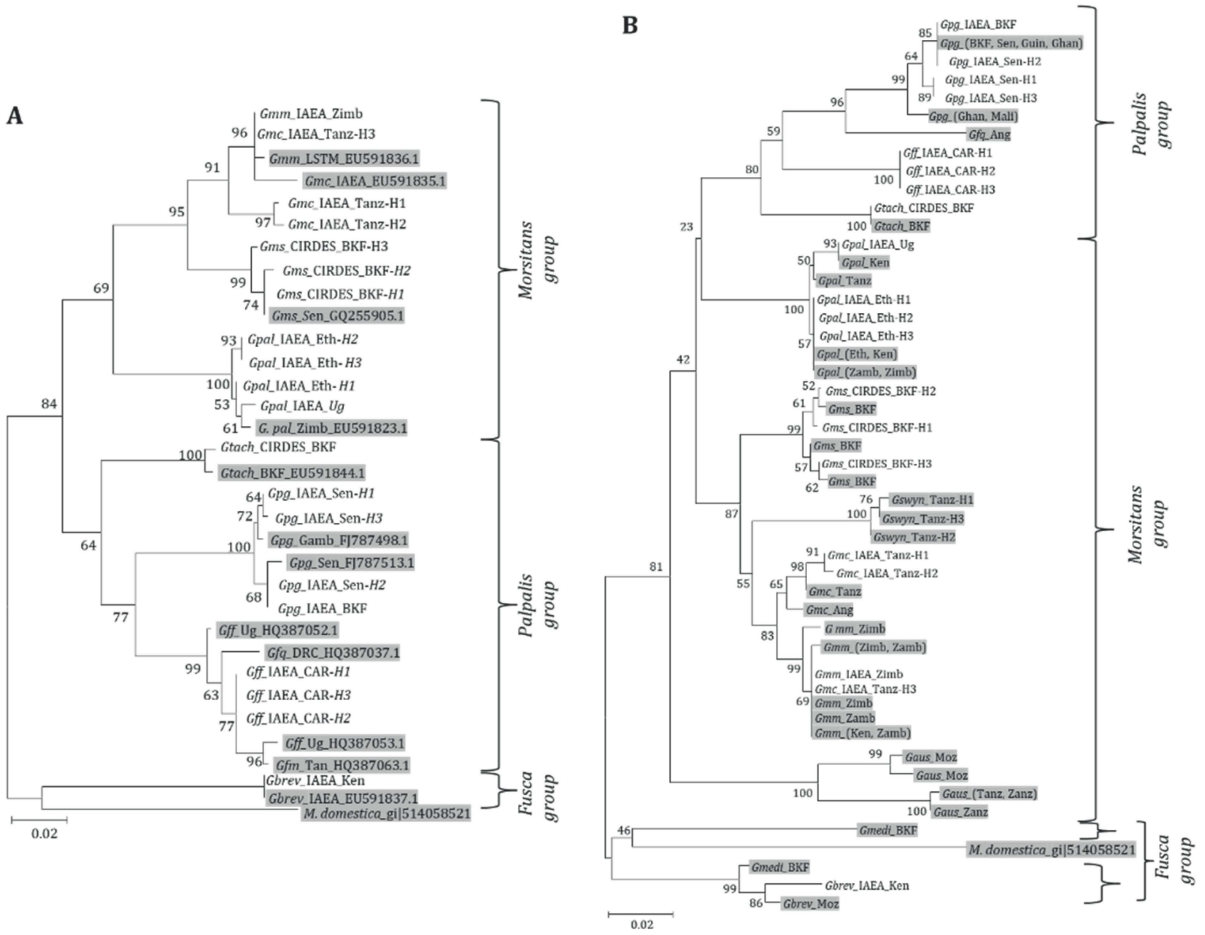


Figure 6: Molecular Phylogenetic analysis of laboratory populations, published sequences, and selected samples from collections deriving from wild, by Maximum Likelihood method using a COI gene fragment. Analysis of the laboratory populations (not highlighted) in comparison to; **A**), the already published COI sequences of different *Glossina* species available in the NCBI database (highlighted in grey) and **B**) to the *Glossina* species samples collected from the field (highlighted in grey). *Musca domestica* was used as outgroup. The numbers at each node represent bootstrap proportions based on 1000 replications. Abbreviations (Species): *Gaus; G. austeni Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gfq; G. f. quanzensis, Gmedi; G. medicorum, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes; Gpg; G. p. gambiensis, Gtach; G. tachinoides, Gswyn; G. swynnertoni*, (Countries): Ang; Angola, BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ghan; Ghana, Guin; Guinea, Ken; Kenya, Moz; Mozambique, Sen; Senegal, Tanz; Tanzania, Zamb; Zambia, Zanz; Zanzibar, Zimb; Zimbabwe, H= Haplotype.

However, *COI* cannot clearly resolve closely related species (subspecies or complex species), as was the case of the *G. m. centralis* haplotype 3 (H3) which has similar *COI* gene sequence to the *G. m. morsitans*. The same case was found in *G. f. quanzensis* from Angola, which is more closely related to the *G. p. gambiensis* rather than to the rest of the *G. fuscipes* (**Figure 6B**).

Development of a multi-marker species identification approach

Based on the initial data derived from the laboratory colonies, focus was directed to the discriminative power of the combined use of ITS1, microsatellite markers Gmm14/A10, and the *Wolbachia* status (both cytoplasmic and chromosomal), utilizing only agarose gel electrophoresis. Previous findings (Dyer *et al.,* 2008) as well as the findings of this study, suggested that the length of the ITS1 amplicon should be sufficient to identify most of the species analyzed, except in the case of *G. m. centralis/G. m. submorsitans* group and the *G. m. morsitans/G. swynnertoni/ G. brevipalpis* (**Figure 2**). To differentiate *G. m. centralis* from *G. m. submorsitans*, *Wolbachia* infection status (cytoplasmic) can be used, since only *G. m. centralis* is infected (**Figure 4**). Further, to differentiate *G. m. morsitans* from *G. brevipalpis/ G. swynnertoni*, the *G. m. morsitans* – specific chromosomal introgression of the *Wolbachia* 16S *rRNA* gene can be used (**Figure 4**). Additionally, since some of the *G. brevipalpis* samples are not infected with *Wolbachia* and can be mixed with *G. swynnertoni* samples, the use of microsatellite marker Gmm14 can distinguish them from *G. swynnertoni* (**Figure 3B**). These results are summarized in **Table 4**.

Table 4: Analysis of ITS1 sequence length, microsatellite markers and *Wolbachia* status in *Glossina* laboratory populations

***Glossina* species**	**Place of origin**	**No.**	**ITS1 expected size**	***Wolbachia***		**Microsatellites**		**Correctly identified samples**
				cytoplasmic	chromosomal	A10	Gmm14	
Gpal	IPCL, Ug	8	920	0.0 % (0/8)	0.0 % (0/8)	-	+	8/8
	IPCL, Eth	8		12.5 % (1/8)	0.0 % (0/8)	-	+	8/8
Gmm	IPCL, Zim	8	775	75 % (6/8)	100 % (8/8)	-	+	8/8
Gswyn	Tanz	24	775	0.0 % (24/24)	0.0 % (24/24)	-	+	21/24
Gmc	IPCL, Tanz	8	~800 + ~150	100 % (8/8)	0.0 % (0/8)	-	+	8/8
Gms	CIRDES, BKF	12	~800 + ~150	0.0 % (0/12)	0.0 % (0/8)	-	+	12/12
Gpg	IPCL, Sen	8	543	0.0 % (0/8)	0.0 % (0/8)	+	+	8/8
	IPCL, BKF	8		0.0 % (0/8)	0.0 % (0/8)	+	+	8/8
Gff	IPCL, CAR	8	618	12.5 % (1/8)	0.0 % (0/8)	Partial	+	8/8
Gbrev	IPCL, Ken	8	778	75 % (6/8)	0.0 % (0/8)	Partial	-	8/8
Gtach	CIRDES, BKF	12	597	0.0 % (0/12)	0.0 % (0/8)	-	+	12/12

-: no amplicon detected, +: the expected amplicon was detected in all individuals screened, Partial: the expected amplicon was detected, but not in all individuals screened. Abbreviations (Species): *Gbrev; G. brevipalpis, Gff; G. f. fuscipes, Gmm; G. m. morsitans, Gmc; G. m. centralis, Gms; G. m. submorsitans, Gpal; G. pallidipes, Gpg; G. p. gambiensis, Gtach; G. tachinoides, Gswyn; G. swynnertoni,* (Countries): BKF; Burkina Faso, CAR; Central Africa Republic, Eth; Ethiopia, Ug; Uganda, Ken; Kenya, Sen; Senegal, Tanz; Tanzania, Zimb; Zimbabwe.

With these results, an approach that can be used to differentiate *Glossina* species is summarized in **Figure 7.** Following this approach and without using any morphological data, all ten laboratory colonies (representing eight species) were accurately resolved.

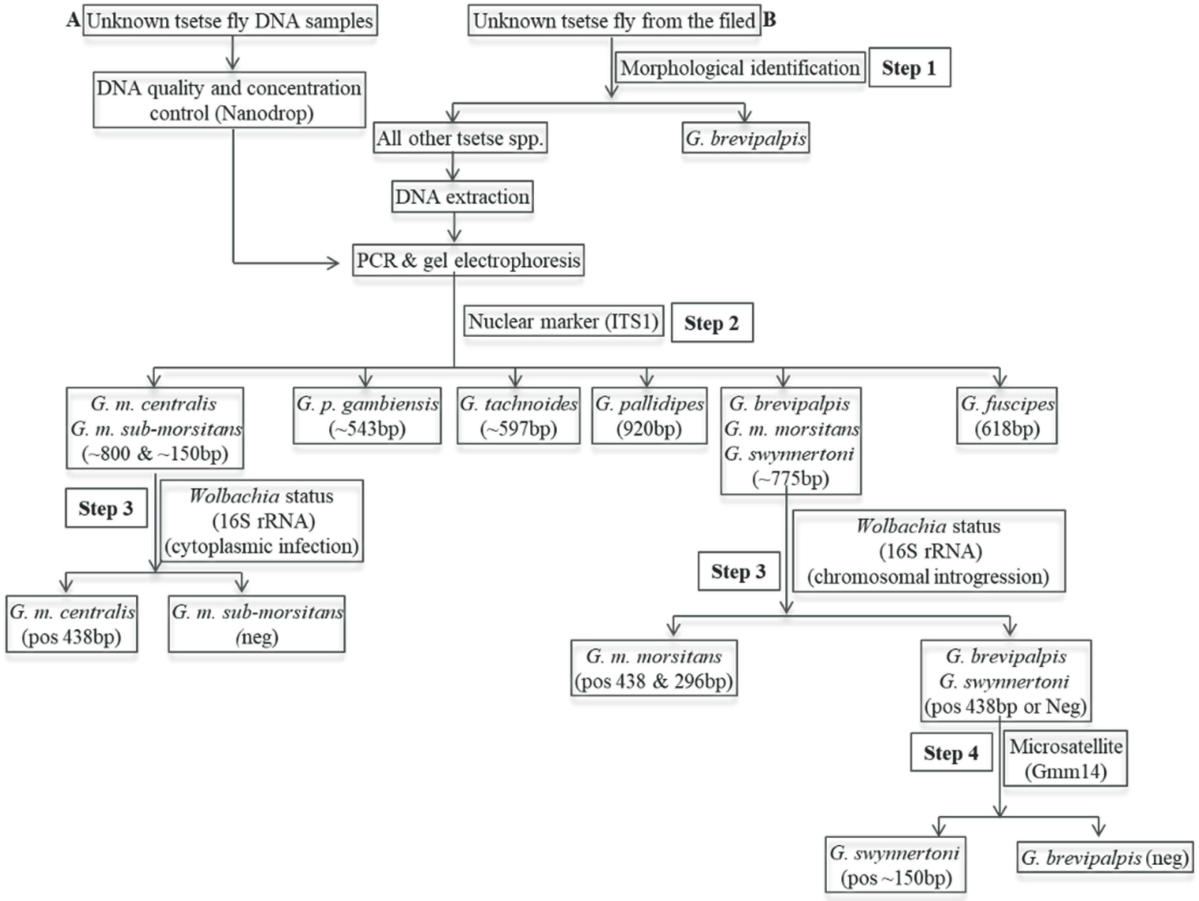


Figure 7: A multi-marker-based approach to distinguish tsetse species, based on agarose gel electrophoresis. This approach relies on the amplicons (size and number) of ITS1, presence/absence of microsatellite Gmm14 or the *Wolbachia* specific 16S rRNA amplicons (both cytoplasmic and chromosomal).

The 'blind test' using ITS1, selected microsatellite markers, and Wolbachia

To further test the resolution power of this combined approach, a 'blind test' of randomly selected DNAs of individuals collected from the field and available from the DNA base of the IPCL was performed. The first step was the application of the ITS1 marker. A total of 2695 individuals were genotyped and 2662 (98.78 %) were

assigned to the expected taxon (**Table 5**), based on the information available during the sample collection in the field. For 33 individuals, there was a discrepancy between data obtained upon collection and ITS1 profile. Specifically, data from collection sites of 0.57 % of the *G. p. gambiensis* samples (seven out of 1267), 7.94 % of the *G. m. submorsitans* samples (22 out of 277), 0.13 % of the *G. tachinoides* samples (one out of 799), and 12.5 % of the *G. swynnertoni* samples (three out of the 24), were not in agreement with the molecular identification (**Table 5**). These samples were further analyzed to confirm their identities using *Wolbachia* infection status, the amplicon profile of microsatellite markers A10 and Gmm14, and the sequencing data of *COI* gene. The combined use of the four classes of markers, along with data of the geographical distribution of *Glossina* species verified the taxon of these samples, showing that they were cases of either misidentification in the field or subsequent mislabeling (**Table 5**). The combined use of these markers resulted in correct identification of all samples.

In this study, four field-collected species that had no laboratory colonies as reference were included (*G. austeni, G. f. quanzensis, G. medicorum,* and *G. swynnertoni*). However, the estimated size of ITS1 amplicons were in accordance with that expected from previous studies. The pattern of ITS1 was sufficient to differentiate both *G. austeni* (amplicon of 633 bp) from all other taxa of this study, although this amplicon size is similar to the *G. fuscipes* amplicon size (633 bp). *G. f. quanzensis* could not be differentiated from *G. f. fuscipes*, based on the single agarose gel electrophoresis of the ITS1 amplicon. *G. medicorum* gave two amplicons, with one having a size between 600 and 700 bp, and the other being close to the one expected from previous studies (~880 bp). However, in the samples used for the current analyses, the amplicon of lower molecular weight (600 -700 bp) was more robust and consistent than the expected one. *G. swynnertoni* provided a unique combined profile: (a) the COI sequencing data place these samples close to *G. m. centralis* and *G. m morsitans* (**Figure 6B**), (b) the ITS1 profile (amplicon size) is similar or identical to *G. m. morsitans* and *G. brevipalpis* (**Figure 2**) and (c) it has complete absence of both cytoplasmic and chromosomal *Wolbachia* (**Figure 4**).

Table 5: Validation of Tsetse species from field-collected samples using *Glossina* ITS1

Tsetse field-collected species	Expected band size	Tested flies	Correctly identified flies		Misidentified flies		Band size of the misidentified samples	Corrected identification
			No.	%	No.	%		
G. pallidipes	920	69	69	100	0	0	-	-
G. m. morsitans	775	13	13	100	0	0	-	-
G. m. centralis	~800 + 150	85	85	100	0	0	-	-
G. p. gambiensis	**543**	**1267**	**1260**	**99.44**	**7**	**0.56**	**800 + 150**	***G. m. submorsitans*[1]**
G. f. fuscipes	618	52	52	100	0	0	-	-
G. f. quanzensis	618	3	3	100	0	0	-	-
G. m. submorsitans	**~800 + 150**	**277**	**255**	**92.06**	**22**	**7.94**	**597**	***G. tachinoides*[2]**
G. brevipalpis	775	6	6	100	0	0	-	-
G. tachinoides	**597**	**799**	**798**	**99.87**	**1**	**0.13**	**800 + 150**	***G. m. submorsitans*[1]**
G. austeni	700	14	14	100	0	0	-	-
G. medicorum	~850+~650	86	86	100	0	0	-	-
G. swynnertoni	**775 bp[4]**	**24**	**21**	**87.5**	**3**	**12.5**	**920**	***G. pallidipes*[3]**
total		2,695	2,662	98.78	33	1.22		

In grey scale: field collections lacking reference laboratory populations. In bold: field collections where discrepancies between data deriving from collection sites and molecular identification was observed. [1]Based on the ITS1 profile, non-amplification of microsatellite A10, complete absence of the cytoplasmic infection of *Wolbachia*, and the geographical distribution of tsetse species, and so the samples were identified as *G. m. submorsitans*. [2]Based on the ITS1 profile, non-amplification of A10 microsatellite marker, absence of cytoplasmic and chromosomal *Wolbachia*, and the geographic distribution of tsetse species, these individuals were identified as *G. tachinoides*. [3]Based on the ITS1 profile, *COI* profile, amplification of both A10 and Gmm14 microsatellite markers, absence of cytoplasmic and chromosomal *Wolbachia*, and the geographic distribution of tsetse species, these individuals were identified as *G. pallidipes*. [4]For *G. swynnertoni*, there was no ITS1 amplicon expected from previous studies. The one generated in the present study is stated as 'expected'.

Of interest is the combined use of ITS1 and *Wolbachia* to differentiate among the subspecies of *G. morsitans.* As shown in **Table 4**, *G. m. morsitans* has a distinct ITS1 profile and the presence of the chromosomal introgression of *Wolbachia. G. m. centralis* and *G. m submorsitans*, which share the same characteristic ITS1 pattern can be differentiated by the presence of an active *Wolbachia* infection that is only present in *G. m. centralis*. To support this, 85 field-collected individuals belonging to *G. m. centralis* (Angola and Tanzania), which had the same ITS1 profile, were also 100 % infected with *Wolbachia* (**Table 6**).

Table 6: *Wolbachia* status of selected *Glossina* field collections

Field collected tsetse species	*Wolbachia* status				
	Cytoplasmic			Chromosomal	
	No.	%	Estimation	No.	%
G. pallidipes	0/57	0	no PCR amplicon, no infection	0/57	0
G. m. centralis	85/85	100	strong PCR amplicons, fixed infection	0/85	0
G. p gambiensis	15/78	19.2	weak PCR amplicons, sporadic	0/78	0
G. f. fuscipes	2/52	3.8	weak PCR amplicons, sporadic	0/52	0
G. f. quanzensis	1/3	33.3	weak PCR amplicons, sporadic	0/3	0
G. brevipalpis	3/6	50	strong PCR amplicons, not fixed infection	0/6	0
G. austeni	7/7	100	strong PCR amplicons, fixed infection	0/7	0

Regarding *Wolbachia* status of the other field-collected samples, *G. austeni* was 100 % infected, *G. brevipalpis* did not show a fixed infection pattern (though with strong PCR amplicons in some of the individuals), and three other species also presented non-fixed infection patterns and with weak PCR amplicons (*G. f. fuscipes, G. f. quanzensis,* and *G. p. gambiensis*). *G. pallidipes* did not show any evidence of *Wolbachia* infection (**Table 6).**

Discussion

The main objective of this study was to develop and evaluate a convenient and cost-effective approach to identify *Glossina* species at the molecular level (i.e. PCR and gel electrophoresis). Taking together results from laboratory and field samples, the ITS1 amplicon produced eight size variants that could easily be recognized in 2.5 % agarose gel electrophoresis. The ITS1 amplicons generated from this study are in accordance with previously published ITS1 sequenced species (Dyer *et al.,* 2009; 2008; 2011) (**Table 7**).

These profiles successfully identified five species (*G. pallidipes, G. p. gambiensis, G. tachinoides, G. austeni,* and *G. medicorum*). The three remaining ITS1 profiles clustered seven taxa in three different groups. The *G. m. morsitans / G. swynnertoni / G. brevipalpis* group, the *G. m. centralis / G. m. submorsitans group,* and the *G. f. fuscipes / G. f. quanzensis* group. Some ITS1 amplicons, representing different species from the reference laboratory colonies, were sequenced to confirm the actual amplicon size (data not shown). Regarding the *G. fuscipes* subspecies, there were no well-characterized material besides *G. f. fuscipes* at the IPCL, but few field-collected individuals were available for *G. f. quanzensis* that shared the same ITS1 profile with *G. f. fuscipes.* Dyer *et al.* (2011) previously developed ITS1 diagnostic primer pairs and diagnostic assays that can differentiate among the three subspecies of *G. fuscipes* (*fuscipes, quanzensis,* and *martinii*) as shown in **Table 7**. Since there were no reference laboratory materials available for the two of the three *fuscipes* subspecies, no attempts were made to identify these taxa. For further analyses, several microsatellite markers were screened to identify some species-specific markers that could be used as diagnostic markers for those specific species. Cross-species amplification of microsatellite markers is an indication of the phylogenetic relation among different species. The more closely related species are expected to share a higher number of cross amplified microsatellite markers which can also be regarded as an indicator of their genetic proximity. This property has already been exploited in *Glossina* species for *de novo* development of markers as shown in **Table 8** (Baker and Krafsur, 2001; Brown *et al.,* 2008; Dyer *et al.,* 2008; Luna *et al.,* 2001; Ouma *et al.,* 2003; 2006; Solano *et al.,* 1997).

Table 7: ITS1 size variants in tsetse species as published in previous studies

Taxon	ITS1 size variant												Reference
	Original primer pairs							PCR assays with modified primers					
	880	778	919	633	597	618	543	~240	~240+ ~330	234+ 239	234+ 417	339	
G. medicorum	+												(Dyer *et al.*, 2008)
G. brevipalpis		+											
G. pallidipes			+										
G. austeni				+									
G. tachinoides					+								
G. f. quanzensis						+							
G. f. fuscipes						+							
G. p. gambiensis							+						
G. p. palpalis						+							
G. p. palpalis								+	+				(Dyer *et al.*, 2009)
G. f. quanzensis										+			(Dyer *et al.*, 2011)
G. f. martinii											+		
G. f. fuscipes												+	

Table 8: Microsatellite markers' cross species amplification in different *Glossina* taxa as referred in previous publications.

Microsatellite	*Glossina* species													Ref
	Gpp	*Gff*	*Gtach*	*Gms*	*Gmm*	*Gpal*	*Gswyn*	*Gaus*	*Gbrev*	*Glong*	*Gfus*	*Glong*	*Gmc*	*Gpg*	
55.3	171-175	181-185	-	-	-	nt	nt	nt	nt	nt	nt	nt	nt	nt	(Solano *et al.*, 1997)
19.62	170-174	174-182	-	-	-	nt	nt	nt	nt	nt	nt	nt	nt	nt	
69.22	194-200	192-192	-	-	-	nt	nt	nt	nt	nt	nt	nt	nt	nt	

Microsatellite	*Glossina* species													Ref
	Gpp	*Gff*	*Gtach*	*Gms*	*Gmm*	*Gpal*	*Gswyn*	*Gaus*	*Gbrev*	*Glong*	*Gfus*	*Glong*	*Gmc*	*Gpg*	
Gmm8	+	+	+	nt	125-131	+	+	+	-	-	-	nt	nt	nt	(Baker and Krafsur, 2001)
Gmm14	**-**	**-**	**-**	**nt**	**153-211**	**+**	**-**	**-**	**-**	**-**	**-**	**nt**	**nt**	**nt**	
Gmm15	-	-	-	nt	185-195	+	-	-	-	-	-	nt	nt	nt	
Gmm22	-	-	-	nt	133-145	+	-	-	-	-	-	nt	nt	nt	
Gmm5B	**-**	**-**	**-**	nt	155-175	**-**	**-**	**-**	**-**	**-**	**-**	nt	nt	nt	
Gmm9B	-	-	-	nt	140-180	-	-	-	-	-	-	nt	nt	nt	
GmsCAG16	-	-	-	nt	120-140	-	-	-	-	-	-	nt	nt	nt	
GmsCA16C	+	+	+	nt	200-210	+	+	+	+	+	+	nt	nt	nt	
GmsCAG2	+	+	-	nt	130-145	+	+	+	-	-	-	nt	nt	nt	
GmsCAG17B	+	+	+	nt	+	+	+	+	**-**	**-**	**-**	nt	nt	nt	
GmsCAG29B	-	-	-	nt	175-190	+	-	-	-	-	-	nt	nt	nt	
GpCAG133	+	+	+	nt	185-205	+	+	+	**-**	**-**	**-**	nt	nt	nt	
Gmm127	+	+	+	nt	295-301	+	+	-	-	-	-	nt	nt	nt	
Pgp1	124	+	+	nt	nt	nt	nt	-	nt	nt	nt	nt	nt	nt	(Luna *et al.*, 2001)
Pgp8	192	+	-	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp11	178	+	-	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp13	201	+	+	nt	nt	nt	nt	-	nt	nt	nt	nt	nt	nt	
Pgp17	191	+	+	nt	nt	nt	nt	-	nt	nt	nt	nt	nt	nt	

Microsatellite	*Glossina* species													Ref
	Gpp	*Gff*	*Gtach*	*Gms*	*Gmm*	*Gpal*	*Gswyn*	*Gaus*	*Gbrev*	*Glong*	*Gfus*	*Glong*	*Gmc*	*Gpg*	
Pgp20	194	+	+	nt	nt	nt	nt	-	nt	nt	nt	nt	nt	nt	
Pgp22	279	+	+	nt	nt	nt	nt	-	nt	nt	nt	nt	nt	nt	
Pgp24	215	+	+	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp28	103	+	+	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp29	237	+	+	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp33	208	+	+	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp34	364	+	+	nt	nt	nt	nt	-	nt	nt	nt	nt	nt	nt	
Pgp35	202	+	+	nt	nt	nt	nt	+	nt	nt	nt	nt	nt	nt	
Pgp38	225	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Pgp37	217	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
GpA19a	nt	+	nt	+	-	142-189	+	+	+	nt	nt	+	+	nt	
GpA23b	nt	+	nt	+	+	172-215	+	+	+	nt	nt	+	+	nt	
GpB6b	nt	+	nt	+	-	187-224	+	+	+	nt	nt	+	+	nt	
GpB20b	nt	+	nt	+	+	139-200	+	+	+	nt	nt	+	+	nt	(Ouma *et al.*, 2003)
GpC5b	nt	+	nt	+	+	187-239	+	+	+	nt	nt	+	+	nt	
GpC10b	nt	+	nt	+	+	283-314	+	+	+	nt	nt	+	+	nt	
GpC26b	nt	+	nt	+	+	168-201	+	+	+	nt	nt	+	+	nt	
GpD18b	nt	+	nt	+	-	220-229	+	+	+	nt	nt	+	+	nt	
GpB115	nt	-	nt	+	+	133-177	+	+	-	nt	nt	+	+	_	(Ouma *et al.*, 2006)
GpC101	nt	+	nt	+	+	186-230	+	+	+	nt	nt	+	+	+	

Microsatellite	*Glossina* species														Ref
	Gpp	*Gff*	*Gtach*	*Gms*	*Gmm*	*Gpal*	*Gswyn*	*Gaus*	*Gbrev*	*Glong*	*Gfus*	*Glong*	*Gmc*	*Gpg*	
GpC107	nt	+	nt	+	+	202-217	+	+	+	nt	nt	+	+	+	
A10	**-**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**nt**	**+**	(Dyer *et al.*, 2008)
Gff_B8	nt	183-217	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	(Brown *et al.*, 2008)
Gff_C107	nt	189-245	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_D6	nt	259-279	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_D109	nt	153-177	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_A3	nt	227-258	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_A6	nt	257-267	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_A9	nt	170-174	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_A112	nt	121-133	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_B101	nt	268-308	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
Gff_A10	nt	184-213	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	

Allele size or range is only given for the species where microsatellites were originally developed. +: presence of amplicon, -: absence of amplicon, nt: not tested, in bold: the two microsatellite markers selected to be included in the genotyping approach presented in this study. Abbreviations: *Gpp; G. p. palpalis, Gff; G. f. fuscipes, Gtach; G. tachinoides, Gms; G. m. submorsitans, Gmm; G. m. morsitans, Gpal; G. pallidipes Gswyn*; *G. swynnertoni, Gaus; G. austeni, Gbrev; G. brevipalpis, Gfuscip; G. fuscipleuris, Glong; G. longipennis, Gmc; G. m. centralis, Gpg; G. p. gambiensis.*

As previously reported (Dyer *et al.,* 2008), microsatellite A10 can be used to distinguish *G. p. gambiensis* from *G. tachinoides* which showed similar (but not identical) ITS1 length. Moreover, microsatellite Gmm14 can successfully differentiate *G. brevipalpis* from all other species in this study, which was crucial since it shared an identical (or similar) ITS1 profile with *G. m. morsitans* and *G. swynnertoni*. The two remaining 'black boxes' are the *G. m. morsitans / G. swynnertoni* and the *G. m. centralis / G. m. submorsitans* groups. However, based on the current (and previous) data, these species can be distinguished based on the *Wolbachia* profile. *G. m. morsitans* is up to now the only *Glossina* species that has a *Wolbachia* chromosomal insertion that gives a characteristic 16S *rRNA* amplicon of 296 bp and *G. swynnertoni* samples tested did not produce this amplicon. This fixed horizontal gene transfer of *Wolbachia* in *G. m. morsitans* laboratory colony, agreed with already published results (Brelsfoard *et al.,* 2014; Doudoumis *et al.,* 2013). Regarding the last group, *G. m. centralis* has a fixed *Wolbachia* infection (cytoplasmic), while *G. m. submorsitans* seems to lack *Wolbachia.*

Except the *G. m. centralis* that harbored a fixed *Wolbachia* infection and *G. m. morsitans* that showed a fixed chromosomal insertion, all other laboratory colonies were shown to be either *Wolbachia*-free (*G. p. gambiensis, G. m. submorsitans,* and *G. tachinoides*) or had varying levels of *Wolbachia* infection (*G. pallidipes, G. m. morsitans*, *G. f. fuscipes,* and *G. brevipalpis*). These data agree with previous studies about the *Wolbachia* infection status of laboratory colonies and natural populations of *Glossina* species (Alam *et al.,* 2012; Doudoumis *et al.,* 2013; 2017; 2012; Symula *et al.,* 2013). The presence of *Wolbachia* in some of the *G. pallidipes* flies from Ethiopia and its absence from all Uganda *G. pallidipes* flies suggests that geographical origin of a species might impact the *Wolbachia* infection status of the species. The presence or absence of *Wolbachia* infection in the same species from different geographical areas has been previously reported (Alam *et al.,* 2012; Doudoumis *et al.,* 2013; 2017; 2012); however, many of these cases are both low prevalence and low titer infections (**Table 9**). The biological, ecological and evolutionary significance of such infections remains to be resolved.

Table 9: *Wolbachia* status in different *Glossina* taxa as referred in previous publications.

Taxon	*Wolbachia*		Reference
	Cytoplasmic	Chromosomal	
G. m. morsitans	Low to fixed	Fixed	(Doudoumis *et al.,* 2012)
G. pallidipes	low	Absent	
G. austeni	Medium to fixed	Absent	
G. p. palpalis	Absent	Absent	
G. p. gambiensis	Absent to low	Absent	
G. brevipalpis	Low to medium	Absent	
G. f. fuscipes	Absent	Absent	
G. m. centralis	Fixed (small sample)	Absent	
G. f. fuscipes	Low to medium	not tested	(Alam *et al.,* 2012)

Further, sequencing of some of the mitochondrial genes supported the phylogeny of the three *Glossina* taxonomic groups. Different haplotypes within some of the analyzed *Glossina* species were revealed for the *COI* gene sequence. Although the sequencing of the mitochondrial markers showed differences among the *Glossina* species and even within populations from different geographical areas, these sequences alone could not distinguish some of the species. For instance, the *G. m. centralis H3 COI* and 16S *rRNA* gene sequences were similar to the *G. m. morsitans* sequences. Additionally, mitochondrial markers can be considered as 'compromised' in cases of closely related species. In such cases, mitochondrial haplotypes may have a completely different phylogenetic history than nuclear DNA (Alam *et al.,* 2011; Saridaki and Bourtzis, 2010) For these reasons, sequencing of mitochondrial markers was not included as a tool in the approach followed in the present study. Furthermore, the mitochondrial gene sequencing was excluded in order to keep the protocol cheap, quick, and easy to apply. Therefore, this study clearly suggests that the combined use of ITS1, selected microsatellite markers, and *Wolbachia* status (cytoplasmic infection and chromosomal introgression) provides a reliable and cost-effective approach that can be applied for the identification of many *Glossina* species without need of DNA sequencing.

Conclusions

The integration of nuclear and symbiotic markers in this study could clearly discriminate among some different economically important *Glossina* species. The correct identification at least at the species level is critical for the application of SIT and requires large numbers of individuals, especially in cases of morphologically indistinguishable subspecies and sympatric species. This study avoided using sequencing and/or specialized PCR assays to keep the identification test easy to apply, analyze and cost effective. Although there are now modern tools available that can support molecular taxonomy (genome wide sequencing for example), these tools cannot as yet be used cost effectively on numerous individuals. Therefore, the approach used in this chapter can be considered as adequate to support species identification, especially in African countries where quick decision making, and planning may be needed, depending on the data derived from trap collections. In addition, the correct tsetse species/ subspecies identification may assist the development of integrated pathogen management strategies (e.g. viruses such as SGHVs) for different tsetse species that may present variability in their susceptibility to pathogen infections (see **Chapter 4**).

Acknowledgements

The authors would like to thank Mark Vreysen for his useful comments on the manuscript and his overall support on this study. This work was funded by FAO/IAEA Joint Program of Nuclear Techniques in Food and Agriculture (Coordinate Research Project No. D42015).

Chapter 4

Hytrosavirus genetic diversity and eco-regional spread in *Glossina* species

This chapter has been modified from:

Meki, I.K., Kariithi, H.M., Ahmadi, M., Parker, A.G., Vreysen M.J.B., Vlak, J.M., van Oers, M.M. and Abd-Alla A.M.M. Hytrosavirus genetic diversity and eco-regional spread in *Glossina* species. BMC Microbiology, (*in press*).

Abstract

The management of *Glossina pallidipes* (Diptera; Glossinidae) in Africa by the sterile insect technique (SIT) has been hindered by infections of lab-bred colonies of this species with *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV; *Hytrosaviridae* family). This virus can significantly decrease productivity of the *G. pallidipes* colonies. The question is to what extent GpSGHV is present in other *Glossina* species in the field where the virus prevalence is much lower and may have diverged over time and space. In this chapter, three highly diverged genes and two variable number tandem repeat regions (VNTRs) of the GpSGHV genome were used to identify the viral haplotypes and their phylogenetic relatedness in seven *Glossina* species. These *Glossina* species were obtained from 29 African locations and their identities were confirmed using the multi-markers discussed in Chapter 3. GpSGHV was detected in all analyzed *Glossina* species using PCR. The highest GpSGHV prevalence was found in *G. pallidipes* colonized at FAO/IAEA Insect Pest Control Laboratory (IPCL) that originated from Uganda (100%) and Tanzania (88%), and a lower prevalence in *G. morsitans morsitans* from Tanzania (58%) and Zimbabwe (20%). Whereas GpSGHV was detected in 25-40% of *G. fuscipes fuscipes* in eastern Uganda, the virus was not detected in specimens of neighboring western Kenya. Most of the identified 15 haplotypes were restricted to specific *Glossina* species in distinct locations. Seven haplotypes were found exclusively in *G. pallidipes*. The reference haplotype H1, which corresponds to the Ugandan GpSGHV strain (GpSGHV-Uga), was the most widely distributed among the species as well as in different locations but was not found in *G. swynnertoni*. The 15 haplotypes clustered into three distinct phylogenetic clades, of which the largest contained seven haplotypes, that were detected in six *Glossina* species. The *G. pallidipes*-infecting haplotypes H10, H11 and H12 (from Kenya) clustered with H7 (from Ethiopia), which presumably corresponds to the recently sequenced Ethiopian GpSGHV strain (GpSGHV-Eth). These four haplotypes diverged the most from the reference H1 (GpSGHV-Uga). Haplotypes H1, H5 and H14 formed three main genealogy hubs, potentially representing the ancestors of the 15 haplotypes. These data implicate that *G. pallidipes* is a significant driver for the generation and diversity of GpSGHV variants. This information may provide control guidance when new tsetse colonies are established and hence, for improved management of the virus in tsetse rearing facilities that maintain multiple *Glossina* species.

Introduction

Management of insect vectors using the sterile insect technique (SIT) within the context of area-wide integrated pest management (AW-IPM) approaches, requires mass-production of high quality insects that must outcompete wild males for mating virgin wild females (Vreysen *et al.,* 2011). These non-viable matings eventually lead to the decline of the target insect population and reduction in the occurrence of the trypanosomosis disease they transmit to animals and human. The successful eradication of a population of the tsetse fly species *Glossina austeni* Newstead (Diptera; Glossinidae) on Unguja Island in Zanzibar, using an AW-IPM approach with an SIT component (Vreysen *et al.,* 2000) elicited efforts to apply a similar approach to eradicate *G. pallidipes* from the Southern Rift Valley region of Ethiopia (Alemu *et al.,* 2007). However, like in many insect mass-production facilities where viral diseases challenge production of high quality insects (Gouli *et al.,* 2011), infections of *G. pallidipes* colonies with the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV; *Hytrosaviridae* family) hindered full implementation of the SIT component of the Ethiopian program (Abd-Alla *et al.,* 2007; 2011).

The inbreeding in long-term colonized insects reduces their genetic diversity, i.e. fly populations become genetically more and more homozygous (Ugelvig and Cremer, 2012), which in return may promote their susceptibility to pathogen infections. In the case of GpSGHV infections in tsetse flies colonies, the virus transmission mainly horizontal, is further stimulated by the close interactions of conspecifics tsetse species, the membrane feeding regimes, and the conducive environments created by the high tsetse densities as reported in high density ecosystems of other insects such as bees mass-production facilities (López-Uribe *et al.,* 2016). Altogether these factors result in life-history trade-offs between immune and reproductive functions, which in turn contribute to reduced fitness-related traits of individual insects and colony productivity at large (Lazzaro and Little, 2009; Siva-Jothy *et al.,* 2005). In nature, where tsetse flies blood-feed on live animals, the observed GpSGHV infections are largely asymptomatic. This could be due to the low population densities and solitary habitats of different tsetse species which would prevent fly to fly contact. It could also be that the diseased (symptomatic) flies reduce the chance

of detection due to their short life span (Abd-Alla *et al.,* 2011; 2016; Kariithi *et al.,* 2013a; Sang *et al.,* 1997).

Currently, two GpSGHV strains induce distinctive pathologies in *G. pallidipes* flies, one in the mass-rearing facility at Kaliti in Ethiopia (Ethiopian strain with high SGH prevalence), and the other at the FAO/IAEA Insect Pest Control Laboratory (IPCL) in Seibersdorf, Austria (Ugandan strain with low SGH prevalence) (Abd-Alla *et al.,* 2016). It is not known why GpSGHV infections have such a devastating impact only on colonized *G. pallidipes*, despite the rearing of this species together with multiple *Glossina* species in the same tsetse production facilities. Kariithi *et al.,* (2017b) reported that GpSGHV-infected *G. m. morsitans* expressed more antiviral proteins than symptomatically infected *G. pallidipes* flies. For example, the reactive oxygen species and components of the phagocytic engulfment system were among the overexpressed proteins in *G. m. morsitans*. The expression of such antiviral proteins in *G m. morsitans* and the reduced expression these genes in symptomatically infected *G. pallidipes* indicates that this *G. pallipides* is immunocompromised in its response to GpSGHV. Being the most susceptible species to GpSGHV, one may hypothesize *G. pallidipes* as a key species that drives the evolution and the inter-species spread of GpSGHV in tsetse mass rearing facilities.

To manage the prevalence of SGH in tsetse mass-production facilities, it is necessary to understand the diversity, evolution and transmission potential of GpSGHV in *Glossina* species. Kariithi *et al*. (2013b) reported 23 GpSGHV haplotypes in wild *G. pallidipes* flies, but the virus' genetic heterogeneity found in that study was low, and without direct correlation to geographical locations. Except in *G. pallidipes*, the GpSGHV diversity in other *Glossina* species has yet to be investigated. A high prevalence of both asymptomatic and symptomatic GpSGHV infections in *G. pallidipes* may increase the potential of cross-species exposure and transmission of the virus in mass-rearing facilities where multiple tsetse species are reared. Replication of viruses in a new host species may provide opportunities for the virus to adapt and evolve into novel viral haplotypes that may be more pathogenic, due to accumulated mutations over time. For instance, Grubaugh *et al*. (2016) demonstrated that the genetic diversity of West Nile virus (WNV) depended on the mosquito species. The study demonstrated that the southern house mosquito, *Culex*

quinquefasciatus Say supported the evolution of WNV variants exhibiting greater fitness when transferred to avian hosts compared to three other *Culex* species. The study concluded that *C. quinquefasciatus* is the main engine that drives WNV evolution. Although the study involved an RNA virus, the same dynamics may apply to evolution of DNA viruses such as GpSGHV, though most likely at much slower evolutionary rates.

In this chapter, the GpSGHV genetic diversity and prevalence in seven *Glossina* species obtained from different geographical locations throughout Africa was investigated. Based on comparison of the genomes of the two pathogenic strains GpSGHV-Uga and GpSGHV-Eth (Abd-Alla *et al.,* 2016), three of the highly diverged GpSGHV genes were selected, as well as two variable number tandem repeat regions (VNTRs); these five genes were then used to construct phylograms and to search for potential ancestral origins of this hytrosavirus. The data obtained in this chapter, are important for future development of robust strategies to effectively manage GpSGHV infections and sustainably remove SGHV from tsetse mass-rearing facilities.

Materials and methods

Sample collection

Glossina flies were collected between 1994 and 1995 and between 2005 and 2015 from 29 geographical locations in eastern, southern and central African countries. The flies were collected as described by Kariithi *et al*. (2013b). Flies from seven species were analysed in this study, i.e. *G. pallidipes*, *G. morsitans morsitans*, *G. swynnertoni, G. fuscipes fuscipes*, *G. brevipalpis*, *G. palpalis palpalis* and *G. austeni* (**Table 1).** The collected samples were preserved in absolute ethanol, or propylene glycol, shipped to the IPCL of the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Seibersdorf, Austria, and stored at -20^{o}C until further analysis.

Table 1: Details of *Glossina* species sampled in different sites in Africa: samples were collected from different geographical sites in eastern, southern and central African countries for the analysis of GpSGHV prevalence and genetic diversity. Only the samples marked by a star (*) were further analysed for GpSGHV genetic diversity.

Country	Location	Species	Collection date	Latitude	Longitude	Total No.	Prevalence (%)
Field collected samples							
Uganda	Tororo	*G. f. fuscipes*	1994	0°41'34.0"N	34°10'52.0"E	17	6 (35.3%)
	Buvuma Island	*G. f. fuscipes**	1994	0°14'36.7"N	33°16'53.9"E	10	4 (40.0%)
	Kiyindi Island	*G. f. fuscipes*	1994	0°19'20.4"N	32°59'34.2"E	8	2 (25.0%)
	Bagala Island	*G. f. fuscipes*	1994	0°25'15.2"S	32°14'38.1"E	18	5 (27.8%)
Ethiopia	Arba Minch	*G. pallidipes**	2006	6°07'01.2"N	37°01'60.0"E	431	297 (68.9%)
Kenya	Mwea	*G. pallidipes**	2007	0°53'15.9"N	37°37'59.7"E	233	17 (7.3%)
	Mwea N. Park	*G. pallidipes*	2008	0°49'23.2"S	37°37'02.3"E	21	1 (4.8%)
	Katotoi	*G. pallidipes*	2007	0°42'42.7"N	34°18'57.1"E	226	0 (0.0%)
	Meru N. Park	*G. pallidipes*	2008	0°05'18.2"N	38°11'23.8"E	95	1 (1.1%)
	Kiria	*G. pallidipes*	2008	0°31'09.8"S	36°37'27.3"E	20	0 (0.0%)
	Koibos Soi	*G. pallidipes**	2008	0°09'57.9"N	36°06'20.6"E	94	19 (20.2%)
	Mogotio-Emsos	*G. pallidipes**	2008	0°01'00.4"S	35°57'32.7"E	72	14 (19.4%)
	Ruma N. Park	*G. pallidipes**	2007	0°38'44.8"S	34°16'31.8"E	176	3 (1.7%)
	Obekai	*G. f. fuscipes*	2007	0°30'52.5"N	34°12'17.6"E	38	0 (0.0%)
	Ikapolok	*G. f. fuscipes*	2007	0°37'44.9"N	34°18'38.0"E	52	0 (0.0%)
Tanzania	Kwekivu	*G. pallidipes**	2005	5°46'30.5"S	37°23'55.4"E	50	44 (88.0%)

Country	Location	Species	Collection date	Latitude	Longitude	Total No.	Prevalence (%)
Field collected samples							
		*G. m. morsitans**	2005			50	29 (58.0%)
	Kwamume	*G. pallidipes**	2005	5°41'51.9"S	37°52'01.3"E	33	1 (3.0%)
		G. m. morsitans	2005			50	0 (0.0%)
	Ikorongo GR	*G. swynnertoni**	2015	1°54'58.8"S	34°43'49.8"E	48	23 (47.9%)
	Jozani, Zanzibar	*G. austeni*	1994	6°14'28.4"S	39°24'50.3"E	29	6 (20.7%)
Zambia	Mfuwe	*G. pallidipes**	2007	13°04'41.2"S	31°47'26.5"E	201	49 (24.1%)
		*G. m. morsitans**	2007			116	9 (7.8%)
Zimbabwe	Mashumbi	*G. pallidipes**	2006	15°56'13.8"S	29°27'25.7"E	50	1 (2.0%)
		G. m. morsitans	2006			8	0 (0.0%)
	Gokwe	*G. pallidipes**	2006	17°36'14.5"S	28°27'41.1"E	150	10 (6.7%)
		G. m. morsitans	2006			92	23 (25.0%)
	Ruckomechi	*G. pallidipes**	2006	15°50'55.0"S	29°07'30.0"E	97	30 (30.9%)
		*G. m. morsitans**	2006			103	21 (20.4%)
	Makuti	*G. pallidipes*	2006	16°17'59.0"S	29°17'59.9"E	96	0 (0.0%)
		*G. m. morsitans**	2006			99	9 (9.1%)
	Mukondore	*G. m. morsitans*	1995	16°05'22.7"S	29°14'36.0"E	36	18 (50.0%)
	Chiuyi	*G. m. morsitans*	1995	16°6'31.6"S	29°24'33.8"E	36	19 (50.0%)
DRC	Malanga	*G. p. palpalis**	1995	5°33'26.6"S	14°21'00.1"E	52	4 (7.7%)
South Africa	Zululand	*G. brevipalpis**	1995	28°01'07.2"S	32°12'52.6"E	33	5 (15.2%)
		*G. austeni**	1999			53	14 (26.4%)
Laboratory colonised Glossina species							
Uganda	Tororo (IPCL)	*G. pallidipes**	2010	0°41'34.0"N	34°10'52.0"E	48	48 (100.0%)
Kenya	BioRI-KALRO	*G. pallidipes*	2008	1°13'28.0"S	36°38'10.2"E	99	1 (1.0%)
		*G. m. morsitans**	2008			89	16 (17.9%)
Total						**3229**	

DNA extraction, polymerase chain reaction (PCR) and gel electrophoresis

Total DNA was extracted from whole fly bodies of 3,229 individuals of the above-mentioned species using the DNeasy Tissue Kit (QIAGEN Inc, Valencia, CA) following the manufacturer's instructions. PCR amplifications were performed as previously described (Kariithi *et al.,* 2013b). Briefly, final PCR reaction volumes of 25 µl were used containing 12.5 µl of *Taq* PCR Master Mix (*Taq* PCR Master Mix Kit, QIAGEN Inc.), ~50 ng of the isolated DNA template, 1 µl of forward and reverse primers to a final concentration of 0.2 mM per primer and 10 µl of RNase-free water. PCR products were analysed by 1.5% agarose gel electrophoresis according to standard protocols.

Verification of the taxonomic status of tsetse species

Taxonomic status of the *Glossina* flies, determined initially on visual identification in the field, were analysed (eight samples/species) using the optimized multi-markers described in Chapter 3. Briefly, the markers consisted of PCR-based sequencing of the non-coding internally transcribed spacer-1 (ITS1) of the ribosomal DNA (rDNA), and on *Wolbachia* diagnosis. The ITS1 sequence provides differences in the PCR product lengths produced with different tsetse species. Diagnosis of *Wolbachia* infection was applied to further verify the tsetse species, i.e. PCR-detection for the presence or absence of this endosymbiont (Schneider *et al.,* 2013). Laboratory tsetse flies of known taxonomic status (i.e. obtained from the IPCL) were used as positive controls during the species identification by the above-mentioned molecular markers.

Determination of GpSGHV prevalence

To determine GpSGHV prevalence in the randomly collected tsetse samples, PCRs were performed to amplify partial sequences of two conserved viral genes, *odv-e66* (SGHV005) and *dnapol* (SGHV079) using sets of primers as described by Abd-Alla *et al*. (2007). Samples were considered virus-infected if the expected PCR products of at least one of the two viral genes were detected (**Table 2**). The *Glossina* species microsatellite GpCAG133 was used to control the quality of the extracted DNA and the PCR amplifications.

Table 2A: Primers used to determine the prevalence of GpSGHV: The expected PCR products of the selected viral genes are shown.

Target gene	Forward and Reverse primer sequences (5' to 3')	Amplicon size (bp)	Ref.
GpCAG133	GpCAG133F-ATTTTTGCGTCAACGTGA GpCAG133R-ATGAGGATGTTGTCCAGTTT	180-220	(Abd-Alla *et al.,* 2007)
GpSGHV (ODV-*e66*)	GpSGHV_2F- CTTGTCAGCGCCACGTACAT GpSGHV_2R- GCATTCACAGCATCCCAATTTT	401	
GpSGHV (DNA-*pol*)	83F_GTACATATTCGAATGTATTTGCCGTTGCTC 82R_CGGGAGGAGTTGTAATACCCTGTATCAAAG	320	

Table 2B: Primers used for identification of *Glossina* species: Primer sequences used to amplify the selected markers and the expected PCR products.

Target gene	Forward and Reverse primer sequences (5' to 3'	Amplicon size (bp)	Ref.
Nuclear marker (ITS1)	*Glossina*ITS1_for- GTGATCCACCGCTTAGAGTGA *Glossina*ITS1_rev- GCAAAAGTTGACCGAACTTGA	Variable (Species-specific)	(Dyer *et al.,* 2008)
Wolbachia infection status (16S rRNA)	WspecF- YATACCTATTCGAAGGGATAG WspecR- AGCTTCGAGTGAAACCAATTC	438+(296)	(Doudoumis *et al.,* 2012)

Table 2C: List of primers used to determine the genetic diversity of GpSGHV: The expected PCR product and the variations (SNPs, deletions or insertion) between the Ugandan and Ethiopian strains are indicated.

GpSGHV-Uga		GpSGHV-Eth		For and Rev primer sequences	Amplicon size (Eth/Ug) bp	SNPS (bp)	Deletion (bp)	Insertion (bp)
ORF	**Position**	**ORF**	**Position**					
SGHV 009	8631> 10868	SGHV Eth008	8634> 10868	8F_ TTTCCTCCAATTCTTCTCTGGCAGC 8R_ CCACGTCAATGTTGCCTTTCAAATC	1433/1436	22	38	35
SGHV 010	14205< 10894	SGHV Eth009	14184< 10894	11F_GCCGTTTCTTTTCTAATTTCTTCATCTTCGGG 11R_GCTCAATAGTTTAAAGCACTGTAACCGCGTTGATT	1631/1655	35	24	0
SGHV 038	44374< 40853	SGHV Eth039	44292< 40702	32F_ACGCTGAACTAAATTATCGTCATCTACACG 32R_GCACCAATTGAACATGGATTCCGTTAT	1623/1626	28	3	0
VNTR -1	22536> 22814	VNTR -1	22617> 22830	18F_TGGCCCAGCCCTAAATATCTTAATAGCG 17R_CAAAGCTGGGCCATATATTGGGTAGAAATT	511/682	22	170	0
VNTR -2	73504> 73727	VNTR -2	73389> 73702	R2-Nested3F_ GATACGTCTCACTCATACAATC R2-Nested3R_ CATATTACCGACAGAGGGCGTTCAC	812/707	42	0	105

PCR product purification and sequencing

To determine GpSGHV genetic variation in the GpSGHV-positive flies (marked with asterisks (*) in **Table 1**), three putative open reading frames (ORFs) and two VNTRs were selected, based on the differences in the genomes of the virus from Uganda (GpSGHV-Uga; Accession Number: EF568108) and from Ethiopia (GpSGHV-Eth; Accession Number: KU050077) (Abd-Alla *et al.,* 2008; 2016). The selected ORFs were: SGHV009, SGHV010, SGHV038, and the GpSGHV VNTR-1 and VNTR-2 (corresponding to R1 and R2) loci as described in Abd-Alla *et al.* (2008). PCRs were performed as described above with primers shown in **Table 2**. The PCR amplification conditions for ORFs SGHV009, SGHV010 and SGHV038 were, 5 min at 95°C, 35 cycles of 94°C for 45s, 60°C for 45s and 72°C for 2 min, then 72°C for 10 min. PCR cycling conditions for the VNTR-1 and VNTR-2 were, 5 min at 95°C, 35 cycles of 94°C for 45s, 45°C for 45s and 72°C for 1 min, followed by 72°C for 10 min. PCR products were subsequently purified using the QIAquick PCR purification Kit (Qiagen, Valencia, CA), and sequenced from both ends by the Sanger method (Eurofins Genomics, Ebersberg, Germany) using their respective primer sets.

Phylogenetic analysis

DNA sequence reads from the sequenced PCR products were assembled and aligned using the SeqMan Pro (Lasergene 14, DNASTAR, Inc.). Only sequences with good quality reads in the chromatograms were used for further analyses. Single nucleotide polymorphisms (SNPs), deletions or insertions of sequences were determined based on the GpSGHV-Uga genome as the reference. The ORF Finder platform (https://www.ncbi.nlm.nih.gov/orffinder/) was used to identify the ORF of SGHV009, SGHV010 and SGHV038 viral genes for all the samples. The nucleotide sequences were translated using the BioEdit program (Hall, 2011) to identify the synonymous and non-synonymous mutations. Both nucleotide and the translated amino acid sequences of individual viral genes were used to determine phylogenetic relationships amongst the GpSGHV haplotypes infecting the seven *Glossina* species. Here, a haplotype is defined as a population of closely related genetic variants resulting from mutations events. The sequences were aligned and trimmed using

ClustalW on MEGA6 using default settings (Tamura *et al.,* 2013). Concatenated sequences of the three genes and the two VNTRs were used for phylogenetic analyses using Maximum-Likelihood (ML) based on the General Time Reversible (TR) model with gamma distributed rates (Rodríguez *et al.,* 1990) with 1000 bootstrap replications. Samples that could not be sequenced in all the selected genes and VNTRs were marked with '?' to indicate missing sequence in the concatenated sequence alignment. It should be noted that attempts to PCR-amplify these samples using flanking primer sets also failed. To determine the number of GpSGHV haplotypes present amongst the *Glossina* populations, samples presenting the same sequence were categorised as a single haplotype.

Estimation of gene genealogies

The Arlequin software version 3.5 (Excoffier *et al.,* 2005) was used to compare the genetic differences of the haplotypes and their relationships (mutation events between the haplotypes). The Arlequin program was used to visualize the mutational events i.e. the number of mutational differences (single nucleotide polymorphisms and deletions/insertions) between two haplotypes. The deletions or insertions that occurred as a block of repeat unit in the VNTRs sequences were interpreted as one mutation event. The Arlequin output files were used to visualize the haplotype network on the HapStar program version 0.7 (Teacher and Griffiths, 2011). Hapstar uses a spring model algorithm by automatic repulsion of disconnected haplotype branch nodes and the connected ones to an optimal format. The haplotype network was then exported as a scalable vector graphics (SVG) and loaded into Inkscape graphics editor software v 0.92.1 (Bah, 2011) for additional text, colours and patterns.

Results

Prevalence of GpSGHV infection in wild Glossina species

GpSGHV prevalence was tested for 3,229 flies collected from 29 geographical locations and belonging to seven tsetse species (**Table 1**). The GpSGHV prevalence of some of the *G. pallidipes* individuals used in the current study was previously reported by Kariithi *et al.* (2013b). In this study, the number of sampling locations

were extended and other *Glossina* species were included. GpSGHV prevalence was determined by PCR amplifications of two conserved viral genes the *odv-e66* (SGHV005) and *dnapol* (SGHV079), which were not applied in the genetic diversity analysis. Morphological identification of tsetse species is challenging and sometimes inaccurate. Therefore, on the samples that were positive for GpSGHV infection, PCR-generated ITS1 amplicons were used (as described in **Chapter 3**) to assess the taxonomic status of the seven-tsetse species (**Table 1**). In the current study, the ITS1 PCR products were of the expected sizes, which confirmed the status of *G. pallidipes* (920 bp), *G. f. fuscipes* (618 bp), *G. brevipalpis* (778 bp), *G. p. palpalis* (618 bp) *and G. austeni* (633 bp) (**Figure 1A**).

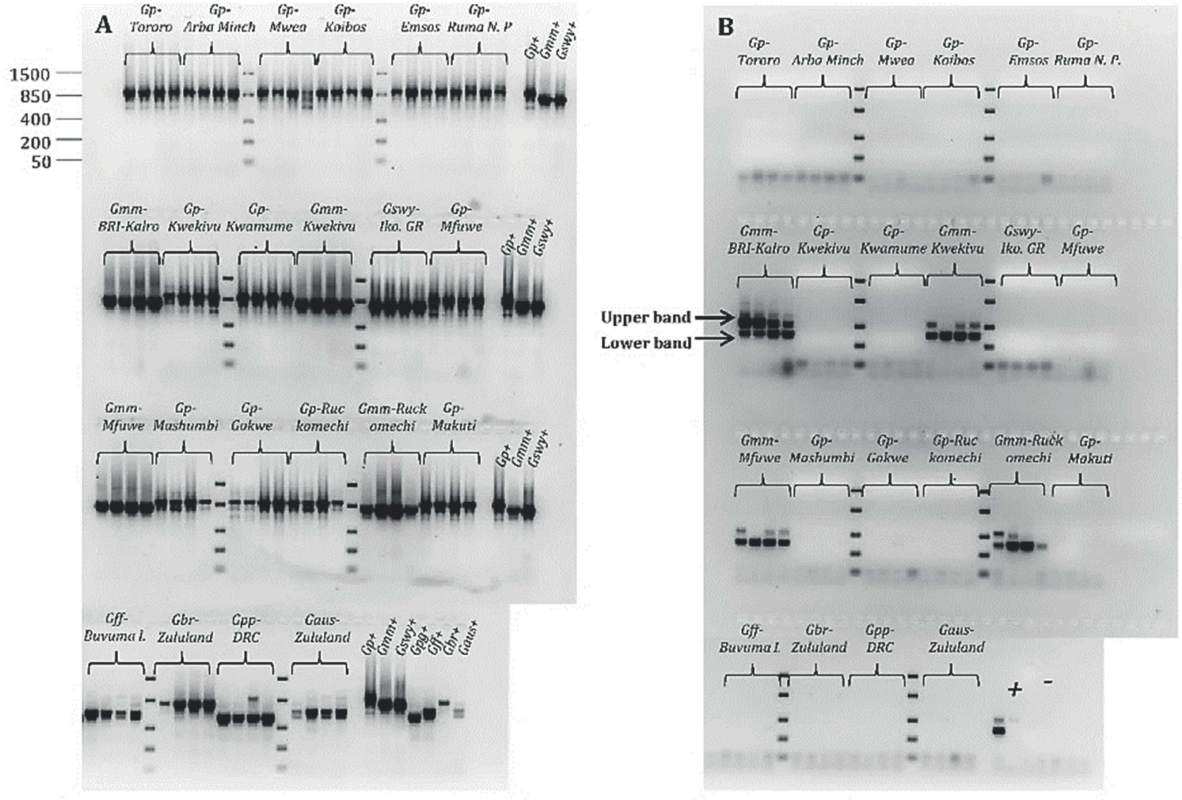


Figure 1: Analysis of PCR-amplification products on 1.5% agarose gels: Panels A and B show the different band sizes of ITS1 sequences and *Wolbachia* diagnosis for different *Glossina* species, respectively. Four samples were analysed to represent each species from different geographical regions. The positive controls were *Glossina* species from the IPCL laboratory colonies whose identities are known. The sizes of the DNA ladder bands used for both gel images are indicated. Abbreviations: Gp; *G. pallidipes*, Gmm; *G. m. morsitans*, Gswy; *G. swynnertoni*, Gff; *G. f. fuscipes*, Gbr; *G. brevipalpis*, Gpp; *G. p. palpalis*, and Gaus; *G. austeni*.

These sizes were consistent with ITS1 PCR product sizes of known tsetse species reared at the IPCL as reported in Chapter 3. The *G. m. morsitans* and *G. swynnertoni* with equal ITS1 sequence lengths (775 bp) were distinguished by the presence of the endosymbiont *Wolbachia* in *G. m. morsitans*, and its absence in *G. swynnertoni* (**Figure 1B**). *Wolbachia* integration and prevalence in *G. m. morsitans* was evidenced by the presence of lower (296bp) and the upper (438bp) bands on the agarose gel (**Figure 1B**). The variations in the *Wolbachia* prevalence agreed with previous studies in field and laboratory tsetse populations (Doudoumis *et al.,* 2013).

The GpSGHV prevalence was highest in the *G. pallidipes* colonized at the IPCL that originated from Tororo, (Uganda; 100%), followed by Kwekivu (Tanzania; 88%), Arba-Minch (Ethiopia; 68.9%), and Ruckomechi (Zimbabwe; 30.9%) (**Table 1**). The virus was not detected in *G. pallidipes* populations from Kiria and Katotoi in Kenya. The prevalence of the virus in *G. m. morsitans* flies was highest in the field-collected samples from Kwekivu (Tanzania; 58%), Chiuyi and Mukondore (Zimbabwe; 52% and 50%, respectively) compared to the laboratory colonised *G. m. morsitans* from BRI-KALRO (17.9%). The virus prevalence varied widely amongst the Ugandan *G. f. fuscipes* specimens (40%, 35.3%, 27% and 25% in Buvuma, Tororo, Bagala Island and Kiyindi Island, respectively), whereas the virus was not detected in the Kenyan *G. f. fuscipes* specimens. The virus prevalence was likewise high (47.9%) in the *G. swynnertoni* specimens from the Ikorongo Game Reserve in Tanzania, but lower in the populations of *G. austeni* (26.4%) and *G. brevipalpis*, (15.2%) from KwaZulu Natal, South Africa, and in specimens of *G. p. palpalis* from Malanga, Democratic Republic of the Congo (7.7%) (**Table 1**). These results provide evidence that the GpSGHV is present in multiple tsetse species, to varying degrees under laboratory and field conditions.

Geographical distribution of GpSGHV haplotypes

The VNTR-2 sequences, which were successfully obtained from all individual GpSGHV-positive samples in the seven-tsetse species revealed 14 viral haplotypes (**Figure 3A**). Sequence analysis using the alignments of the concatenated sequences of the three conserved genes and the two VNTRs revealed 15 GpSGHV haplotypes due to sequence differences found in SGHV009 of *G. pallidipes* samples from Mwea, Kenya (**Figure 3B**). However, the relationship between the other 14

haplotypes did not change when the analysis was performed using either the VNTR-2 or concatenated sequences. SGHV009, SGHV010 and SGHV038 and the VNTR-1 of some of the samples could not be sequenced due to failure to amplify the region with the PCR conditions used in the current study. VNTR-2 was the only successfully sequenced region in all the representative samples. The distribution of the identified haplotypes over the *Glossina* species varied depending on the geographical locations (**Table 3**). Haplotype H1, which corresponded to the reference GpSGHV-Uga strain (Abd-Alla *et al.,* 2008), was found in all examined *Glossina* species except in *G. swynnertoni*. This haplotype was found in tsetse populations sampled in 10 different locations in seven of the eight countries (except in Ethiopia) (see **Table 3**). Seven of the 15 GpSGHV haplotypes (H2, H3, H6, H9, H10, H12 and H15) were restricted to *G. pallidipes* specimens in specific locations. Haplotype H5 was detected in *G. pallidipes* and *G. m. morsitans* from Kwekivu, Tanzania, in *G. m. morsitans* from Ruckomechi, Zimbabwe and from Mfuwe, Zambia (**Table 3**). Haplotypes H7, H8 and H11 were each detected in *G. pallidipes* specimens from two distinct locations, while H13 was restricted to Zambian *G. pallidipes* and *G. m. morsitans* specimens. Haplotype H14 was restricted to the *G. swynnertoni* samples from the Ikorongo Game Reserve in Tanzania. Notably, in some cases, the same tsetse species (but not the same individual flies) from the same geographical locations harboured more than one haplotype. For example, *G. pallidipes* from Kenya, were infected with H10 and H11 (Koibos-Soi), and H11 and H12 (Emsos). Similar observations were made in the haplotypes infecting *G. pallidipes* from IPCL originated from Uganda (H1, H2, H3 and H4) (**Table 3**).

Table 3: Descriptions of the 15 GpSGHV haplotypes (abbreviated by 'H') identified in *Glossina* species from different geographical locations: The numbers between the brackets in columns 4 to 8 indicate the number of samples presenting a particular haplotype. Column 10 shows the total number of haplotypes found in each tsetse species in different locations, and the abbreviated names of the GpSGHV haplotypes in this column are indicated in the brackets.

Country	Location	Species	Tested flies	Haplotype (No. of tested samples occurring in the haplotype)					No. of haplotypes per *Glossina* species (haplotype name)
				VNTR-1	VNTR-2	SGHV009	SGHV010	SGHV038	
Uganda	Tororo	*G. pallidipes*	8	H1(8)	H1(5), H2(1), H3(1), H4(1)	H1(8)	H1(8)	H1(8)	4(H1, H2, H3, H4)
	Buvuma Island	*G. f. fuscipes*	9	H1(6)	H1(9)	-	-	-	1(H1)
Ethiopia	Arba Minch	*G. pallidipes*	8	H7(8)	H7(8)	H7(8)	H7(8)	H7(8)	1(H7)
Kenya	Mwea	*G. pallidipes*	5	H9(1)	H9(5)	H9(2), H15(1)	H9(3)	H9(3)	2(H9, H15)
	Koibos Soi	*G. pallidipes*	8	-	H10(7), H11(1)	H11(4)	H11(4)	H11(4)	2(H10, H11)
	Emsos	*G. pallidipes*	8	-	H11(7), H12(1)	H11(2)	H11(2)	H11(2)	2(H11, H12)
	BioRI-KALRO	*G. m. morsitans*	16	H1(12)	H1(16)	-	-	-	1(H1)
	Ruma N. Park	*G. pallidipes*	1	-	H8(1)	-	H8(1)	H8(1)	1(H8)
Tanzania	Kwamume	*G. pallidipes*	1	H6(1)	H6(1)	-	-	-	1(H6)
	Kwekivu	*G. pallidipes*	8	H5(4)	H5(8)	H5(6)	H5(6)	H5(6)	1(H5)
		G. m. morsitans	17	H1(1)	H1(6), H5(11)	H1(4)	H1(4)	H1(4)	2(H1, H5)

Country	Location	Species	Tested flies	Haplotype (No. of tested samples occurring in the haplotype)					No. of haplotypes per *Glossina* species (haplotype name)
				VNTR-1	VNTR-2	SGHV009	SGHV010	SGHV038	
	Ikorongo GR	*G. swynnertoni*	3	-	H14(3)	H14(3)	H14(3)	H14(3)	1(H14)
Zambia	Mfuwe	*G. pallidipes*	4	-	H13(4)	H1(4)	H1(4)	H1(4)	2(H1, H13)
		G. m. morsitans	9	H1(7)	H13(6), H1(1), H4(1), H5(1)	-	-	-	4(H1, H4, H5, H13)
Zimbabwe	Mashumbi	*G. pallidipes*	1	-	H1(1)	-	-	-	1(H1)
	Gokwe	*G. pallidipes*	1	-	H8(1)	-	-	-	1(H8)
	Ruckomechi	*G. pallidipes*	8	H1(3)	H1(7), H7(1)	H1(3)	H1(3)	H1(3)	2(H1, H7)
		G. m. morsitans	14	H1(12)	H1(13), H5(1)	H1(3)	H1(3)	H1(3)	2(H1, H5)
	Makuti	*G. m. morsitans*	8	-	H1(8)	H1(1)	H1(1)	H1(1)	1(H1)
DRC	Malanga	*G. p. palpalis*	4	H1(3)	H1(4)	-	-	-	1(H1)
South Africa	Zululand	*G. brevipalpis*	5	H1(4)	H1(5)	-	-	-	1(H1)
		G. austeni	6	H1(2)	H1(5), H4(1)	-	-	-	2(H1, H4)

Single nucleotide polymorphisms (SNPs) and nonsynonymous mutations among GpSGHV haplotypes

After analysing the GpSGHV haplotype distribution amongst the *Glossina* species from the different geographical locations (**Table 4**), synonymous and nonsynonymous single nucleotide polymorphisms (SNPs) were determined in the amplified nucleotide sequences. The analysis revealed a high rate of SNPs, and a few deletions and insertions (**Table 4**). Haplotype H1, which is similar to the reference GpSGHV-Uga did not harbour any mutations. Most of the deletions and insertions were found in the VNTRs, with the VNTR-2 presenting most of the inter- and intra-haplotype variations. The VNTRs mutations observed were mostly patterns in repeat polymorphisms, i.e. additions or deletions of repeat units, rather than single nucleotide mutations (**Figure 2A-F**).

VNTR-2 of H7, which infected the Ethiopian (Arba Minch) and Zimbabwean (Ruckomechi) *G. pallidipes* populations, harboured the most intra-haplotype variations (i.e. 24 SNPs, and 126 bp and 20 bp insertions and deletions, respectively) (**Table 4**). Compared to the GpSGHV-Uga reference sequence, most of the nonsynonymous mutations of 19 of 178, 12 of 475 and 10 of 489 amino acids in SGHV009, SGHV010 and SGHV038, respectively, were found in the H7 infecting the *G. pallidipes* populations from Ethiopia (Arba Minch) and Zimbabwe (Ruckomechi) (**Table 4**). Further, H10, H11 and H12 infecting *G. pallidipes* populations from Kenya (Koibos Soi and Emsos) also presented high mutational variations within the haplotype, with five of 178, 12 of 475 and nine of 489 amino acids of nonsynonymous mutations in SGHV009, SGHV010 and SGHV038, respectively. The greatest number of nonsynonymous mutations was found in the haplotypes infecting *G. pallidipes* populations, followed by those infecting the *G. m. morsitans* populations (see **Table 4**).

Table 4: Analysis of PCR product lengths in nucleotides, single nucleotide polymorphisms (SNPs), insertion and deletions detected in the GpSGHV haplotypes in the *Glossina* species: The polymorphisms were based on the partial sequences of GpSGHV ORFs SGHV009, SGHV010, SGHV038, VNTR-1 and VNTR-2. The 3 numbers in brackets for each haplotype in column four to column eight refer to the number of SNPs, insertions and deletions respectively, relative to the GpSGHV-Uga reference sequences. The samples marked with a question mark (?) were not successfully sequenced. The nonsynonymous mutations of SGHV009, SGHV010, SGHV038 found in each sample are also shown in the last column.

Haplotype	Species	Location	PCR Length: (SNPs, insertion, deletion) (bp)					No. of nonsynonymous mutations (aa)
			VNTR-1	VNTR-2	SGHV009	SGHV010	SGHV038	(SGHV009, SGHV010, SGHV038)
GpSGHV-Uga	Reference	NCBI	370	437	535	1425	1461	(0,0,0)
H1	*G. pallidipes*	Tororo	370: (0,0,0)	437: (0,0,0)	535: (0,0,0)	1425: (0,0,0)	1461: (0,0,0)	(0,0,0)
	G. m. morsitans	Ruckomechi						
	G. m. morsitans	Mfuwe						
	G. m. morsitans	Makuti						
	G. pallidipes	Ruckomechi						
	G. pallidipes	Mashumbi						
	G. m. morsitans	BioRI-KALRO						
	G. m. morsitans	Kwekivu						
	G. f. fuscipes	Buvuma Island						
	G. p. palpalis	Malanga						

Haplotype	Species	Location	PCR Length: (SNPs, insertion, deletion) (bp)					No. of nonsynonymous mutations (aa)
			VNTR-1	VNTR-2	SGHV009	SGHV010	SGHV038	(SGHV009, SGHV010, SGHV038)
	G. brevipalpis *G. austeni*	Zululand Zululand						
H2	*G. pallidipes*	Tororo	370: (0,0,0)	500: (2,63,0)	535: (0,0,0)	1425: (0,0,0)	1461: (0,0,0)	(0,0,0)
H3	*G. pallidipes*	Tororo	370: (0,0,0)	500: (0,63,0)	535: (0,0,0)	1425: (0,0,0)	1461: (0,0,0)	(0,0,0)
H4	*G. pallidipes* *G. m. morsitans* *G. austeni*	Tororo Mfuwe Zululand	370: (0,0,0)	374: (0,0,63)	535: (0,0,0)	1425: (0,0,0)	1461: (0,0,0)	(0.0,0)
H5	*G. pallidipes* *G. m. morsitans* *G. m. morsitans* *G. m. morsitans*	Kwekivu Kwekivu Ruckomechi Mfuwe	416: (12,66,20)	542: (15,105,0)	538: (1,3,0)	1425: (4,0,0)	1467: (12,6,0)	(1,1,4)
H6	*G. pallidipes*	Kwamume	415: (13,65,20)	458: (14,21,1)	?	?	?	?
H7	*G. pallidipes* *G. pallidipes*	Arba Minch Ruckomechi	272: (8,1,99)	542: (24,126,20)	571: (14,36,0)	1425: (16,0,0)	1458: (26,0,3)	(19,12,10)
H8	*G. pallidipes* *G. pallidipes*	Gokwe Ruma N. Park	?	483: (47,46,0)	?	1425: (14,0,0)	1460: (14,0,0)	(?,9,2)

Haplotype	Species	Location	PCR Length: (SNPs, insertion, deletion) (bp)					No. of nonsynonymous mutations (aa)
			VNTR-1	VNTR-2	SGHV009	SGHV010	SGHV038	(SGHV009, SGHV010, SGHV038)
H9	*G. pallidipes*	Mwea	431: (11,61,0)	521: (30,84,0)	523: (7,0,12)	1425: (11,0,0)	1460: (20,0,1)	(3,7,5)
H10	*G. pallidipes*	Koibos Soi	?	458: (40,21,0)	529: (6,3,9)	1425: (15,0,0)	1461: (21,0,0)	(5,12,9)
H11	*G. pallidipes* *G. pallidipes*	Koibos Soi Emsos	?	416: (36,0,21)	529: (6,3,9)	1425: (15,0,0)	1461: (21,0,0)	(5,12,9)
H12	*G. pallidipes*	Emsos	?	416: (35,0,63)	529: (6,3,9)	1425: (15,0,0)	1461: (21,0,0)	(5,12,9)
H13	*G. pallidipes* *G. m. morsitans*	Mfuwe Mfuwe	?	374: (3,0,63)	535: (0,0,0)	1425: (0,0,0)	1461: (0,0,0)	(0,0,0)
H14	*G. swynnertoni*	Ikorongo GR	?	503: (36,67,0)	532: (3,0,3)	1425: (14,0,0)	1467: (16,6,0)	(1,9,6)
H15	*G. pallidipes*	Mwea	431: (11,61,0)	521: (30,84,0)	532: (3,3,6)	1425: (11,0,0)	1460: (20,0,1)	(2,7,5)

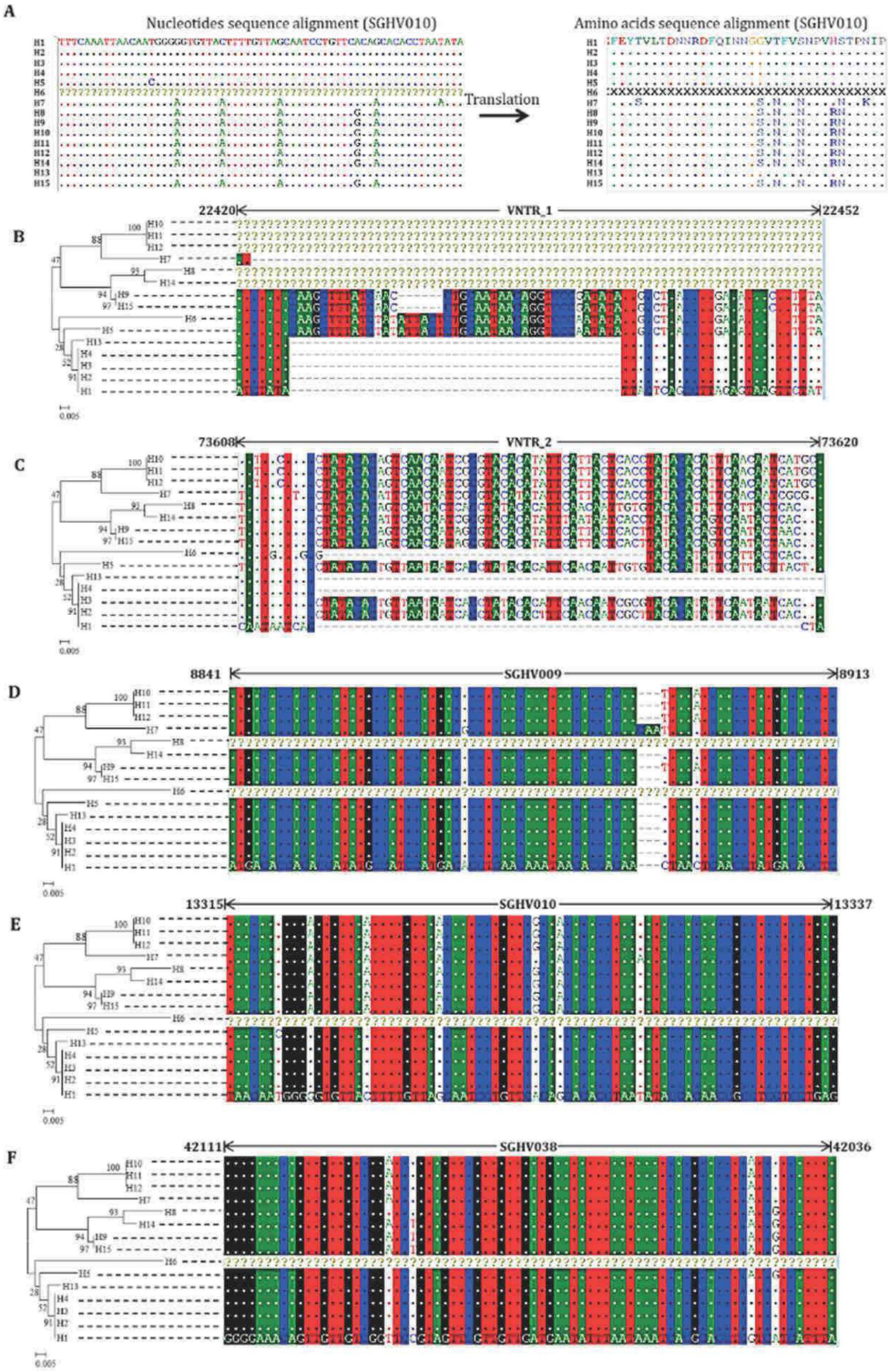


Figure 2: Sequence alignment of GpSGHV haplotypes in *Glossina* species: A). Partial alignment of SGHV010 sequences showing the single nucleotide mutations that lead to non-synonymous mutations. Multiple alignment of VNTR-1 (B), VNTR-2 (C), SGHV009 (D), SGHV010 (E) and SGHV038 (F), showing different GpSGHV haplotypes and their phylogenetic relatedness. The positions are based on the reference sequence (GpSGHV-Uga). Abbreviations; H= Haplotype.

Phylogenetic analysis of GpSGHV haplotypes

Analysis of the phylogenetic relationships of the various GpSGHV haplotypes revealed three distinct clades consisting of haplotypes from different locations (**Figure 3A and B**). The largest clade (clade 1) consisted of haplotypes H1, H2, H3, H4, H5, H6 and H13 infecting six out of seven *Glossina* species derived from most of the geographical locations. Clade 2 consisted of H8, H9, H14 and H15, which infected *G. pallidipes* from Mwea, Ruma (Kenya), Gokwe (Zimbabwe) and *G. swynnertoni* from the Ikorongo Game reserve (Tanzania). The third clade included H7, H10, H11 and H12 infecting *G. pallidipes* samples from Koibos-Soi, Emsos (Kenya), Arba Minch (Ethiopia) and Ruckomechi (Zimbabwe) (**Figure 3**).

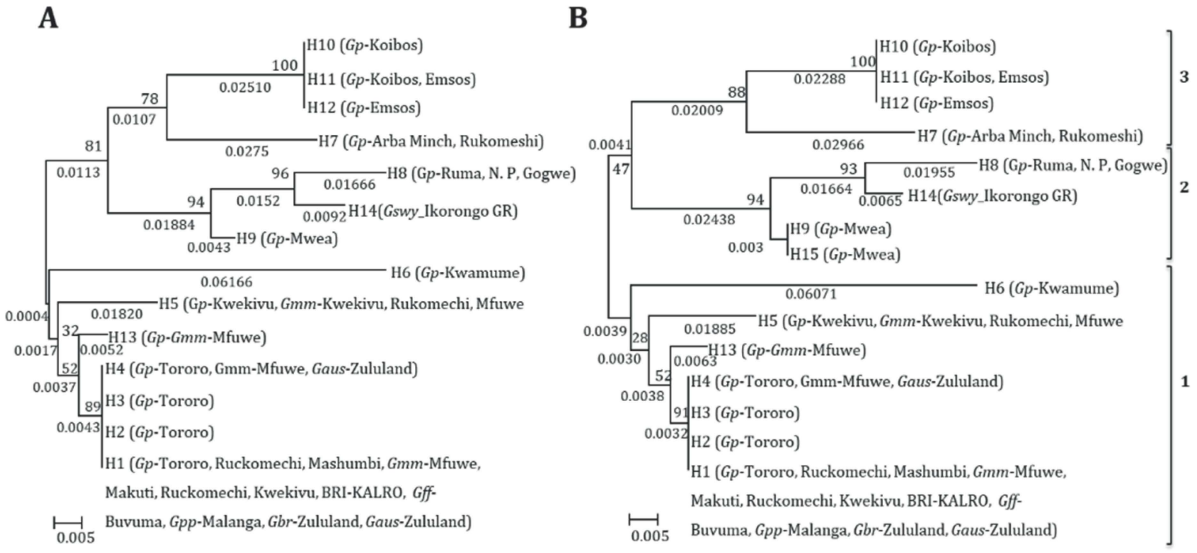


Figure 3: GpSGHV haplotypes in *Glossina* species: Maximum Likelihood (ML) phylogenetic tree for the GpSGHV strains from different geographical locations in Africa using (A) VNTR-2 and (B) concatenated sequences of VNTR-1, VNTR-2, ORF009, ORF010 and ORF038. ML bootstrap values based on 1,000 replicates are shown on the branches. Abbreviations: Gp (*G. pallidipes*), Gmm (*G. m. morsitans*), Gswy (*G. swynnertoni*), Gff (*G. f. fuscipes*), Gbr (*G. brevipalpis*), Gpp (*G. p. palpalis*), and Gaus (*G. austeni*).

The clustering of the Kenyan haplotypes (H10, H11 and H12, infecting *G. pallidipes*) in clade 3, which was supported by 100% bootstrap values, appeared to be closely related to the Ethiopian H7. H7 is presumably the GpSGHV-Eth strain, whose

genome has been fully sequenced, and which might be more pathogenic than the reference GpSGHV-Uga strain (H1) (Abd-Alla *et al.,* 2016).

Potential ancestry of GpSGHV haplotypes

To gain insights into the evolutionary history of GpSGHV, which can be seen as a series of mutation events leading to the various haplotypes, the genealogies (or gene trees) (Hudson, 1990) using both VNTR-2 alone (**Figure 4A**) and the concatenated gene sequences of the above-mentioned three genes and the two VNTRs were analysed (**Figure 4B**). The haplotype genealogies did not differ when analysed using either the VNTR-2 alone or using the concatenated sequences of all the five candidate genes. In order to include also H15, which is similar to H9 in VTNR-2 sequence and only differ at SGHV009 sequence, the concatenated sequences were used for the analysis. The topology of the star-like genealogies revealed three potential ancestral origins of the various GpSGHV haplotypes (**Figure 4**), which were largely in agreement with the clustering observed in the phylogeny (**Figure 3**). Due to its large host range and wide geographical representations, H1 was presumed to be the ancestral origin of all the 15 haplotypes (i.e. shared parental DNA sequences). This would be in line with the wide distribution of H1 described above (in six of the seven *Glossina* species, originating from 10 of geographical locations (**Figure 3**). Further, H6 (infecting *G. pallidipes* only), H4 (infecting *G. pallidipes*, *G. m. morsitans* and *G. austeni*), and H13 (infecting *G. pallidipes* and *G. m. morsitans*) potentially trace back to H1 by eight and 11 mutation events, respectively. H5 was the second potential ancestral haplotype origin. Notably, based on the analysed genes in this study, H2 and H3 (infecting the IPCL *G. pallidipes* populations) presumably coalesced from H1 and H5 by 11 and 10 mutation events, respectively (Figure 4; see also **Table 4**). On the other hand, H14, which infects the *G. swynnertoni* from the Ikorongo Game reserve in Tanzania, can be interpreted to be the ancestral origin of H7, H8, H9, and H10. Additionally, H9 and H15 (**Figure 4B**) potentially coalesced from H5 and H14 by 7-8 and 15-16 mutation events, respectively. Finally, H10, H11 and H12 from Kenya showed the highest divergence from H1, which as mentioned above, corresponded to the reference GpSGHV-Uga.

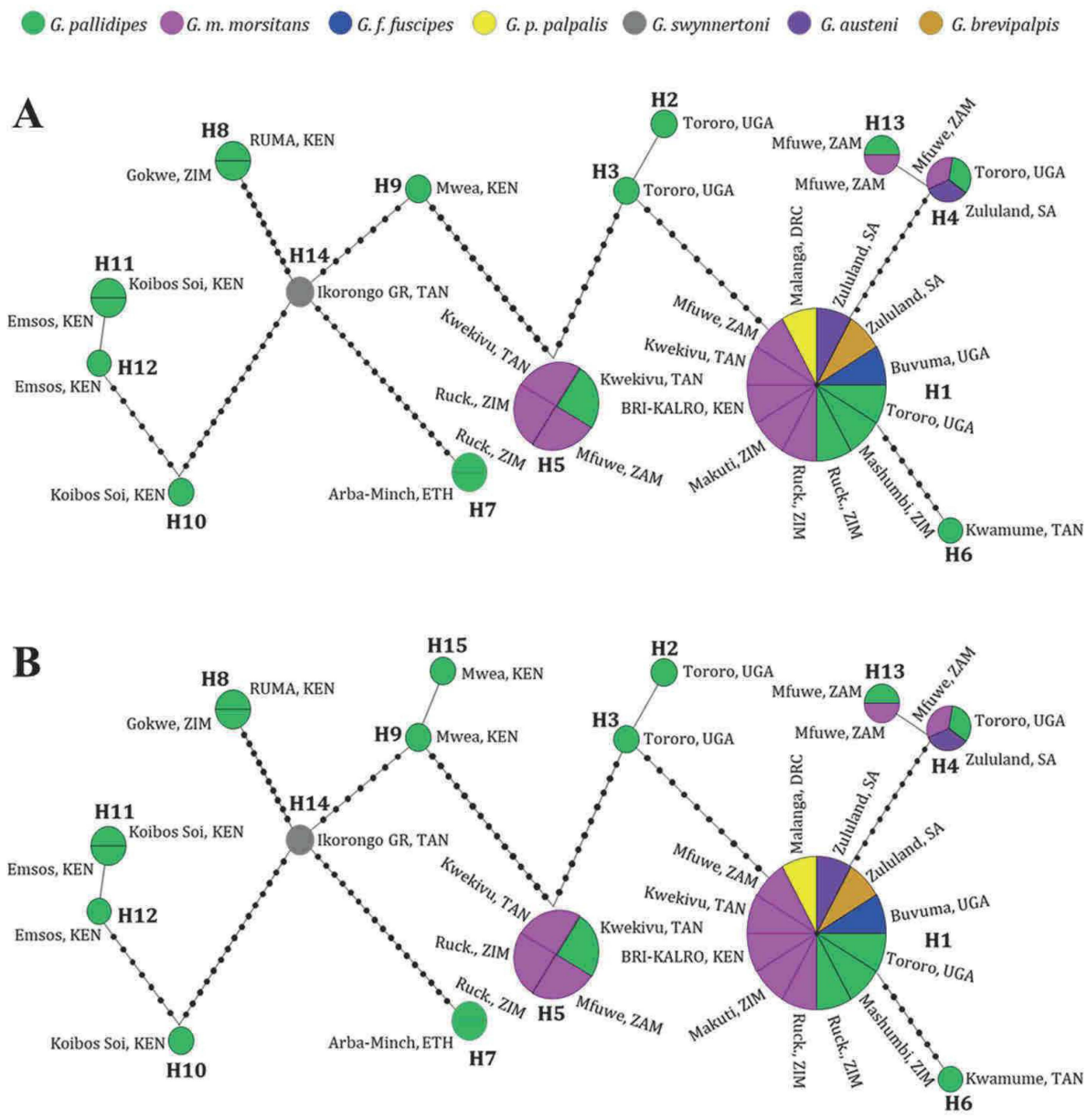


Figure 4: GpSGHV haplotype network in *Glossina* species: The haplotype network generated based on the ML tree generated based on GpSGHV VNTR-2 sequence. The black dots on the lines represent mutations events between the haplotypes. The different colours represent the *Glossina* species. Abbreviations; KEN; Kenya, TAN; Tanzania, ETH; Ethiopia, ZIM; Zimbabwe, ZAM; Zambia, UGA; Uganda, SA; South Africa, DRC; Democratic republic of Congo, Ruck; Ruckomechi, BRI-KALRO; Biotechnology Research institute-Kenya agricultural and livestock research organization, GR; Game reserve.

Discussion

This chapter provides the first evidence that all seven *Glossina* species examined harboured GpSGHV. The ability of GpSGHV to infect multiple *Glossina* species is important because the virus could hamper future SIT efforts as part of AW-IPM programmes against specific *Glossina* species in various sub-Saharan African countries. This finding is especially relevant for tsetse mass production facilities where multiple tsetse species are reared and often receive their blood meals using the same membranes in successive feeding cycles (Abd-Alla *et al.,* 2013). This feeding regime increases risks of virus transmission within and between the species. This cross-species virus transmission could result in viral amplifications and SGH outbreaks in laboratory colonies or mass rearing facilities (due to horizontal transmission of infectious virus particles via saliva during the *in vitro* membrane feeding (Kariithi *et al.,* 2013a)). Additionally, the cross-species virus transmission, could result in the generation within the original and new hosts of virus variants capable of efficient spreading amongst multiple host species (Parrish *et al.,* 2008).

The finding that GpSGHV prevalence was highest in *G. pallidipes* irrespective of the geographical locations (e.g. 100% in IPLC colony and 88% in Tanzania field samples) implies that the virus is present at high frequency in this species. This finding supports the notion that this virus most probably has a recent relationship with *G. pallidipes* compared to other tsetse species analysed in this study, perhaps due to virus-host interactions that influence genetic drift and selection as reported for arbovirus infection in different mosquito species (Grubaugh *et al.,* 2016). Next to *G. pallidipes*, the virus prevalence was high amongst populations of *G. m. morsitans* (e.g. 58% in Tanzania and 20% in Zimbabwe), and *G. f. fuscipes* (25-40% in Uganda). It should be noted that SGH symptoms were first reported in the 1930's amongst *G. pallidipes* populations in the Umfolozi Game Reserve, KwaZulu Natal, South Africa (Whitnall, 1934), and later in the early 1970's the causative virus was observed in *G. morsitans* and *G. f. fuscipes* in Tanzania and Uganda, respectively (Jenni and Steiger, 1974a; 1974b). Later, these virus particles were associated with SGH in *G. pallidipes* (Jaenson, 1978). Decades later (in 1993), the virus was reported in *G. brevipalpis* populations in Kenya (Shaw and Moloo, 1993). Overall, based on the chronological history since the initial discovery of SGH

in *Glossina* (Jaenson, 1978), unlike others tsetse species, *G. pallidipes* is evidently the most common *Glossina* species to which GpSGHV has not yet evolutionarily adapted to (Kariithi *et al.,* 2013a). It is yet to be determined why the virus is more pathogenic to *G. pallidipes* as compared with other *Glossina* species. It is obvious that a well-established evolutionary relationship between the virus and the host will result in a stable status, where the virus can be present in the host without affecting the host's general fitness or causing disease symptoms. What is known is that pathogens, including viruses, can specifically modulate their host-environment infections to favour their transmission (Mauck *et al.,* 2012).

The largest number of nonsynonymous mutations (in all the three genes) was found within the GpSGHV haplotypes infecting *G. pallidipes* populations from Ethiopia, Zimbabwe (H7), and Kenya (H10, H11 and H12). The nonsynonymous mutations provide a preview of the evolutionary path that can shape the genetic structure of viral haplotypes (Duffy *et al.,* 2008). Of the two GpSGHV strains whose genomes have been fully sequenced, the Ethiopian strain (GpSGHV-Eth) had a higher (> 85%) SGH prevalence in *G. pallidipes* originating from Ethiopia, than the Ugandan strain (GpSGHV-Uga) infecting *G. pallidipes* from Uganda (10% SGH prevalence), despite the two colonies being maintained in the same insectary conditions in IPCL (Abd-Alla *et al.,* 2016; 2013; 2010b). Whereas higher SGH incidences may not necessarily reflect higher pathogenesis, there are indications that this may be the case. It could as well be that the *G. pallidipes* flies from Ethiopia are less tolerant to the virus infection compared to the long-domesticated *G. pallidipes* colony in IPCL (Abd-Alla *et al.,* 2013). In this study, VNTR-2 revealed the highest rate of deletions and insertions of repeat units. VNTRs are amongst the most discriminating of the genotyping methods and have been used in pedigree analysis of disease-causing pathogens due to their roles in rapid genome evolution and adaptations (Gemayel *et al.,* 2012). Although VNTRs usually generate neutral genetic variations (Schlötterer, 2000), some VNTRs can alter critical biological functions. For instance, if localized near or within gene promoter regions, VNTRs may affect transcription of downstream genes by affecting the number of transcription factor binding sites or inducing changes in spacing between critical promoter elements (van Belkum *et al.,* 1998). In addition, it has been reported that polymorphism in VNTR loci contributes

to genome evolution (Gemayel *et al.,* 2012). In this study, the sequence analysis using either the VNTR-2 or the concatenate sequences of the three genes and two VNTRs revealed the same haplotypes. This indicates that VNTR-2 can be a suitable tool/microsatellite to discriminate GpSGHV haplotypes. Of the three genes used in this study, SGHV010 and SGHV038, which code for putative desmoplakin-like protein and maltodextrin glycosyltransferase, respectively, are both virion tegument proteins (Abd-Alla *et al.,* 2016; Kariithi *et al.,* 2013c). For some DNA viruses (e.g. herpesviruses), tegument proteins have been described that are essential for virus replication (Mocaski, 2007). The third gene (SGHV009) is known to be homologues to viral regulatory proteins. Although the mutations in the three ORFs and the two VNTRs represent a small subset of the GpSGHV genome to make robust conclusions, hypothetically, these mutations might affect the pathogenesis of GpSGHV-Eth as compared to the GpSGHV-Uga. It would be interesting to sequence the genomes of the GpSGHV strains circulating amongst the Zimbabwean and Kenyan *G. pallidipes* populations. This will help determine their genetic differences, compared to already sequenced GpSGHV strains. The nonsynonymous mutations found in these haplotypes need further investigations to elucidate their impacts on the virus pathobiology.

The central genealogy hubs (based on the haplotype network) occupied by haplotypes H1 (infecting six of the seven analysed *Glossina* species), H5 (infecting three *Glossina* species), and H14 (infecting only *G. swynnertoni*) provided insight into the potential ancestral origins and evolution of GpSGHV in *Glossina*. In addition, haplotype H1 is presumed to be the best potential ancestor of the virus among the three haplotypes due to its wide host-range and geographical representation. Hence, it was difficult and beyond the scope of this chapter to trace the origin and evolution of the virus through the species host or the geographical locations. However, an accurate assessment of the origin and evolution of large dsDNA viruses needs to be based on whole genomes, including primary genomic sequence comparisons, genome organizations and gene content (Shackelton and Holmes, 2004). Recently, it was shown that the two GpSGHV strains (GpSGHV-Uga and GpSGHV-Eth), although similar in nucleotide sequence (98.1%), differ in their genomes in terms of the numbers of ORFs (with insertions and deletions of

entire ORFs), and SNPs within the genes (Abd-Alla *et al.,* 2016). The three genes and the two VNTRs were selected for these analyses because they exhibited the most significant differences between the two virus strains. The hypothesis that H1 and H5 are potentially of ancestral origin is supported by previous findings by Kariithi *et al.* (2013b), who found the same haplotypes to occupy similar positions in the genealogical network. However, H5 from the previous study was found in *G. pallidipes* from Kenya, while in the current study H5 infected both *G. m. morsitans* and *G. pallidipes* from Tanzania, and *G. m. morsitans* from Zimbabwe and Zambia. In the current study, H14 (infecting *G. swynnertoni*) was identified as an additional possible ancestral GpSGHV origin. Further studies are necessary to characterize the haplotypes found in Kenya, which was found to be phylogenetically related to the GpSGHV-Eth, as well as to H14. Several Kenyan authors reported the occurrence of SGH symptoms in wild-caught *Glossina* species (Kariithi *et al.,* 2013a). Since the occurrence of SGH symptoms is an exception (Boucias *et al.,* 2013), especially in wild tsetse populations, one could conclude that the GpSGHV strains circulating amongst the various tsetse populations in Kenya could be as pathogenic as the GpSGHV-Eth. Although multiple GpSGHV haplotypes infecting the same tsetse species from the same geographical location were detected, the occurrence of multiple GpSGHV haplotypes in single individuals was not tested. However, this phenomenon in the *Glossina*-GpSGHV system cannot be ruled out since multiple virus haplotypes in the same individual has been reported in other systems such as *Drosophila* C virus system (Kapun *et al.,* 2010). Infection by multiple genotypes of nucleopolyhedrovirus has also been reported in *Spodoptera frugiperda* which contributed to the diversity of the virus (Clavijo *et al.,* 2010).

The findings presented in this chapter suggest that compared to other tsetse species, *G. pallidipes* might be the most recent host for GpSGHV. The large number of haplotypes observed in *G. pallidipes* suggests that the virus is still in the process of adapting to the host, which also partially explains why SGH symptoms were first observed in this species. This indicates that the original GpSGHV host species could be any other tsetse species that has yet to present overt SGH symptoms or is even infected at levels that are too low to be detected by conventional PCR. Hypothetically, the virus might on rare occasions be transmitted horizontally

between individuals and species when tsetse flies acquire a blood meal on the same animal in the field. In this scenario, infectious GpSGHV particles can pass from infected to uninfected flies via salivary secretions as up to 10^6 viral genome copies are secreted by an infected symptomatic fly in a 10-15 min blood meal feeding event during membrane feeding in the laboratory (Abd-Alla *et al.,* 2013; 2010b; 2011). These secreted virus particles can be infectious as evidenced by the reduction in virus copy numbers in flies fed with new blood at every feed, compared to flies fed under the normal feeding regime of feeding several sets of cages on the same tray of blood. (Abd-Alla *et al.,* 2013; 2010b; 2011). In the field, tsetse flies aggregate on specific parts of the host to feed (Spath, 2000; Van Den Abbeele *et al.,* 2010) and produce pharmacologically active saliva components that are deposited by the flies at the feeding site to interfere with host responses such as vasoconstriction and thrombocyte aggregation. This creates a blood pool at the bite site and maintain blood fluidity as well as reducing the blood diffusion rate (Caljon *et al.,* 2010). This may reduce the dilution by the host animal at the bite site of any infectious viral particles released via the saliva of infected flies and hence increase the chances of horizontal virus transfer to the flies feeding in the bite site proximity. This hypothesis has been discussed previously (Abd-Alla *et al.,* 2011), but needs experimental validation. Virus transmission through a shared food source has been demonstrated in the closest relative of GpSGHV, the MdSGHV that infects houseflies, whereby healthy flies became infected after they were fed on food contaminated by infected flies (Geden *et al.,* 2008; Lietze *et al.,* 2009). Similar modes of virus transmission have been reported in other insect viruses such as Israeli acute paralysis virus (IAPV) in bumblebees (Singh *et al.,* 2010b).

Conclusions

This chapter has demonstrated that the GpSGHV diversity is higher in *G. pallidipes* compared to other *Glossina* species. However, the high virus diversity in *G. pallidipes* from the current study differed with the results obtained in the previous study by Kariithi *et al.,* (2013a), which was based on conserved virus genes (*p74*, *pif-1*, *pif-2*, *pif*-3, and *dnapol*). In addition, the results appear to support the concept that GpSGHV has over evolutionary times reached a stable but dynamic equilibrium

with *Glossina* species other than *G. pallidipes*. In *G. pallidipes* the virus seems to be undergoing co-adaptation, thus accounting for the higher prevalence and diversity. This concept is also supported by the fact that it is only in *G. pallidipes*, that under certain laboratory settings, support symptomatic SGHV infections. In the natural tsetse populations, SGH symptoms are rarely observed. Some of the immune responses that may be compromised in *G. pallidipes* and further explored in **Chapters 5 and 6**. Taken together, this study presents VNTR-2 as a potential candidate to distinguish virus haplotypes since it was successfully sequenced in all the analyzed individuals. This is as opposed to the use of the concatenated sequences that had missing sequences of VNTR-1, SGHV009, SGHV010 and SGHV038 in some individuals due to unsuccessful attempts to amplify these candidate genes.

The finding that GpSGHV infects all *Glossina* species included in the current and previous studies underscores the importance of taking appropriate measures to ensure that field-derived biological material to establish new tsetse colonies for mass rearing is free of GpSGHV infections. A positive note on virus management in tsetse mass rearing is that it is highly likely that the key GpSGHV genes critical for the virus infections and transmission are conserved over haplotypes, implying that a common strategy can be used to mitigate virus infections in multiple tsetse species. This approach is supported by successful control of the GpSGHV using antiviral drugs, which target the viral *dnapol* gene (Abd-Alla *et al.,* 2012; 2013; 2014). With the potential of the evolving viral genotypes with enhanced infection and transmission dynamics in insect mass production facilities, the data presented herein are essential for future development of robust strategies against new GpSGHV strains. It is therefore recommended that different tsetse species should be reared in separate insectaries (or under appropriate conditions) to avoid horizontal transmission of GpSGHV from one species to another during membrane feeding.

Acknowledgements

The authors acknowledge Hasim Mohamed of IPCL for technical assistance in sample handling, preparations and data management. This research was supported by the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture,

Vienna, Austria, (Coordinate Research Project No. D42015) and Wageningen University and Research, The Netherlands.

Chapter 5

RNA interference-based antiviral immune response against the salivary gland hypertrophy virus in *Glossina pallidipes*

This chapter has been modified from:

Meki, I.K., Kariithi, H.M., Parker, A.G., Vreysen M.J.B., Ros, V.I.D., Vlak, J.M., van Oers, M.M. and Abd-Alla A.M.M. RNA interference-based antiviral immune response against the salivary gland hypertrophy virus in *Glossina pallidipes.* BMC Microbiology, (*in press*).

Abstract

Glossina pallidipes salivary gland hypertrophy virus (GpSGHV; *Hytrosaviridae*) is a non-occluded dsDNA virus that specifically infects the adult stages of *Glossina* species (Diptera: Glossinidae). GpSGHV infections are usually asymptomatic, but unknown factors can trigger a switch to an acute symptomatic infection state, which is characterized by the salivary gland hypertrophy (SGH) syndrome. SGH is associated with decreased fecundity that can ultimately lead to a colony collapse. It is uncertain how GpSGHV is maintained amongst *Glossina* spp., populations. However, the RNA interference (RNAi) machinery, a conserved antiviral defense in insects, is hypothesized to be amongst the host's mechanisms that could maintain the GpSGHV in an asymptomatic (persistent or latent) infection state. This chapter investigated the involvement of RNAi during GpSGHV infections by comparing the expression of three key RNAi machinery genes, *Dicer* (*DCR*), *Argonaute* (*AGO*) and *Drosha*, in artificially virus-injected individuals compared to PBS-injected controls. Comparisons were made of the expression levels of these genes between asymptomatically and symptomatically infected *G. pallidipes* flies. Further assessments were made on the impact of *AGO2* knockdown on virus infection by RT-qPCR quantification of four selected GpSGHV genes, i.e. *odv-e66, dnapol*, maltodextrin glycosyltransferase (a tegument gene) and SGHV091 (a capsid gene). These analyses show that in response to hemocoelic injections of GpSGHV into *G. pallidipes* flies, increased virus replication was accompanied by significant upregulation of the expression of *AGO1, AGO2* and *DCR2*, and a moderate increase in the expression of *Drosha* compared to the PBS-injected controls. Furthermore, compared to asymptomatically infected individuals, symptomatic flies had significantly lower transcript levels for *AGO1*, *AGO2* and *Drosha*, but a moderate increase in the expression of *DCR2*. Compared to the controls, knockdown of *AGO2* did not have a significant impact on virus infection in the flies as evidenced by unaltered transcript levels of the selected GpSGHV genes. The observed upregulation of the expression of the RNAi genes implicate involvement of this machinery in controlling GpSGHV infections and the establishment of symptomatic GpSGHV infections in *Glossina*. These findings provide a strategic foundation to understand GpSGHV infections and to control latent (asymptomatic) infections in *Glossina* spp. and thereby control SGHVs in insect production facilities.

Introduction

Tsetse flies (Diptera: Glossinidae) are naturally infected by the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV, family *Hytrosaviridae*), a large double-stranded DNA (dsDNA) virus pathogenic specifically to *Glossina* spp., (Abd-Alla *et al.,* 2009a; 2010a). Although the majority of colonized and wild tsetse fly species are asymptomatically infected by GpSGHV (low virus titers), some unknown factors can trigger symptomatic infections (high virus titers) in *G. pallidipes* (Abd-Alla *et al.,* 2009b). This, in turn, is associated with the occurrence of overt salivary gland hypertrophy (SGH) symptoms (Kariithi *et al.,* 2013a; 2013b). As discussed in Chapter 4, *G. pallidipes* not only has high virus prevalence, but also harbors more virus haplotypes compared to other tsetse species. In the mass rearing of *G. pallidipes*, SGH epizootics reduce fly survival and productivity, and have caused the collapse of three colonies in the past; two in the Insect Pest Control Laboratory (IPCL) in Seibersdorf, Austria (in 1987 and 2001), and one in the mass rearing facility in Kaliti, Ethiopia, in 2012 (Abd-Alla *et al.,* 2013). These GpSGHV-induced effects have significantly compromised the implementation of the sterile insect technique (SIT), a component of area-wide integrated pest management (AW-IPM) strategies designed for the eradication of *G. pallidipes* from the Southern Rift Valley of Ethiopia (Alemu *et al.,* 2007).

Although it is uncertain how the virus is maintained within wild and lab-bred tsetse populations, three hypothetical scenarios may account for the maintenance of asymptomatic GpSGHV infection state. The first is a persistent infection whereby the virus remains in specific host cells with low-level production of progeny virions, but without causing substantial cell damage (Boldogh *et al.,* 1996). The second scenario is a latent infection state, during which viral genomes and maybe some viral proteins are present in the infected host cells of certain organs, but without detectable production of infectious viral particles (Goic and Saleh, 2012). In the third scenario, the virus can exist in both persistent and latent infection states at the same time, but in different tissues (Goic and Saleh, 2012). Persistent infection in the salivary gland (SG) cells is accompanied by a low number of virions (10^2 viral genome copies/fly) released by asymptomatic flies via saliva during feeding (Abd-Alla *et al.,* 2010b). In addition, detection of viral DNA in other tissues such as the tracheal cells without

detectable viral gene transcripts (Kariithi *et al.,* 2017a) may reflect a latent infection state. In any case, the persistent or latent GpSGHV infection in *G. pallidipes* potentially represents a homeostatic equilibrium between the host's immune system and the viral escape strategies. Consequently, the viral infection is kept under control (asymptomatic state) but is not completely eliminated from the fly.

Amongst the possible host mechanisms that keep GpSGHV under control is the insect's RNA interference (RNAi) machinery, which regulates both host and viral gene expression by use of small RNAs that bind to their complementary messenger RNA (mRNA) targets (Kingsolver *et al.,* 2013; van Rij, 2008). This hypothesis is based on evidence from various studies indicating that the RNAi machinery is a conserved antiviral defense mechanism, which is not only active against RNA viruses but is also against several groups of large dsDNA viruses that infect insects, including the GpSGHV-related baculoviruses and nudiviruses, but also ascoviruses and iridoviruses (Burand and Hunter, 2013). RNAi is mediated through three pathways: short interfering RNA (siRNA), microRNA (miRNA) and Piwi-interacting RNA (piRNA) pathways (Bronkhorst *et al.,* 2012; Ding, 2010). The siRNAs are processed in the cytoplasm by the ribonuclease III enzyme Dicer-2 (DCR2) from exogenous double stranded (ds) RNAs (e.g. dsRNAs that arise as viral replication intermediates or from overlapping transcripts). These siRNAs destroy (viral or cellular) single-stranded RNAs (ssRNAs) in a sequence-specific manner (Lan *et al.,* 2016). The miRNAs, on the other hand, are processed by DCR1 from cellular or viral pre-miRNAs, originating from DNA components of nuclear replicating viruses that are processed in the nucleus by the RNase III enzyme Drosha (Skalsky and Cullen, 2010). The miRNAs are then exported to the cytoplasm where they post-transcriptionally regulate cellular or viral protein expression, thereby modulating developmental and physiological processes of the host, as well as virus infection (Asgari, 2013; Kincaid and Sullivan, 2012). The piRNAs are processed via diverse pathways independent of DCR proteins (Nandety *et al.,* 2015), and are involved in the regulation of cellular genes and the activity of transposons (Weick and Miska, 2014). The piRNAs may also have a role in antiviral strategies, as has been suggested for arboviruses (Miesen *et al.,* 2016). In the RNAi biogenesis pathways, the siRNAs, miRNAs and piRNAs are loaded into Argonaute proteins 2,

1 and 3 (AGO2, 1 and 3), respectively, that mediate the RNAi by either cleavage or degradation of target RNAs (AGO2), translation repression (AGO1), or epigenetic modifications (AGO3) (Carmell and Hannon, 2004). However, some of the above-mentioned enzymes may participate in two or more of these pathways. For instance, in *Drosophila melanogaster*, DCR1 is involved in the both siRNA and miRNA pathways, while DCR2 is only involved in the siRNA pathway. Additionally, *Drosophila* AGO1 and AGO2 enzymes may participate in both the siRNA and miRNA pathways (Tomoyasu *et al.,* 2008).

The siRNA-mediated RNAi pathway is a potent antiviral immune pathway in insects (Nandety *et al.,* 2015; Zambon *et al.,* 2006) and is implicated in controlling the replication of RNA and DNA viruses (Bronkhorst *et al.,* 2014; 2012; Jayachandran *et al.,* 2012; Kemp *et al.,* 2013; Sabin *et al.,* 2013). In addition, it has been shown for several viruses that the knockdown of RNAi pathway components leads to increased viral replication. For instance, loss-of-function mutations in *DCR2* enhanced the susceptibility of *Helicoverpa armigera* to infection by *H. armigera* single nucleopolyhedrovirus (HearNPV) (Jayachandran *et al.,* 2012). Similar observations were made in *Drosophila* during infection by Flock House virus (FHV), *Drosophila* C virus (DCV), and Sindbis virus (SINV) (Galiana-Arnoux *et al.,* 2006). In addition to controlling viral replication, the siRNA pathways have also been implicated in establishing persistent virus infections (Goic *et al.,* 2013). To establish persistent infections, viral fragments generated during viral DNA genome replication or that have been generated through reverse transcription from RNA viruses are integrated into the host genome. When these inserted fragments are transcribed they may be processed by DCR2 into virus-derived siRNAs (vsiRNAs); the vsiRNAs are then loaded into AGO2 to mediate specific cleavage of viral mRNAs, leading to persistent infection (Bronkhorst and van Rij, 2014; Gammon and Mello, 2015; Goic *et al.,* 2013). In the case of GpSGHV infections, the outcome of RNAi-based immune responses would hypothetically be restriction of viral replication and prevention of the development of overt SGH. If this is indeed the case, then the virus and the host would progress into a stable equilibrium of a persistent or latent infection state, which may account for the widespread chronic asymptomatic

GpSGHV infections in many tsetse species, particularly in colonized flies (Abd-Alla *et al.,* 2010b).

This chapter evaluated whether GpSGHV infection induces an RNAi response in *G. pallidipes* and whether this would downregulate the development of SGH symptoms and instead induce a covert infection state (persistent or latent). To accomplish this, comparative analyses was performed on the expression of *AGO*, *DCR* and *Drosha* between artificially (intra-hemocoelic) virus injected and uninfected (PBS injected) individuals, and between asymptomatic and symptomatic infected flies (with overt SGH symptoms). It should be noted that artificial injection of the virus does not result in overt SGH in the same (parental) generations rather the SGH symptoms are expressed in the progeny flies (Boucias *et al.,* 2013). These bioassays were complemented by testing the impact of downregulation of a key component in the siRNA pathway (AGO2), on GpSGHV infection in *G. pallidipes*. The data obtained in this chapter offer a rationale for similar studies on other *Hytrosaviridae* family members and may open novel strategies to manage SGHVs in insect production facilities.

Materials and methods

Tsetse fly and virus injections

The *G. pallidipes* flies were obtained from the colony maintained at the Joint FAO/IAEA IPCL, Seibersdorf, Austria. *G. pallidipes* was selected for this study because, unlike other tsetse species, it shows both asymptomatic and symptomatic virus infections (Abd-Alla *et al.,* 2009b). The experimental *G. pallidipes* flies were maintained in controlled insectaria with 70-80% relative humidity, 24 ± 1°C temperature and 12 h photo-phase. The flies were fed for 10-15 min, 3 times per week on defibrinated bovine blood using an *in-vitro* membrane feeding system (Feldmann, 1994). The virus inoculum was prepared from a single pair of hypertrophied salivary glands (with overt SGH symptoms) dissected from *G. pallidipes* male flies; viral titers in the gland homogenates (in PBS) were estimated by qPCR as described previously (Abd-Alla *et al.,* 2009b; Boucias *et al.,* 2013). The experimental flies were injected with 2 µl of the virus inoculum estimated to contain ~10^6 virus genome copies per µl.

Identification of core RNAi gene orthologs in Glossina

To determine whether the *G. pallidipes* genome contains the key RNAi pathway genes, *AGO*, *DCR* and *Drosha* sequences were retrieved from the VectorBase database (Giraldo-Calderón *et al.,* 2015) using as query sequences the annotated homologous gene sequences of *G. morsitans morsitans* (International Glossina Genome Initiative, 2014) (BLASTp; e-value $\leq 10^{-2}$). To determine the conservation of RNAi in *Glossina* species, sequences of these core genes for *G. fuscipes fuscipes*, *G. palpalis palpalis*, *G. austeni* and *G. brevipalpis* were similarly retrieved from VectorBase. The functional domain architecture of the retrieved *AGO*, *DCR* and *Drosha* sequences was analyzed using the ScanProsite tool (De Castro *et al.,* 2006). Homologous search of the retrieved sequences was performed to determine the relatedness of these sequences with those of *D. melanogaster* for which RNAi mechanisms and pathways have been demonstrated. Multiple alignments of protein-coding loci of the identified gene sequences were performed in BioEdit (Hall, 2011). Phylogenetic analysis was performed with MEGA6 using default settings for Maximum-Likelihood (ML) based on the General Time Reversible model with gamma distributed substitution rates with 1000 bootstrap replications (Tamura *et al.,* 2013).

Analysis of the expression of core RNAi genes in virus-injected G. pallidipes by RT-qPCR

To investigate the impact of GpSGHV on the core RNAi genes in *G. pallidipes*, two groups of teneral flies (newly-eclosed; non-fed; 50 females and 50 males per group) were injected with either the virus inoculum as described above, or phosphate buffered saline (PBS) as control. Following the injections, four females and four males were sampled within 1 h post injection and at 7, 14 and 21 days post injection (dpi). Total RNA was extracted from individual whole fly bodies using Trizol reagent (Invitrogen, Paisley UK) according to the manufacturer's instructions. Contaminating DNA was removed from the extracted RNA by treating the samples with DNase 1 (Invitrogen, Paisley, UK), after which the concentration of the RNA was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). Complementary DNAs (cDNAs) were synthesised using the SuperScript® III Reverse Transcriptase kit (Invitrogen, Paisley UK)

following the manufacturer's instructions. The iQ SYBR green supermix (Bio-Rad laboratories, Hercules, CA) was used for RT-qPCR analysis. The viral infection was assessed by quantifying the expression of GpSGHV *odv-e66*, a conserved, late viral gene (highly expressed upon viral genome replication), followed by expression analysis of *AGO*, *DCR* and *Drosha* transcripts, using the PCR cycling conditions: 95°C for 3 min, followed by forty cycles of 95°C for 10 s, 60°C for 1 min, then 95°C for 1 min and 55°C for 1 min, using the primers shown in **Table 1**. The tsetse housekeeping gene *β-tubulin* was used to normalize gene expression.

Differential expression of RNAi genes in asymptomatic and symptomatic flies

To determine the differential expression of *AGO*, *DCR* and *Drosha* between asymptomatic and symptomatic infected flies, 10-day old F_1 progeny flies produced by virus injected mothers were screened under a stereo microscope for the occurence of diagnostic SGH symptoms. Total RNA was extracted from whole bodies of eight asymptomatic and eight symptomatic infected flies (four females and four males). The viral infection level was estimated by quantification of the GpSGHV *odv-e66* transcripts, followed by expression analysis of the three genes as described above. These expression analyses were replicated three times (biological replicates).

Design of dsRNA constructs and prediction of off-targets

The optimal regions on the *AGO2* mRNA for the synthesis of dsRNA constructs were determined by siRNA design software (default setting) (Naito and Ui-Tei, 2012), which uses three predictive steps; (i) selection of functional siRNA sequence, (ii) selection of siRNA sequence with reduced off-target effects and (iii) elimination of near-perfectly matched off-target genes. The identification of off-targets was performed by BLAST (BLASTn; e-value $\leq 10^{-2}$) search of VectorBase. Following identification of the siRNAs, primers for dsRNAs synthesis were designed to flank the most effective siRNAs based on the above-described steps and a T7 promoter sequence added on each primer (See **Figure 1A** and **B** for *AGO2*, and **C** and **D** for the *tsetse EP* gene).

Table1: Sequences of the primers used in synthesis of dsRNAs and for expression analysis by RT-qPCR

Target gene	Primer name	Primer sequence (nt) – Primers are listed 5'- to -3'	Ref
		A. Primers for dsRNAs synthesis	
Argonaute 2	AGO-2T7-F	TAATACGACTCACTATAGGGGTCTTAGCATCCAACAACCA	This study
	AGO-2T7-R	TAATACGACTCACTATAGGGTGTCTATGCCGCACTCTTTC	
Tsetse EP	TseEPT7-F	TAATACGACTCACTATAGGGCTACGATAAATATGTCCCTC TAAT	Modified from (Walshe *et al.,* 2009)
	TseEPT7-F	TAATACGACTCACTATAGGGATCGGGCAAACCCTCAAC	
		B. Primers for q-RT-PCR	
Argonaute 1	AGO-1qPCR-F	CAACTGCTCGTTCGGCTCCA	This study
	AGO-1qPCR-R	GGCAAAACTCGTCCTCTTACTTCCA	
Argonaute 2	AGO-2qPCR-F	CGTTGGATGATGGCACAAAGATG	
	AGO-2qPCR-R	GCTGCCTGATGTGATGCAATTC	
Argonaute 3	AGO-3qPCR-F	GCACAACTAGCAGAGATGACAGATAC	
	AGO-3qPCR-R	TGCAGGGCAATCTTTTGGACAAT	
Dicer 2	DCR-2qPCR-F	GTAGAGCGAAGATACACGGCTAAA	
	DCR-2qPCR-R	CACCATAAATTGCGGCCTAATGAC	
Drosha	DroshaqPCR-F	TCAAAACCAAGGACAGAGCGGA	
	DroshaqPCR-R	GCAAACGGGGAAAAAGGCAAAC	
Tsetse EP	TseEPqPCR-F	ACCGTTCGTTCGCTTTACTAC	Modified from (Walshe *et al.,* 2009)
	TseEPqPCR-R	ACCAGCAGCCGTTTGACTTTC	
GpSGHV (*odv-e66*)	GpSGHVqPCR-F	CAAATGATCCGTCGTGGTAGAA	(Abd-Alla *et al.,* 2009b)
	GpSGHVqPCR-F	AAGCCGATTATGTCATGGAAGG	
GpSGHV (Maltodextrin glycosyltransferase)	GpSGHV32F	ACGCTGAACTAAATTATCGTCATCTACACG	This study
	GpSGHV31R	CACAGAATCGTCATCATCATCATCTACAGA	
GpSGHV (capsid protein)	GpSGHV92F	TATATTGTAATCCACGACCGGAAACTGAAC	
	GpSGHV91R	TCGGTAGGCGTGAATGAACGTTTT	
β-Tubulin (tsetse)	Tse-TubqPCR-F	GATGGTCAAGTGCGATCCT	(Caljon *et al.,* 2009)
	Tse-TubqPCR-R	TGAGAACTCGCCTTCTTCC	

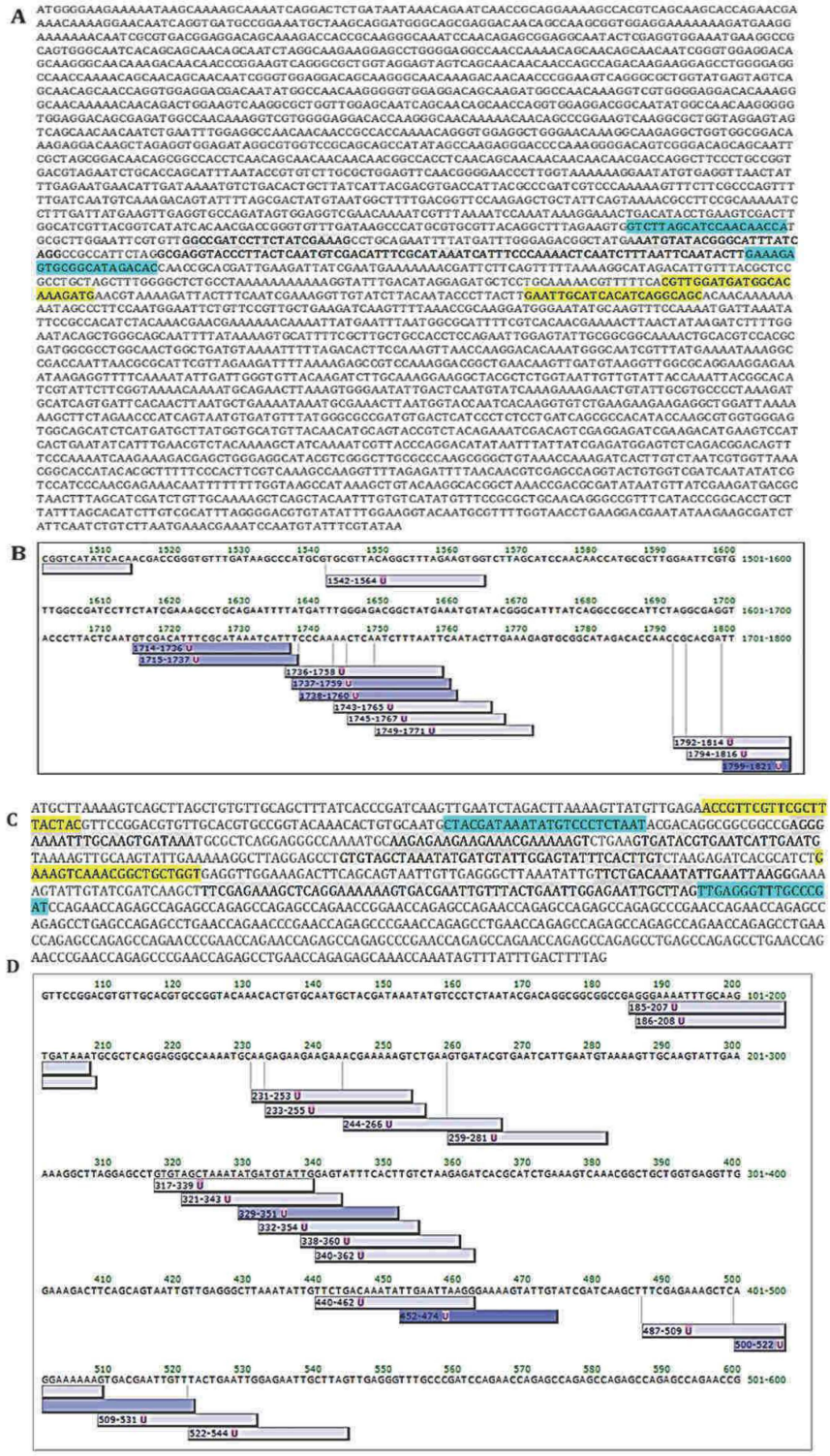


Figure 1: Output of *AGO2* and *tsetse EP* dsRNA design and off-target prediction. (A) *AGO2* mRNA sequence, (B) graphical view of effective *AGO2* siRNAs candidates, (C) *Tsetse EP* mRNA sequence, (D) graphical view of effective *tsetse EP* siRNAs candidates. The primers flanking the knockdown sequence are highlighted in cyan, grey highlights are the predicted siRNAs and the yellow highlights are the qPCR primers.

Synthesis of dsRNAs

To generate dsRNA to knockdown *AGO2*, total DNA was isolated from *G. pallidipes* using the Qiagen DNeasy Blood and Tissue kit (QIAGEN Inc, Valencia, CA). The extracted DNA was subsequently used to produce T7 promotor-tailed PCR amplicons of *AGO2* using primers designed to contain 5'-T7 promotor sequences (See **Table1**). These primers allowed dsRNAs transcription using the Hiscribe T7 Quick high yield RNA synthesis kit (New England Biolabs, UK) according to the manufacturer's instructions. Template DNA was removed from the transcription reaction by DNase treatment, as described in the transcription kit. The synthesized dsRNAs were purified using MEGAclear columns (Ambion, ThermoFisher Scientific, USA) and eluted in 50 µl nuclease free water. The *tsetse EP* gene, an immune response gene with extensive glutamic acid-proline dipeptide repeats, that has been successfully knocked down in tsetse, (Haines *et al.,* 2010; Walshe *et al.,* 2009) was used to assess the efficiency of the knockdown treatment (i.e. by measuring the expression of the *tsetse EP* gene).

Injections of flies with GpSGHV and dsRNAs

To investigate the impact of *AGO2* knockdown on GpSGHV infection, teneral *G. pallidipes* flies were divided into three groups each consisting of 40 females and 40 males and offered one blood meal. After 48 h, two of three groups were injected with 4 µl of either *AGO-2* or *tsetse EP* dsRNAs (2.5 µg/µl dsRNA in RNase free water) (See the procedure in **Figure 2**). The selection of this dsRNA dose was optimized for effective knockdown based on previous bioassays on dsRNA-mediated gene knockdown in tsetse flies (Haines *et al.,* 2010; Walshe *et al.,* 2009). The third group of flies (an additional negative control) was injected with RNase-free water. For the injections, flies were anaesthetized by chilling (~ 5 min) on ice, and subsequently injected in the dorsolateral surface of the thorax. Five days after the dsRNAs/RNase-free water injections, half of the injected flies (20 females and 20 males) from each group were injected with 2 µl of the virus suspension as described above, while the other half were injected with PBS. This time point (i.e. 5 days post dsRNA injection) was selected because successful gene knockdown in tsetse has been shown to occur after ~3 dpi (Haines *et al.,* 2010), implying that the 5 dpi in this case ensured that the virus was injected after successful knockdown. To

monitor the impact of *AGO2* knockdown on GpSGHV infection, 3 females and 3 males were sampled from each of the above described treatment groups at 1-hour post injection, and at 7, 14 and 21 dpi (**Figure 2**). The samples were stored at -20°C until further analysis as described below.

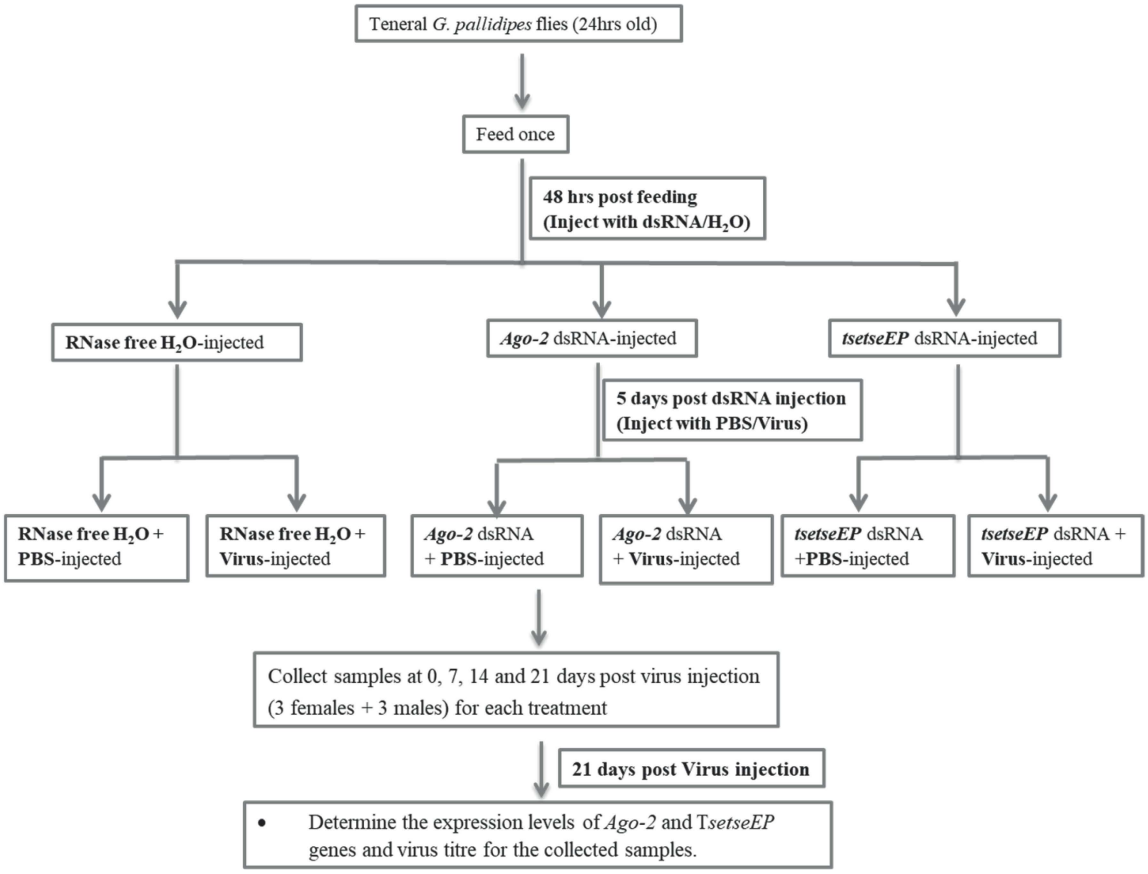


Figure 2: Step by step procedure to determine the effect of Knock-down of *AGO2* on virus infection in *G. pallidipes*. Teneral *G. pallidipes* flies were collected and offered one blood meal. 48 h later the flies were divided onto three groups and injected with 4 µl *AGO2* dsRNAs, *tsetse EP* dsRNAs or RNase-free water. Five days later, each of the 3 groups was divided into two and each injected with either 2 µl of the virus suspension or with PBS. Three females and 3 males were sampled from each of the described treatments at 1 h post injection, and at 7, 14 and 21 days post injection to determine the effect of *AGO2* knockdown on virus infection.

Analysis of the impacts of AGO2 knockdown on GpSGHV replication

To assess the effect of *AGO2* knockdown of GpSGHV replication, total RNA was extracted and cDNA synthesized as described above from the frozen fly samples collected from different time points post dsRNA and virus injection. The efficiency

of gene knockdowns was assessed by quantifying (by RT-qPCR) *AGO2* and *tsetse EP* gene transcripts using the qPCR primer sets listed in **Table 1**. The impact of *AGO2* knockdown on GpSGHV infection was assessed by RT-qPCR quantification of mRNA transcripts of the selected conserved GpSGHV genes; the *per os* infectivity *odv-e66* (SGHV005) gene, *dnapol* (SGHV079) gene involved in DNA replication, a tegument gene (SGHV038) and capsid gene (SGHV091) (Abd-Alla *et al.,* 2008; 2016). Note that a clear correlation between the GpSGHV*odv-e66* gene transcripts and the total virus copy numbers has been previously reported (Abd-Alla *et al.,* 2009b), which may demonstrate the impact of *AGO2* knockdown on virus replication.

Statistical analysis

All quantitative RT-qPCR results were representative of at least three independent biological experiments, each with three technical replicates. Statistical differences in the expression of the above described host and viral genes between the different treatments and the controls were performed with RStudio v1.0.143 (RStudio, 2015) (R v3.4.0 (R-Core, 2015)) using the packages lattice v0.20-35 (Sarkar, 2008) and MASS v7.3.47 (Venables and Ripley, 2002) The obtained data were visualized using the ggplot2 v2.2.1 package (Wickham, 2009) available within the RStudio platform. Data was checked for normality and transformed where necessary using the Box-Cox routine. The data was log transformed where the confidence interval of lambda includes 0 and transformed with $(x^{\lambda}-1)/\lambda$ in other cases. T-tests were used for the comparison of RT-qPCR data.

Results

The Argonaute family in Glossina species

The analyses of the genomes of *G. pallidipes*, *G. m. morsitans*, *G. f. fuscipes*, *G. p. palpalis*, *G. austeni* and *G. brevipalpis* resulted in the identification of AGO 1, 2 and 3 in all these species (**Table 2**), the key components of the RNA induced silencing complex (RISC); AGOs activate and cleave target mRNA within the RISC complex (Carmell *et al.,* 2002). This result underscores the conservation of RNAi machinery in *Glossina* species.

Table 2: VectorBase accession numbers of *Argonaute*, *Dicer* and *Drosha* genes of *Glossina* species and *D. melanogaster*.

Spp.	Argonaute 1	Argonaute 2	Argonaute 3	Dicer 1	Dicer 2	Drosha
Gp	GPA I022202	GPA I002659	GPA I022224	-	GPA I041589	GPA I009042
Gmm	GMO Y010338	GMO Y004940	GMO Y010351	GMO Y008446	GMO Y001890	GMO Y008669
Gff	GFU I031750	GFU I006141	GFU I039869	GFU I018989	GFU I024311	GFU I012078
Gpp	GPP I043499	GPP I035929	GPP I041119	GPP I007107	-	GPP I000118
Gau	GAU T002476	GAU T035389	GAU T027143	GAU T008865	-	GAU T013637
Gbr	GBR I043708	GBR I017817	GBR I017128	-	GBR I010244	GBR I016708
Dmel	NM_ 166020.2	NM_ 140518.3	NM_00104 3162.3	NM_ 079729.3	NM_ 079054.5	NM_ 058088.4

Spp. (species), Gp (*G. pallidipes*), Gmm (*G. m. morsitans*), Gff (*G. f. fuscipes*), Gbr (*G. brevipalpis*), Gpp (*G. p. palpalis*), Gaus (*G. austeni*) and Dmel (*D. melanogaster*).

Phylogenetically, the three AGO proteins segregated into distinct clusters with their orthologs in *D. melanogaster*, which corresponded to the siRNA, miRNA and piRNA pathways of the RNAi machinery (**Figure 3A**). The phylogenetic clustering was supported by robust bootstrap values. Additionally, similar to the *D. melanogaster* AGO family proteins, their orthologs in *Glossina* species contained the critical functional domains, i.e. the PAZ domain (for dsRNA binding) and the PIWI domain (executioner of the RNase activity) (**Figure 3B**). These results strongly suggest that the three RNAi machinery pathways are functional in *Glossina* spp.

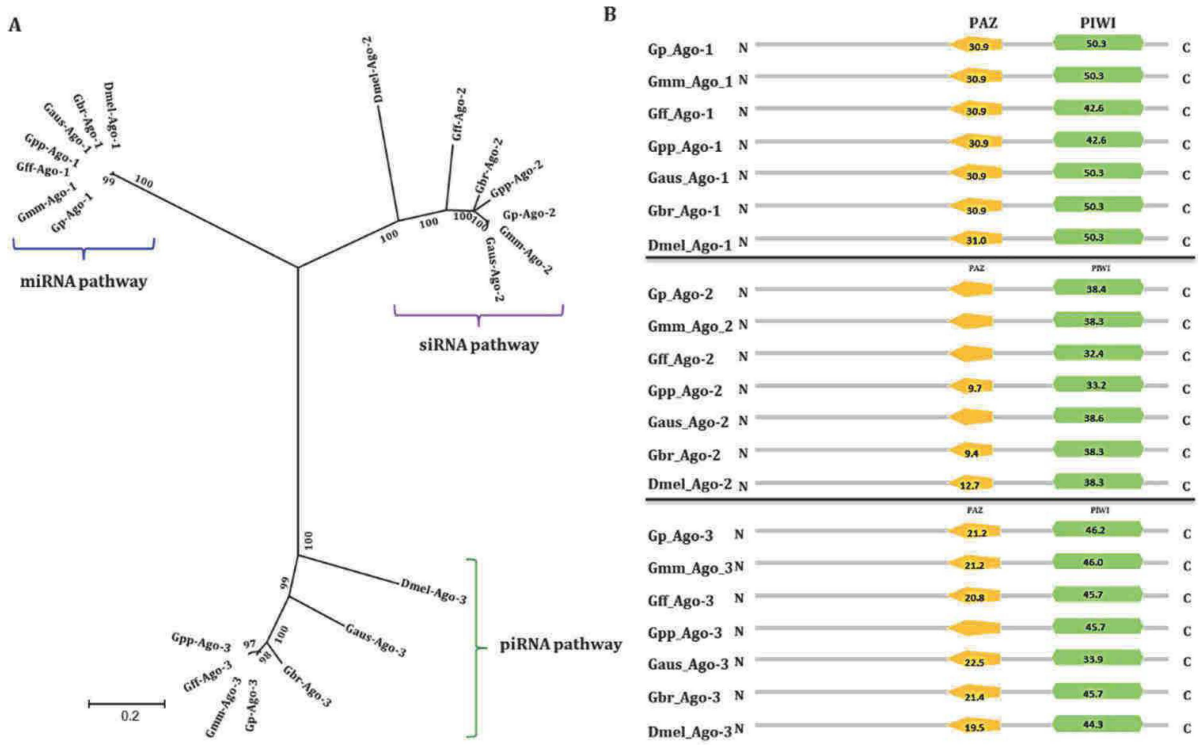


Figure 3: Phylogeny and domain analysis of Argonaute proteins. A) Maximum-likelihood based phylogenetic analysis (1000 bootstrap replicates) of Argonaute amino acid sequences of six tsetse species based on full length alignment with D. melanogaster as an outgroup. B) Domain architecture of Argonaute proteins. The numbers on the domains are the scores produced by the ScanProsite search compared to the PROSITE protein domain database. All the tsetse AGO1, AGO2, and AGO3 proteins show similarity in the domain architecture to their orthologs in D. melanogaster (Dmel-AGO1, Dmel-AGO2 and Dmel-AGO3), respectively. Abbreviations; AGO (Argonaute), Gp (*G. pallidipes*), Gmm (*G. m. morsitans*), Gff (*G. f. fuscipes*), Gbr (*G. brevipalpis),* Gpp (*G. p. palpalis*), Gaus (*G. austeni*) and Dmel (*D. melanogaster*).

The Dicer family in Glossina species

The bioinformatics analyses of the DCR protein family did not result in the identification of the homolog to the *Drosophila* DCR1 protein in *G. pallidipes* and *G. brevipalpis*, but the other *Glossina* species included in this study contained a DCR1 protein homolog (**Figure 4A**). However, homologs to the *Drosophila* DCR2 protein were present in both *G. pallidipes* and *G. brevipalpis*, suggesting that DCR2 might be involved in both siRNA and miRNA pathways, at least in these species. However, DCR2 was lacking in *G. p. palpalis* and *G. austeni* (**Figure 4A**), implying that in these two species DCR1 might be involved in both siRNA and miRNA

pathways. Homologs to both DCR1 and DCR2 were found only in *G. m. morsitans* and *G. f. fuscipes* suggesting that in these species they may be involved in two separate pathways (e.g. miRNA and siRNA) as reported in *Drosophila* (Lee *et al.*, 2004). *Drosha* was present in all six-tsetse species investigated (**Figure 4A**). The analysis also revealed the presence of all the functional motifs in the identified DCR (N-terminal helicases, DCR- dsRBF, PAZ, two C-terminal RNA III, and dsRBD) and *Drosha* (C-terminal RNA III and the dsRBD) protein homologs, which were organized as in their orthologs in *D. melanogaster* (**Figure 4B**).

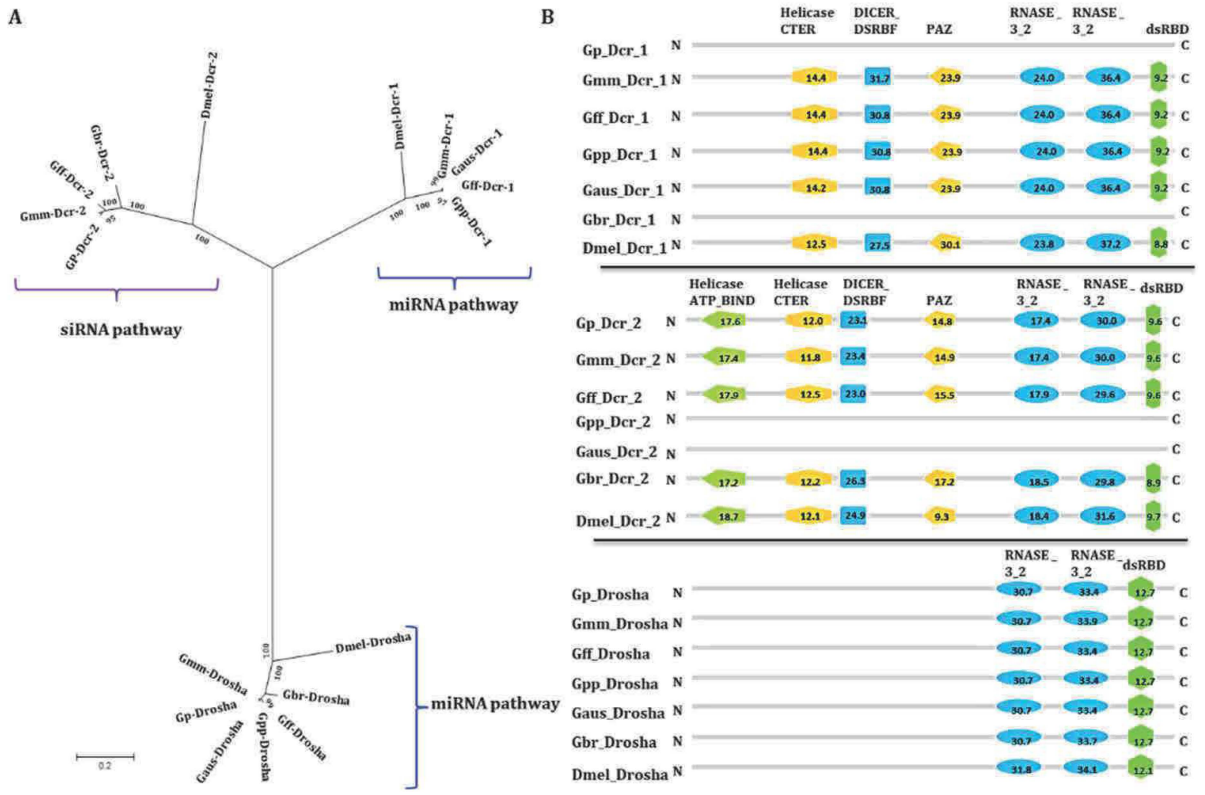


Figure 4: Phylogenetic and domain analysis of Dicer and Drosha proteins. A) Maximum-likelihood based phylogenetic analysis (1000 bootstrap replicates) of Dicer and Drosha amino acid sequences of six tsetse species based on full length alignment with *D. melanogaster* orthologs as outgroup. B) Domain architecture of Dicer and Drosha proteins. Some of the tsetse species had either DCR1 or DCR2 proteins, but Drosha was found in all the species. The numbers on the domains are the scores produced by the ScanProsite search compared to the PROSITE protein domain database. All DCR1, DCR2 and Drosha proteins show similarity in the domain architectures to Dmel-DCR1, Dmel-DCR2 and Dmel-Drosha, respectively. Abbreviations; DCR (Dicer), Gp (*G. pallidipes*), Gmm (*G. m. morsitans*), Gff (*G. f. fuscipes*), Gbr (*G. brevipalpis*), Gpp (*G. p. palpalis*), Gaus (*G. austeni*) and Dmel (*D. melanogaster*).

Notably, no major differences (phylogenetic position and domain architecture) were observed between the DCR1 or DCR2 protein sequences of the tsetse species containing one or both DCR proteins (**Figure 4A and 4B**). The presence of the functional domains in DCR and Drosha homologs could imply the conservation and functionality of both siRNA and miRNA pathways in tsetse. As indicated above, the piRNA pathway is independent of DCR implying that function of this pathway may not be affected by the presence or function of this protein.

Expression levels of AGO, DCR and Drosha in virus-injected G. pallidipes

Having identified the core genes involved in the RNAi machinery pathways, assays were performed to determine whether GpSGHV infection induces an RNAi response. Using *G. pallidipes* as the model species, and due to the absence of *DCR1* in this *Glossina* species (See **Figure 4**), the expression levels of the *AGO* family (*AGO1*, *AGO2* and *AGO3*), *DCR2*, and *Drosha* were determined in flies injected with the virus suspensions compared to the PBS-injected control flies. Additionally, a correlation was made on the expression levels of these RNAi-related genes to the level of virus replication by quantifying the expression levels of GpSGHV *odv-e66*, a conserved and late viral gene. Compared to the control (PBS-injected) fly group, the expression of *odv-e66* increased significantly with time ($t = 8.657$; d.f. = 44; $P < 0.001$) (**Figure 5A**), implying active replication and late gene expression of the virus in the injected flies. This increased virus replication was accompanied by a significant increase in the expression level of *AGO1* ($t = 2.306$; d.f. = 44; $P = 0.026$) and *AGO2* ($t = 3.334$; d.f. = 44; $P = 0.00174$) but not *AGO3* ($t = 1.651$; d.f. = 44; $P = 0.106$), of which the *AGO2* (involved in siRNA pathway) was the most upregulated (compare panels B, C and D in **Figure 5**). Similar to the *AGO* genes, *DCR2* gene, which may be involved in both siRNA and miRNA pathways in *G. pallidipes,* was also found to be significantly upregulated ($t = 3.968$; d.f. = 44; $P < 0.001$) in response to the virus injection (**Figure 5E**). However, unlike *AGO* and *DCR*, the expression levels of *Drosha* showed no significant increase ($t = 0.601$; d.f. = 44; $P = 0.551$) in the virus-injected flies compared to the levels observed in the PBS-injected flies (**Figure 5F**). This suggests that *Drosha*, part of the miRNA pathway, is not involved in the immune response against a lytic infection by GpSGHV.

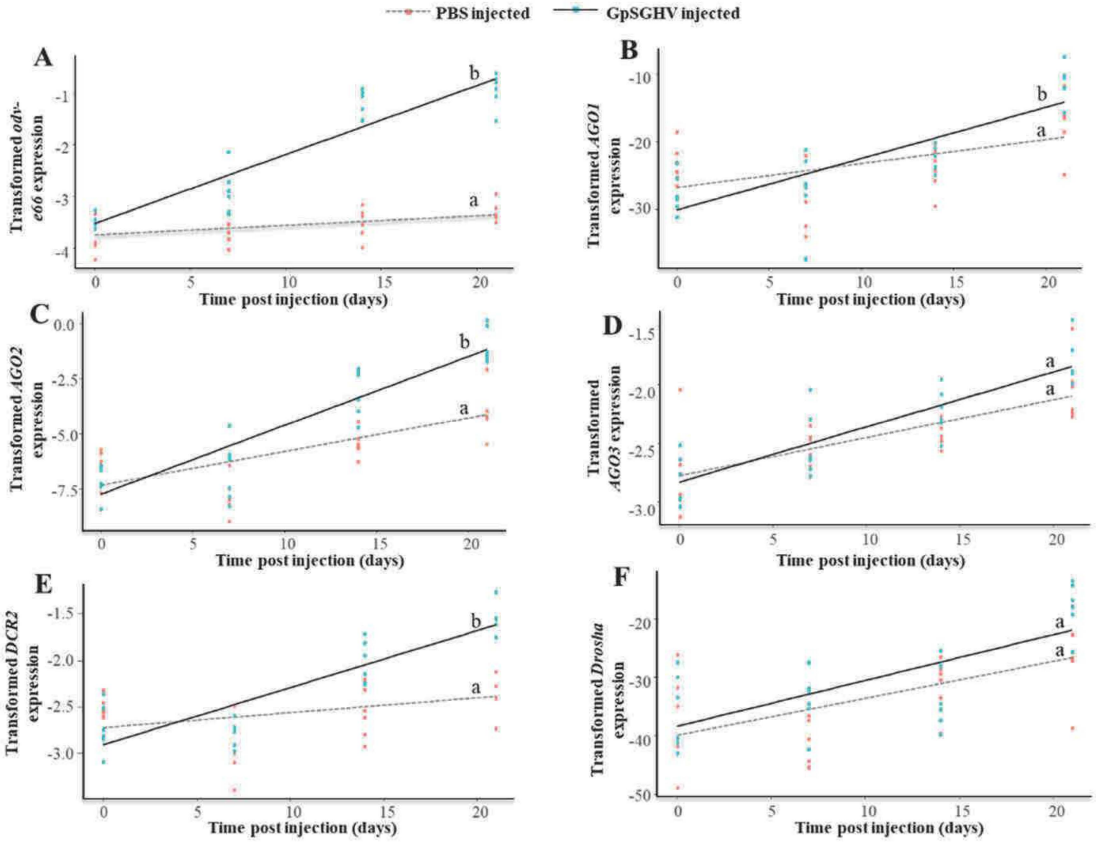


Figure 5: Relative expression of GpSGHV *odv-e66* and RNAi pathway genes post GpSGHV (black line) or PBS (grey dotted line) injection in *G. pallidipes* flies. A) GpSGHV *odv-e66*; B) *AGO1*; C) *AGO2*; D) *AGO3*; E) *DCR2*; and F) *Drosha*. Gene expression was quantified by RT-qPCR of the RNA extracted from whole fly bodies. Gene expression values were normalized to *β–tubulin* and transformed by the Box-Cox process. The expression levels of *AGO*1, *AGO2* and *Drosh*a were transformed using the lambda (λ) values (Expression$^{\lambda}$ - 1)/λ), while virus *odv-e66*, *AGO*3 and *DCR2* expressions were log transformed (log(Expression). The results from PBS and virus injection marked with the same lower-case letter do not differ at the 0.05 level.

Expression levels of AGO, DCR and Drosha in symptomatic G. pallidipes

Comparison was also made on the expression levels of the *DCR* and *AGO* family genes in the virus-injected *G. pallidipes* flies described above with the expression levels in symptomatic (flies with overt SGH symptoms and high virus titers) and asymptomatically infected individuals (flies with low virus titers) (t = 16.72; d.f. = 10; P < 0.001) (**Figure 6A**). There was significant downregulation of *AGO1* (t = -5.454; d.f. = 10; P < 0.001), *AGO2* (t = -3.899; d.f. = 10; P = 0.00363) and *Drosha*

(t = -3.549; d.f. = 10; P = 0.00623) in symptomatic *G. pallidipes* flies as compared to asymptomatically infected flies (**Figure 6B**).

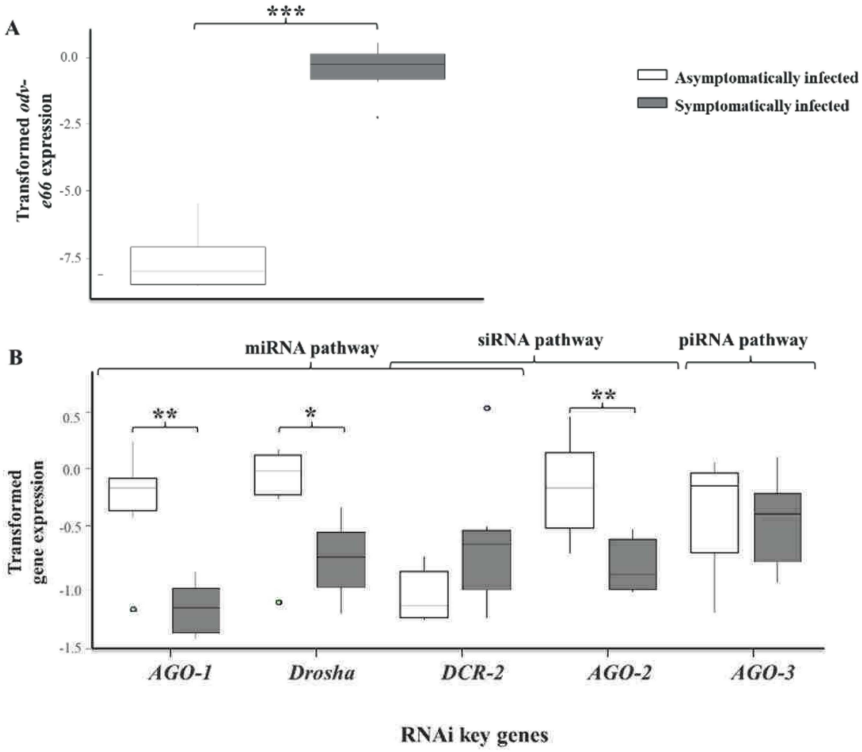


Figure 6: Comparative expression analysis of GpSGHV *odv-e66* and RNAi pathway genes in asymptomatically and symptomatically infected *G. pallidipes* flies. A) Virus *odv-e66* expression and B) RNAi genes expression. Gene expression was quantified by RT-qPCR of the RNA extracted from whole fly bodies. Gene expression values were normalized to *β–tubulin* and transformed by the Box-Cox process (log(Expression). The RNAi pathways in which the genes may be involved are also shown. Open boxes = asymptomatic flies; grey boxes = symptomatic flies. Asterisks indicate the statistical significance: ***P < 0.001, **P < 0.01, *P < 0.05.

There was no difference in expression of *DCR2* (t = 1.318; d.f. = 10; P = 0.2202), or *AGO3* (t = -0.858; d.f. = 10; P = 0.413) between the asymptomatically and symptomatically infected flies (**Figure 6B**).

Effect of AGO2 knockdown on virus infection in G. pallidipes

Assessment was made on the impact of *AGO2* knockdown on GpSGHV infection. The *AGO2* gene, which is involved in the siRNA pathway, was chosen for knockdown largely because its expression levels were significantly modulated in both the virus-injected flies (upregulated; see **Figure 5C**) and symptomatic infected

flies (downregulated; See **Figure 6B**). Compared to the flies injected with nuclease-free water, injection of flies with the dsRNAs did not cause any difference in mortality rate. The injection of dsRNAs specific for *AGO2* and *tsetse EP* in addition to PBS or virus injection resulted in a significant decrease in the expression levels of both *AGO2* (AGO2dsRNA/PBS cf water/PBS: t = -4.265; d.f. = 42; $P < 0.001$, AGO2dsRNA/virus cf water/virus: t = -3.543 d.f. = 42; $P < 0.001$) (**Figure 7A**) and *tsetse EP* (TsetseEPdsRNA/PBS cf water/PBS: t = -5.392; d.f. = 40; $P < 0.001$, TsetseEPdsRNA/virus cf water/virus: t = -6.798; d.f. = 40; P = 0.0034) (**Figure 7B**), compared to the water-injected control flies.

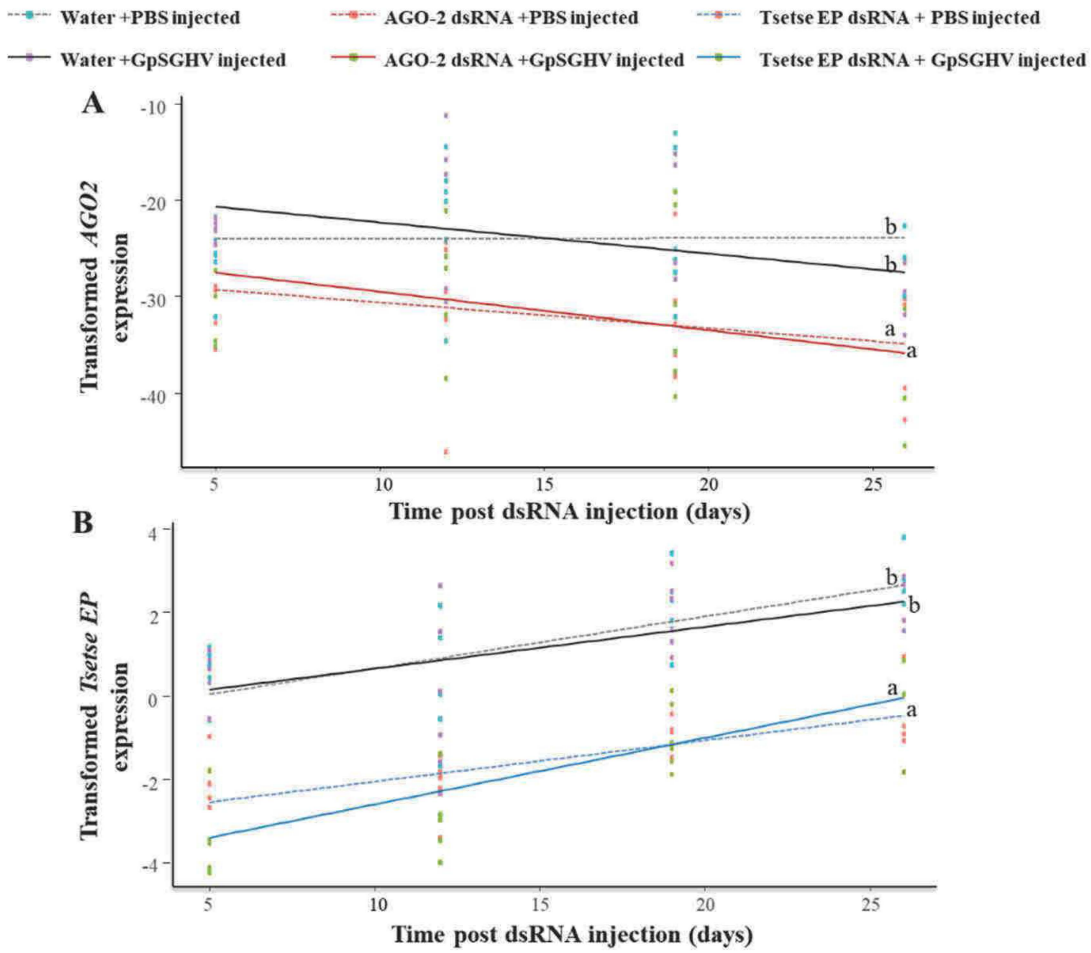


Figure 7: Validation of knockdown of *AGO2* and *tsetse EP* (control) genes in *G. pallidipes*. RT-qPCR expression analysis of: A) A*GO2* and B) *tsetse EP* post PBS/virus injection, following *AGO2* and *tsetse EP* dsRNAs injection, respectively, compared to water injected flies (negative controls). Gene expression values were normalized to *β–tubulin* and transformed by the Box-Cox process (Expressionλ - 1)/λ). Regression lines marked with the same lower-case letter do not differ at the 0.05 level.

Then the effect of *AGO2* knockdown on virus infection was assessed by quantifying expression levels of the selected viral genes, *odv-e66*, *DNApol,* SGHV038 and SGHV091. *AGO2* knockdown did not have a significant impact on the transcript levels of any of the selected viral genes; *odv-e66* (t = -1.861; d.f. = 119; P = 0.391), *DNApol* (t = -0.422; d.f. = 119; P = 0.674), SGHV038 (t = -0.179; d.f. = 119; P = 0.858) and SGHV091 (t = -0.877; d.f. = 119; P = 0.382) compared to the controls (**Figure 8**).

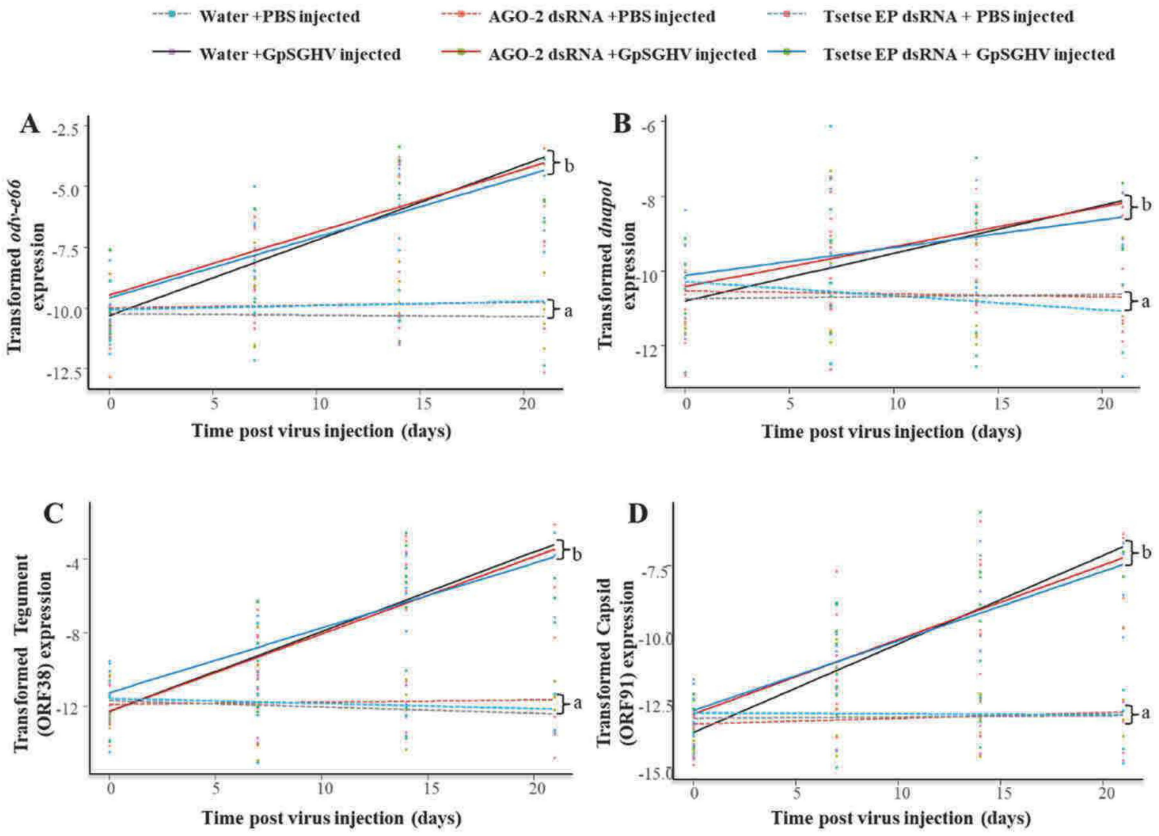


Figure 8: Effect of *AGO2* silencing on selected GpSGHV transcript levels in *G. pallidipes*, following *AGO2* knockdown. RT-qPCR expression analysis of (A) GpSGHV *odv-e66* gene, (B) GpSGHV *dnapol* gene, (C) GpSGHV tegument gene and (D) GpSGHV capsid gene post virus injection, following *AGO2* and *tsetse EP* (control) dsRNAs injection. Gene expression values were normalized to *β–tubulin* gene and transformed by the Box-Cox process (log(Expression)). Regression lines marked with the same lower-case letter do not differ at the 0.05 level.

As expected, knockdown of *tsetse EP*, which is not associated with the RNAi machinery, did not affect the expression levels of these selected viral genes.

Discussion

This chapter investigated the potential involvement of the RNAi machinery during GpSGHV infections in *G. pallidipes* by quantifying the expression of both the host (*AGO*, *DCR* and *Drosha*) and viral genes. The finding that AGO and DCR proteins in *Glossina* species contain the functional domains or motifs known to mediate the RNAi response strongly indicate that the RNAi machinery is functional in tsetse, presumably in a similar version as reported in other insects such as *Drosophila* (Wang *et al.,* 2006). It is known that the presence or absence of these functional domains in AGO or DCR proteins affects the efficiency of the RNAi response. For instance, Gu *et al.* (2012) discovered that although AGO2 lacking the PAZ domain interacts with duplex siRNAs, the truncated protein was unable to unwind the siRNAs or eject the passenger RNA strands. The passenger strand should be degraded or ejected from the siRNA duplex to allow the guide strand to be incorporated into the RISC complex and target the mRNA (Nandety *et al.,* 2015). Moreover, the results from the phylogenetic analyses confirmed that the predicted AGO proteins from the various tsetse species analyzed clustered with their homologs in *Drosophila*.

The presence of only *DCR2* in *G. pallidipes* and *G. brevipalpis*, and only *DCR1* in *G. p. palpalis* and *G. austeni* suggests that in these species only one of the respective proteins is involved in the siRNA and miRNA pathways. In contrast, genes for both DCR1 and DCR2 proteins were found in *G. m. morsitans* and *G. f. fuscipes*. Given that in *Drosophila* DCR1 and DCR2 are involved in the miRNA and siRNA pathways, respectively (Lee *et al.,* 2004), it is possible that these two proteins are involved in these pathways in *G. m. morsitans* and *G. f. fuscipes* as well. Alternatively, the genes might be involved in both pathways, as suggested for the other *Glossina* species as there were no differences in DCR1 or DCR2 protein sequences of species with one or both DCR proteins, which could be linked to the RNAi pathways. The DCR proteins can be involved in both siRNA and miRNA pathways (Carmell and Hannon, 2004) or be involved in the separate pathways (Lee *et al.,* 2004). Taken together, the identification of genes for AGO and DCR proteins may, in principle, be an indication of a robust RNAi silencing response in *Glossina* species (Meister and Tuschl, 2004; Parker and Barford, 2006).

Currently, the mechanism(s) enabling the covert (asymptomatic) GpSGHV infection in *G. pallidipes*, and the reactivation from viral persistence/latency to overt symptomatic infection that is associated with overt SGH symptoms (Boucias *et al.,* 2013) are poorly understood. The finding of significant upregulation of the expression of *AGO2* in GpSGHV-injected flies in this study, suggests that the virus infection induces the host's siRNA-mediated response, presumably to inhibit the virus infection. Notably *DCR2*, which might be involved in both siRNA and miRNA pathways in *G. pallidipes*, was similarly upregulated during virus infection. The upregulated expression of both *AGO2* and *DCR2* post virus injection in *G. pallidipes*, which are key components in the dsRNA-mediated gene silencing in several insects, suggests a functional RNAi-mediated innate immunity response in *Glossina* species. However, more work is required to elucidate the precise details of this pathway in *Glossina* as well as the involvement of the RNAi machinery in other members of the *Hytrosaviridae* family.

In contrast to the above-mentioned increased levels of *AGO2* and *DCR2* in virus-injected flies, the comparative analysis of the expression of the two genes between asymptomatic and symptomatic infections showed a significant downregulation of the expression of *AGO2*, but insignificant upregulation of *DCR2* in the flies exhibiting diagnostic SGH symptoms. The high expression of siRNA pathway genes in the virus-injected flies suggests a tight control of the virus via the RNAi response during asymptomatic GpSGHV infections. However, during symptomatic infections as evidenced by increased virus titers, the siRNA pathway may be compromised (as supported by the low expression of *AGO2*) enabling the virus to escape the RNAi-mediated innate immunity, thereby increasing virus titers and in turn causing the detectable SGH symptoms. A similar outcome has been documented in the case of the African malaria mosquito, *Anopheles gambiae*, where dsRNA-mediated silencing of *AGO2*, which functions in conjunction with *DCR2* in this mosquito, resulted in increased O'nyong-nyong virus (ONNV) viral loads (Keene *et al.,* 2004). It should be noted that, due to its involvement in both the siRNA and miRNA pathways, *DCR2* was not considered a suitable candidate for the knockdown assays in this study. Therefore, the potential involvement of dsRNA-mediated gene silencing (siRNA pathway) in *G. pallidipes* was assessed by knockdown of *AGO2*.

Several examples of the function of RNAi in insects have been clearly demonstrated, including for species from the orders Diptera, Dictyoptera, Isoptera, Hymenoptera and Orthoptera (Huvenne and Smagghe, 2010; Swevers and Smagghe, 2012). In the case of tsetse flies, Walshe *et al.,* (Walshe *et al.,* 2009) showed that micro-injection of dsRNA into 6-8-day-old *G. m. morsitans* flies could persistently silence expression of *tsetse EP,* a gene that is demonstrated to protect the fly against establishment of trypanosome infections in the midgut (Haines *et al.,* 2010). In the current study, the knockdown of *AGO2* in *G. pallidipes* did not alter the transcript levels of the selected GpSGHV genes implying that *AGO2* knockdown had no effect on the GpSGHV infection. These findings contrast with previous results, for instance in *Drosophila melanogaster* flies deficient in the DCR2 protein showed increased susceptibility to infection by members of three different RNA virus families; i.e. FHV (*Nodaviridae*), DCV (*Dicistroviridae*), and SINV (*Togaviridae*) (Galiana-Arnoux *et al.,* 2006). In general, many studies have led to the conclusion that flies that contain mutations in genes that encode components of the siRNA pathway (including *DCR2* and *AGO2*) or the Janus kinase/signal transducers and activators of transcription (Jak-STAT) pathway, are not only more sensitive to infection by several viruses, but also harbor higher viral titers than their wild-type counterparts. The Jak-STAT pathway is also a conserved insect innate immune antiviral response (Galiana-Arnoux *et al.,* 2006; Kingsolver *et al.,* 2013; van Rij *et al.,* 2006; Wang *et al.,* 2006; Zambon *et al.,* 2006). In another report, *D. melanogaster* defective for the *AGO2* were found to be hypersensitive to infections by DCV, an infection which also supported a 1,000-fold increased production of progeny virus (van Rij *et al.,* 2006).

The results presented in this chapter showed that reduction of RNAi efficiency in *G. pallidipes* did not cause a significant impact on the fly immunocompetence. Notably, in addition to the RNAi pathway investigated in this study, the Jak-STAT, immune deficiency (Imd) and Toll immune pathways are also involved in elimination of viruses (Costa *et al.,* 2009; Fullaondo and Lee, 2012; Sabin *et al.,* 2010). The presence of these innate immune pathways may imply that following the interruption of the RNAi pathway in *G. pallidipes* flies in this study, alternative antiviral pathways in these flies were able to control the virus infections.

Conclusions

Given the high diversity of pathogens, their elimination by host organisms is challenging and therefore many organisms, including insects, employ multiple mechanisms to deal with them. The RNAi machinery, particularly the siRNA pathway, plays a central role in insects by specifically recognizing and eliminating invading pathogens and other invasive elements such as transposons. The findings in this chapter have elucidated important milestones in the infection of GpSGHV in *Glossina* spp. Key RNAi genes were detected in all the analyzed *Glossina* species, which may indicate a functional antiviral role of RNAi machinery in tsetse flies. In *G. pallidipes*, the model *Glossina* species in this study, the siRNA pathway genes *AGO2* and *DCR2* were upregulated during virus infection, which confirmed the involvement of the RNAi response in the flies' defense against GpSGHV. It was also noted that in *G. pallidipes*, the siRNA pathway and perhaps the miRNA pathway (analyzed further in **Chapter 6**) was compromised during symptomatic infection as evidenced by the low expression of *AGO1, AGO2* and *Drosha*. However, although knockdown of *AGO2* in *G. pallidipes* did not have an impact on virus infection, it would be worthwhile to further investigate the long-term effects of the gene knockdowns on GpSGHV transmission and the expression of SGH symptoms in F_1 progeny produced by the parents with the knockdown. The F_1 generation is of interest since induction of SGH symptoms does not occur in the parental generation of flies that are intra-hemocoelically injected with the virus but are observed in the subsequent F_1 generation produced by injected mothers (Boucias *et al.,* 2013). It may therefore be important to determine whether the dsRNA-mediated gene silencing is a heritable trait, as has been demonstrated in other studies (Rechavi, 2014). The next question would be to determine which factors (exogenous or endogenous) trigger the transition from the asymptomatic to symptomatic state in colonized and wild flies.

Acknowledgements

The authors acknowledge Prof. Serap Aksoy of Yale University for insightful discussions during the conception of the study and Hasim Mohammed, Carmen Marin and Henry Adun for tsetse rearing and technical assistance. This study was funded by Joint FAO/IAEA Division of Nuclear Techniques in Food and

Agriculture, Vienna, Austria, (Coordinate Research Project No. D42015) and Wageningen University and Research, The Netherlands.

Chapter 6

Expression profile of *Glossina pallidipes* microRNAs during symptomatic and asymptomatic infection with hytrosavirus

This chapter has been modified from:

Meki, I.K., İnce, İ.A., Kariithi, H.M., Boucias, D.G., Orhan O., Parker, A.G., Vlak, J.M., van Oers, M.M. and Abd-Alla A.M.M. Expression profile of *Glossina pallidipes* microRNAs during symptomatic and asymptomatic infection with *Glossina pallidipes* salivary gland hypertrophy virus (hytrosavirus). Frontiers in Microbiology, (*DOI: 10.3389/fmicb.2018.02037*).

Abstract

The *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) infects tsetse flies predominantly asymptomatically and occasionally symptomatically. Symptomatic infections are characterised by overt salivary gland hypertrophy (SGH) in mass reared tsetse flies. This coincides with reproductive dysfunctions leading to decreased fecundity that can result in collapse of the colony, which has considerably hindered the control of tsetse vector via the sterile insect technique (SIT). In this chapter, small RNAs were sequenced and profiles in asymptomatically and symptomatically infected *G. pallidipes* flies determined. Thirty-eight host-encoded microRNAs (miRNAs) were present in both the asymptomatic and symptomatic fly profiles, while nine host miRNAs were expressed specifically in asymptomatic flies versus 10 in symptomatic flies. Of the shared 38 miRNAs, 15 were differentially expressed when comparing asymptomatic with symptomatic flies. The most up-regulated host miRNAs in symptomatic flies was predicted to target immune-related mRNAs of the host. Six GpSGHV-encoded miRNAs were identified, five of which were found only in symptomatic flies. These virus-encoded miRNAs may not only target host immune genes but may also participate in viral immune evasion. This evidence of differential host miRNA profile in *Glossina* in symptomatic flies advances our understanding of the GpSGHV-*Glossina* interactions and provides potential new avenues, for instance by utilization of particular miRNA inhibitors or mimics to better manage GpSGHV infections in tsetse mass-rearing facilities, a prerequisite for successful SIT implementation.

Introduction

The *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV; family *Hytrosaviridae*) is a large, rod-shaped dsDNA virus pathogenic to some species of tsetse flies (*Glossina* spp.) (Abd-Alla *et al.,* 2010a). Tsetse flies such as *Glossina pallidipes* infected with GpSGHV can exhibit either asymptomatic or symptomatic infection states; the former is the most prevalent in reared tsetse fly colonies and in nature (Abd-Alla *et al.,* 2010b; Kariithi *et al.,* 2013a). Asymptomatic (presumed latent) infection state has no apparent fitness cost to infected flies, while symptomatic infections are associated with reproductive dysfunctions that sometimes result in collapse of infected fly colonies (Abd-Alla *et al.,* 2010b; 2011). Asymptomatically infected female *G. pallidipes* can vertically transmit the virus to their offspring through milk gland secretions or transovarially (Abd-Alla *et al.,* 2010b). In tsetse mass rearing facilities, horizontal transmission of the virus mainly occurs during collective *in vitro* membrane feeding, whereby symptomatic flies release the virus via the saliva when taking a blood meal (Abd-Alla *et al.,* 2011). Symptomatic infections in *G. pallidipes* are characterised by detectable salivary gland hypertrophy (SGH). The symptomatic infections also result in testicular degeneration in males and ovarian abnormalities in females, which leads to decreased fecundity of the colony (Abd-Alla *et al.,* 2010b). Therefore, the occurrence of symptomatic GpSGHV in colonies of *G. pallidipes* makes colony maintenance challenging and drastically increases the risk of colony decline or even collapse (Abd-Alla *et al.,* 2013).

Maintenance of healthy tsetse fly colonies is crucial for the application of the sterile insect technique (SIT) to manage tsetse fly populations and African trypanosomosis, the disease these flies transmit (Feldmann *et al.,* 2005). SIT requires mass release of sterile males into the target wild insect population to mate with virgin wild females. These matings will produce no offspring in the target population, which will eventually decline as the population replacement rate is reduced (Knipling, 1959). Symptomatic virus outbreaks in mass-rearing facilities of tsetse species such as *G. pallidipes* are a serious impediment to the implementation of the SIT. This has stimulated research efforts to better understand virus-host interactions at the molecular level and to identify the parameters that determine whether GpSGHV infections become symptomatic or remain covert (Abd-Alla *et al.,* 2010b). As

discussed in Chapter 5 of this thesis, the RNA interference (RNAi) pathways, which are mediated by short interfering RNA (siRNA) and microRNA (miRNA), are known to modulate virus-host interactions in insects, thereby providing an antiviral defence (van Rij, 2008). P-element induced wimpy (PIWI) testis in *Drosophila*-interacting RNAs (piRNAs) are a separate group of non-coding small RNAs of 25-30 nucleotides (nt) that have been shown to repress transposable elements and regulate cellular genes (Luo and Lu, 2017). The piRNAs have recently been shown to play a role in antiviral strategies in insects against arboviruses (Miesen *et al.,* 2016).

The miRNAs are short (18–24 nt) non-coding RNAs that regulate host or pathogen gene expression post-transcriptionally by binding to complementary regions located mainly in the 3' untranslated regions (3'-UTRs) of targeted mRNAs (Hussain and Asgari, 2010). The miRNAs regulate virus infection and other biological processes in animals, plants and insects (Skalsky and Cullen, 2010). For a number of dipteran insects, it has been shown that the miRNA expression profile changes during virus infection and in this way the expression level of host genes with a role in immunity can be modulated (Lucas and Raikhel, 2013). For instance, in the yellow fever mosquito, *Aedes aegypti*, the host miRNA aae-miR-374 enhanced dengue virus (DENV) infection, while another host miRNA aae-miR-2940 reduced replication of the virus (Asgari, 2014; Zhang *et al.,* 2013a). These findings indicate that miRNAs can either positively or negatively regulate the host defence to pathogen infection. Viruses may also encode miRNAs that target host cellular mRNAs and, in that way manipulate host gene expression and ensure effective virus proliferation. Virus-encoded miRNAs may target host or virus genes in order to maintain a latent infection state (Cullen, 2009; He *et al.,* 2014). For instance, it has been reported that the DNA virus *Heliothis virescens* ascovirus (HvAV) encodes an miRNA, HvAV-miR-1, that targets its own DNA polymerase thereby inhibiting lytic virus infection and maintaining a persistent state of the virus (Hussain *et al.,* 2008).

The role of miRNAs in virus-host interactions has been demonstrated in many insects but there is limited information on how GpSGHV infection affects the miRNA profile in tsetse flies. The hypothesis is that GpSGHV alters both viral- and host-encoded miRNA profile in tsetse flies and that specific miRNAs may play a role in inducing or facilitating SGH in some cases and a latent infection state in other

cases. The research in this chapter was designed to investigate the role of host and virus-encoded miRNAs during GpSGHV asymptomatic and symptomatic infection in the tsetse fly *G. pallidipes*. Deep sequencing of small RNA (sRNA) molecules was used to identify host and GpSGHV-encoded miRNAs and to determine whether these were differentially expressed in asymptomatically and symptomatically infected flies, or not. Furthermore, the mRNA targets of the differentially expressed host miRNAs and the GpSGHV-encoded miRNAs were predicted to investigate their potential roles during GpSGHV symptomatic infection. Finally, using inhibitors and miRNA mimics, the functional significance of some miRNA was experimentally validated. This study presents important information on the interaction between GpSGHV and *G. pallidipes* miRNAs and provides potential avenues to further study the mechanisms of immune response during GpSGHV infections in tsetse flies.

Materials and methods

Tsetse flies, GpSGHV preparation and injection

The *G. pallidipes* flies were obtained from the colony maintained at the Joint FAO/IAEA Insect Pest Control Laboratory (IPCL), Seibersdorf, Austria. The flies were maintained in an environment-controlled insectary at 23±1°C, 75-80 % relative humidity, and a 12 h photo-phase. The flies were fed for 10-15 min, three times per week on defibrinated bovine blood using an *in-vitro* membrane feeding system (Feldmann, 1994). The virus inoculum used was the Ugandan GpSGHV isolate (GpSGHV-Uga) prepared from one pair of salivary glands dissected from *G. pallidipes* flies showing overt SGH and homogenised in phosphate buffered saline (PBS) as described previously (Boucias *et al.,* 2013). As a *G. pallidipes* colony free of overt SGH symptoms has been established in the IPCL, to obtain symptomatic flies, experimental flies were intra-hemocoelically injected with 2 µl of the virus inoculum, which was estimated to contain ~10^6 virus genome copies per µl of the virus suspension by quantitative polymerase chain reaction (qPCR) as described previously (Boucias *et al.,* 2013). The progeny of these artificially infected mothers were used for the experiments that required symptomatically infected flies (confirmed microscopically during salivary gland dissections). A lack of a virus-free *G. pallidipes* flies at the IPCL colony precluded inclusion of a non-infected control

group in the assays. Instead, flies of the same age as those of the above-described symptomatic fly group were directly sourced from the SGH-free colony and used for the asymptomatic (control) group.

RNA isolation, small RNA library construction and deep sequencing

Prior to RNA extraction, flies were individually dissected to confirm their SGH status (symptomatic or asymptomatic). The virus presence was not confirmed by PCR since the flies were progeny of GpSGHV-injected *G. pallidipes* parents, which are known to produce only asymptomatically or symptomatically infected progeny (Boucias *et al.,* 2013). Total RNA was extracted using Trizol (Invitrogen) from 10-day old flies, eight flies from each group of asymptomatically or symptomatically infected flies. As there is no *G. pallidipes* colony free of virus at the IPCL, a control group (non-infected) was not included in the assay. To prepare the sRNAs, the extracted total RNA was purified from denaturing polyacrylamide gel and sequentially ligated to the adapters for next generating sequencing (NGS) according to the manufacturer's instructions (Illumina Inc.) before sequencing. The RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies) (Masotti and Preckel, 2006). Two biological replicates from each group of barcoded libraries were then sequenced on Illumina Genome Analyzer Miseq (1X50 run) for 50 cycles to produce 3-5 million reads per library. Raw sequencing data have been submitted to the National Center for Biotechnology Information under the accession number SRP139935.

Small RNA analysis

The CLC genomic workbench version 11.0.1 was used to remove adapter sequences and low-quality sequence reads from the datasets by applying a quality trimming cut off score of 0.05. The sequence reads without the 3' adapters were discarded from the libraries. The small RNA tool of the CLC genomic workbench was used to extract and count unique sRNA reads. Clean sequence reads with lengths ranging from 18 to 30 nt were mapped against the *G. pallidipes* (GCA_000688715.1) genome found on the VectorBase database (Giraldo-Calderón *et al.,* 2015) and against the genome of the Ugandan GpSGHV strain (Accession Number: EF568108), allowing only a mismatch, insertion and deletion costs of 2, 3, and 3 respectively. The Rfam (RNA families) database was used to remove ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear ribonucleic acid RNAs

(snRNAs) and repeats from the sRNA sequences (Kalvari *et al.,* 2017). The remaining sequence reads were uploaded onto the CLC Genomic Workbench 'annotate and merge counts tool' to search for conserved precursors and mature miRNAs using insect miRNA sequences found in the miRBase 22.0 (Griffiths-Jones *et al.,* 2007) as reference. Only the perfectly or near-perfectly (1–2 mismatches) matching sequences were considered to represent conserved miRNAs.

To identify the GpSGHV-encoded miRNAs, the unique reads that mapped to the GpSGHV-Uga genome were combined with 150 nt either upstream or downstream from their position on the virus genome. The RNA secondary structures of the predicted pre-miRNA hairpins were analysed using RNAfold (Lorenz *et al.,* 2011). The hairpins were considered pre-miRNA if a mature miRNA was present in the arm of the hairpin precursor and the secondary structure was stable with low free energy of hybridization.

The putative GpSGHV-encoded miRNAs identified from the sRNA NGS data were then compared to the GpSGHV pre-miRNA hairpins predicted by the VMir software using the GpSGHV-Uga genome (Grundhoff, 2011). The miRNA hairpin prediction by VMir was initially performed with the program default settings and later the values for minimum score and window counts were adjusted to 115 and 30, respectively, to increase the stringency for the hairpin selection as previously described and optimized (Grundhoff, 2011; Hussain *et al.,* 2011).

miRNA differential expression

The miRNA expression for each independent biological replicate was normalized on the CLC genomic workbench using the option 'by totals', which applies tag (number of copies of the sRNAs) per million total RNA reads (TPM). The normalized mean values of the two replicates were used to compare miRNA abundance or expression in asymptomatic and symptomatic libraries. Because of the low abundance of some of the identified miRNAs, only the miRNAs with more than 10 raw reads in the libraries were included in the differential expression analysis. Changes in miRNA expression in the symptomatic versus asymptomatic flies were considered significant when their *P* values were below 0.05. The final fold change values were given in $\log_2$ scale and the miRNAs with $\log_2$-fold change ($\log_2$FC) higher than 0.2 or less than -0.1 were designated as up-regulated and down-regulated respectively in symptomatic flies.

Putative viral and host miRNA target identification and functional analysis

RNA22 and RNAhybrid software packages were used to predict the putative host and virus gene targets of the differentially expressed host- and GpSGHV-encoded miRNAs (Krüger and Rehmsmeier, 2006; Miranda *et al.,* 2006). The miRNA target prediction with the RNAhybrid was performed using default settings with energy threshold set to -20kcal/mol. The RNA22 software *P* value was set to 0.05 and a minimum free energy (mfe) of < -12.0 kcal/mol and the remaining parameters were set to default. Only putative target genes that were predicted by both software packages were selected for further analysis. Gene Ontology (GO) enrichment and pathway analysis of the miRNA-targeted genes was performed using Blast2GO version 5.1.13 (Conesa *et al.,* 2005). To select only the putative immune related genes and their immune pathways, the miRNA targeted genes were further analysed by protein blast (BLASTp; e-value $\leq 10^{-2}$) on the Insect Innate Immunity Database (IIID) (Brucker *et al.,* 2012). Based on the *P* values and the mfe values (lowest *P* value and minimum mfe) for miRNA-mRNA interaction, the top 10 immune related genes targeted by the regulated host miRNAs and the GpSGHV-encoded miRNAs were used to generate host-, GpSGHV-encoded miRNA and host mRNA interaction networks using Cytoscape (Shannon *et al.,* 2003).

RT-qPCR of miRNAs and their predicted putative mRNA targets

To validate the differentially expressed miRNAs during GpSGHV infections, reverse transcription qPCR (RT-qPCR) was used. Total RNA was extracted using Trizol as described above from eight asymptomatically and eight symptomatically infected flies. Complementary DNA (cDNA) was synthesised using the miSCRIPT II RT kit (Qiagen) using the Hiflex buffer which ensures cDNA synthesis of both miRNAs and mRNA molecules. The qPCR was performed using the miScript SYBR Green PCR kit (Qiagen), which includes an miRNA universal reverse primer. The forward primer in the reactions for miRNA quantification was derived from each of the specific miRNA sequences investigated in this study (**Table 1**).

The PCR program used to quantify the miRNAs was; 95°C for 15min, followed by forty cycles of 94°C for 15sec, 55°C for 30sec, and 70°C for 30sec. To determine the impact of the virus regulated host miRNAs on their mRNA transcript levels, the expression levels of the selected top 10 immune related genes targeted by the

miRNAs was assessed in the same asymptomatic and symptomatic flies using the primers listed in **Table 1**.

Table 1: Sequences for the primers used for expression analysis by RT-qPCR and the miR-184-3p mimic and inhibitor sequences.

Target	Primer name	Sequence
Primers for the differentially expressed host miRNAs		
miR-184-3p	miR-184-3p_F	AACTGGACGGAGAACTGATAAGGGC
miR-277-3p	miR-277-3p_F	TTGTAAATGCACTATCTGGTACGAC
miR-7-3p	miR-7-3p_F	CAACAAAATCACTAGTCTTCCA
miR-8-3p	miR-8-3p_F	TAATACTGTCAGGTAAAGATGTC
miR-999-3p	miR-999-3p_F	TGTTAACTGTAAGACTGTGTCT
miR-1-3p	miR-1-3p_F	TGGAATGTAAAGAAGTATGGAGCGA
miR-263-5p	miR-263-5p_F	AATGGCACTGGAAGAATTCACGG
miR-276-5p	miR-276-5p_F	AAGAGCACGGTATGAAGTTCCTA
miR-283-5p	miR-283-5p_F	AAATATCAGCTGGTAATTCTG
miR-6497	miR-6497_F	CGTAACTTCGGGATAAGGATTGGCTCTGAAG
miR-9-3p	miR-9-3p_F	TCATACAGCTAGATAACCAAAGA
Primers for the selected immune genes targeted by miRNAs		
GPAI025158	GPAI025158-qPCR_F	GTATTCCTCACACTTCCTCCAAC
	GPAI025158-qPCR_R	CCACCATAACTGAGAACAGAAGAA
GPAI030501	GPAI030501-qPCR_F	CGATGCTATGGGTTTTCTGCT
	GPAI030501-qPCR_R	TCGCATTTATTACCGCACAACA
GPAI014544	GPAI014544-qPCR_F	GGATGCGAGAACGGGAAATG
	GPAI014544-qPCR_R	CAAACACTCTTCCTGACAAAATGG
GPAI038987	GPAI038987-qPCR_F	CGGATTGGTTTAGTTTCGGTTG
	GPAI038987-qPCR_R	CCACTTCTTCTCTTTTCACTTTCTC
GPAI034557	GPAI034557-qPCR_F	TAATCGCTGGTTGGGTAATGAG
	GPAI034557-qPCR_R	GTTTGTATCTATTCGGTTCCTCCT
GPAI025990	GPAI025990-qPCR_F	GCAATACTTCCCTGTCCATAAC
	GPAI025990-qPCR_R	CTGTCGTCCAACCTTCACTT
GPAI001218	GPAI001218-qPCR_F	ATGAGGTGGATGAAAGTGATAAAGG
	GPAI001218-qPCR_R	CTTCCTCGGGTATGTCAATCAAG
GPAI015640	GPAI015640-qPCR_F	GCATACCTTTTCTGTTGGTTGG
	GPAI015640-qPCR_R	CGAGTTTTGGCTGATGTTTCTAC
GPAI042543	GPAI042543-qPCR_F	CAAATCACGCATAGCCACAAG
	GPAI042543-qPCR_R	AATGGGTTTAGTGGAGGGTTTC
GPAI007448	GPAI007448-qPCR_F	TGCGACAAAAGCTAGATGTAATGGG
	GPAI007448-qPCR_R	AAATCCTCAAACACAGCACCAACA
Primers for the virus and reference gene quantification		
GpSGHV (*odv-e66*)	GpSGHVqPCR-F	CAAATGATCCGTCGTGGTAGAA
	GpSGHVqPCR-F	AAGCCGATTATGTCATGGAAGG
β-Tubulin (tsetse)	Tse-TubqPCR-F	GATGGTCAAGTGCGATCCT
	Tse-TubqPCR-R	TGAGAACTCGCCTTCTTCC
Mir-184 mimic and inhibitor sequences		
miR-184-3p	Mimic	UGGACGGAGAACUGAUAAGGGC
	Inhibitor	GCCCUUAUCAGUUCUCCGUCCA

The virus infection level was estimated by quantifying the expression of the conserved GpSGHV *odv-e66* gene whose transcript levels have been correlated to the total virus copy numbers (Abd-Alla *et al.,* 2009b) using the primers included in **Table 1**. The PCR program to quantify the *odv-e66* and the selected immune genes was; 95°C for 15 min, followed by forty cycles of 94°C for 15 sec, 60°C for 30 sec, and 70°C for 30 sec. Two technical replicates were included for each reaction and all the target genes and miRNA expressions were normalized to *β-tubulin* gene expression using previously described primers (Caljon *et al.,* 2009).

Inhibition of miR-184-3p in G. pallidipes

To investigate the role of the most up-regulated host miRNA and with high abudance in symptomatically infected flies compared to the asymptomatic individuals (i.e. miR-184-3p), an inhibitor and a mimic of this miRNA were synthesised (Thermofisher company; Waltham, USA) and subsequently injected into the flies. The details of the inhibitor and the mimic sequences are included in **Table 1**. Prior to the above-mentioned injections, teneral (24 h post eclosion; un-fed) adult flies were anaesthetised on ice, and then injected in the thorax with 10 pmol (the doze optimized in this study) of either the inhibitor or the mimic or RNase free water (40 females and 40 males per group). Two days post injection with the inhibitor, mimic, or RNase-free water, half of the flies from each group (20 females and 20 males) were injected with PBS (control), and the other half of the flies were injected with 2 µl of the virus inoculum as described above. The ability of GpSGHV-injected female parents to induce symptomatic infections in the progeny depends on the increase in virus titre in parents (Boucias *et al.,* 2013). Therefore, in this study, samples from the parental generation (i.e. 3 females and 3 males), which were collected at zero and 21 days post PBS/virus injection followed by RNA extraction and cDNA synthesis were analsyed using miSCRIPT II RT kit as described above analysed. To validate the success of the inhibitor and the mimic of miR-184-3p, the expression levels of miR-184-3p in the collected samples were assessed using the miScript SYBR Green PCR kit. The effect of injecting miR-184-3p inhibitor or mimic on virus infection was assesed by quantifying the GpSGHV *odv-e66* gene expression on the day of injection and 21 days post PBS/virus injection. The expression levels were normalized to *β-tubulin* gene as described in the previous section.

Removal of adaptors, contaminants and low-quality reads resulted in ~20,166 and ~16,309 reads from the asymptomatic and symptomatic libraries, respectively. Of these, 16,422 and 13,077 clean reads from the asymptomatic and symptomatic libraries respectively, were mapped onto the *G. pallidipes* genome (**Table 2**). A summary of the length distribution of the clean reads that mapped onto the *G. pallidipes* genome are shown in **Figure 1.**

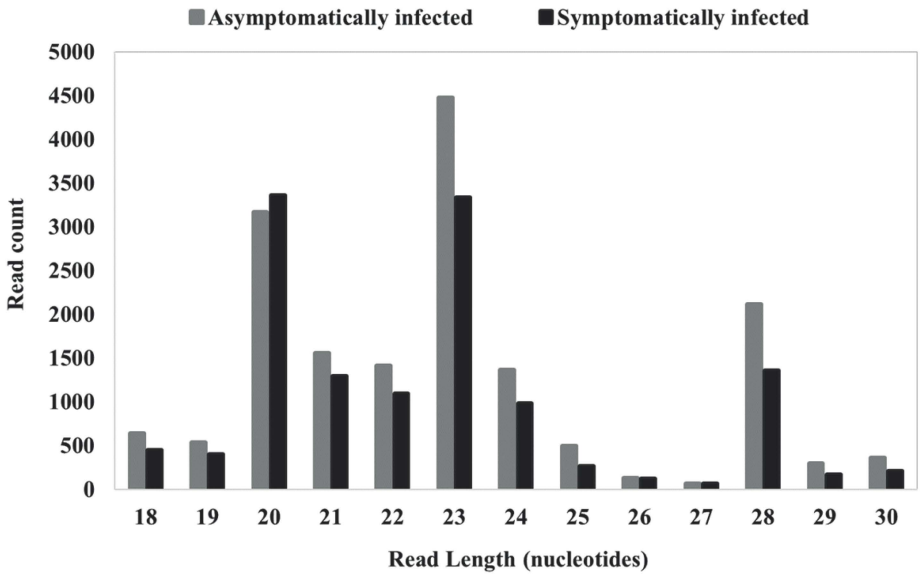


Figure 1: Length distribution of reads mapped onto the *G. pallidipes* genome from asymptomatically (grey bar) and symptomatically (black bar) infected flies.

As evidenced in **Figure 1**, the number of sRNA over length distribution was less in symptomatic library compared to the asymptomatic libraries. Both libraries showed a peak at 20-23 nt, which may represent the class of miRNA or siRNAs, and another frequency peak was observed at a length of 28 nt, which may represent the piRNAs. Mapping of the clean reads onto the GpSGHV genome resulted in three viral sRNA from the asymptomatic compared to 53 sRNAs from symptomatic flies (**Table 2**).

Identification of host- and virus- encoded miRNAs

Analyses of the sRNAs in *G. pallidipes* revealed that 1,675 and 1,207 reads putatively coded for host miRNAs in asymptomatic and symptomatic flies, respectively (**Table 2**). A total of 57 host miRNAs were identified, which were named with sequential numbers following their gene family as designated in the miRBase. Of the 57 host miRNAs, 38 miRNAs were expressed in both asymptomatic and symptomatic flies, while nine and 10 of the miRNAs were found only in asymptomatic and symptomatic flies, respectively (**Table 3**).

Table 3: miRNAs identified in sRNA libraries of asymptomatically and symptomatically infected *G. pallidipes*. Total reads of each of the identified miRNA, their normalized expression levels, the mature miRNA sequence and the miRNA gene families are shown. Abbreviation: Asymp = Asymptomatic, Symp = Symptomatic, R = Replicate

miRNA Name	Mature sequence	Gene family	Raw read counts				Normalized read counts				Normalized mean read counts	
			Asymp R1	Symp R1	Asymp R2	Symp R2	Asymp R1	Symp R1	Asymp R1	Symp R1	Asymp	Symp
miR-276-3p	AAGAGCACGGTATGAAGTTCCTA	MIPF0000124; mir-276	6	3	16	16	1313,77	673,25	1025,71	1349,87	1169,74	1011,56
miR-276-5p	TAGGAACTCTATACCTCGCT	MIPF0000124; mir-276	4	9	23	17	875,85	2019,75	1474,45	1434,24	1175,15	1726,99
miR-1-3p	TGGAATGTAAAGAAGTATGGAG	MIPF0000038; mir-1	97	111	585	322	21239,33	24910,23	37502,40	27166,12	29370,86	26038,18
miR-1-5p	CTTCCATACTTCTTTACATTCCA	MIPF0000038; mir-1	11	6	36	27	2408,58	1346,50	2307,84	2277,90	2358,21	1812,20
miR-277-3p	TAAATGCACTATCTGGTACGAC	MIPF0000156; mir-277	45	27	118	106	9853,30	6059,25	7564,59	8942,88	8708,94	7501,06
bantam-5p	AAATCAGCTTTCAAAATGATCTCAC	MIPF0000153; bantam	7	8	42	25	1532,73	1795,33	2692,48	2109,17	2112,61	1952,25
miR-8-3p	TAATACTGTCAGGTAAAGATGTC	MIPF0000019; mir-8	25	24	156	135	5474,05	5386,00	10000,64	11389,52	7737,35	8387,76
miR-8-5p	CATCTTACCGGGCAGCATTATAA	MIPF0000019; mir-8	3	0	4	0	656,89	0,00	256,43	0,00	456,66	0,00
miR-9-3p	TCATACAGCTAGATAACCAAAGA	MIPF0000014; mir-9	18	20	58	30	3941,32	4488,33	3718,19	2531,00	3829,75	3509,67
miR-9-5p	TCTTTGGTATTCTAGCTGTAAA	MIPF0000014; mir-9	5	3	17	2	1094,81	673,25	1089,81	168,73	1092,31	420,99
miR-31-5p	TGGCAAGATGTCGGCATAGCTG	MIPF0000064; mir-31	10	6	62	44	2189,62	1346,50	3974,61	3712,14	3082,12	2529,32
miR-6497	CGTAACTTCGGGATAAGGATTGGCTCTGA	MIPF0001965; mir-6497	13	0	75	29	2846,51	0,00	4808,00	2446,64	3827,25	1223,32
miR-283-5p	AAATATCAGCTGGTAATTCTG	MIPF0000054; mir-216	2	3	26	16	437,92	673,25	1666,77	1349,87	1052,35	1011,56
miR-7-3p	CAACAAAATCACTAGTCTTCCA	MIPF0000022; mir-7	5	3	25	23	1094,81	673,25	1602,67	1940,44	1348,74	1306,84
miR-996-3p	CTAAATAATCAAGTTCGGTCAACTTT	MIPF0000449; mir-996	0	0	1	0	0,00	0,00	64,11	0,00	32,05	0,00
miR-184-3p	TGGACGGAGAACTGATAAGGGC	MIPF0000059; mir-184	7	9	23	42	1532,73	2019,75	1474,45	3543,41	1503,59	2781,58
miR-263-5p	AATGGCACTGGAAGAATTCACGG	MIPF0000122; mir-263	2	3	14	11	437,92	673,25	897,49	928,04	667,71	800,64

miRNA Name	Mature sequence	Gene family	Raw read counts				Normalized read counts				Normalized mean read counts	
			Asymp R1	Symp R1	Asymp R2	Symp R2	Asymp R1	Symp R1	Asymp R1	Symp R1	Asymp	Symp
miR-999-3p	TGTTAACTGTAAGACTGTGTCT	MIPF0000852; mir-999	4	1	17	10	875,85	224,42	1089,81	843,67	982,83	534,04
miR-10-5p	AAACCTCTCTAGAACCGAATTTG	MIPF0000033; mir-10	0	0	0	1	0,00	0,00	0,00	84,37	0,00	42,18
miR-10-3p	AACAAATTCGGATCTACAGGGT	MIPF0000033; mir-10	1	0	1	1	218,96	0,00	64,11	84,37	141,53	42,18
miR-279-3p	TGACTAGATCCACACTCATT	MIPF0000184; mir-279	1	5	16	14	218,96	1122,08	1025,71	1181,14	622,33	1151,61
miR-305-5p	ATTGTACTTCATCAGGTGCTCTGG	MIPF0000158; mir-305	3	3	7	6	656,89	673,25	448,75	506,20	552,82	589,73
miR-305-3p	CGGCACATGTTGAAGTACATTCAA	MIPF0000158; mir-305	0	0	0	1	0,00	0,00	0,00	84,37	0,00	42,18
miR-281-3p	CTGTCATGGAATTGCTCTCTTT	MIPF0000087; mir-46	2	6	9	5	437,92	1346,50	576,96	421,83	507,44	884,17
miR-281-5p	AAGAGAGCTATCTGTCGACAGTA	MIPF0000087; mir-46	0	0	0	2	0,00	0,00	0,00	168,73	0,00	84,37
miR-100-5p	AACCCGTAAATCCGAACTTGTG	MIPF0000033; mir-10	2	0	3	3	437,92	0,00	192,32	253,10	315,12	126,55
miR-278-3p	TCGGTGGGACTTTCGTCCGTTT	MIPF0000155; mir-278	2	1	3	3	437,92	224,42	192,32	253,10	315,12	238,76
miR-284-5p	CCTGGAATTAAGTTGACTGTGCA	MIPF0000228; mir-284	2	1	4	2	437,92	224,42	256,43	168,73	347,18	196,58
lin-4-5p	TCCCTGAGACCCTAACTTGTGA	MIPF0000303; lin-4	0	0	1	1	0,00	0,00	64,11	84,37	32,05	42,18
lin-4-3p	ACAAGTTTTGATCTCAGGTATA	MIPF0000303; lin-4	0	0	1	0	0,00	0,00	64,11	0,00	32,05	0,00
miR-987-5p	TAAAGTAAATAGTCTGGATTGATG	MIPF0000859; mir-987	0	1	2	0	0,00	224,42	128,21	0,00	64,11	112,21
miR-2-5p	AGCTCATCAAAGCTGGCTGTGATA	MIPF0000049; mir-2	1	0	2	0	218,96	0,00	128,21	0,00	173,59	0,00
miR-33-5p	GTGCATTGTAGTCGCATTGTC	MIPF0000070; mir-33	0	1	3	3	0,00	224,42	192,32	253,10	96,16	238,76
miR-6-3p	TATCACAGTGGCTGTTCTTTTT	MIPF0000119; mir-6	0	0	0	1	0,00	0,00	0,00	84,37	0,00	42,18
miR-87-3p	TTGAGCAAAATTTCAGGTGT	MIPF0000152; mir-87	0	0	1	1	0,00	0,00	64,11	84,37	32,05	42,18
miR-124-5p	GGCATTCACCGCGTGCCTTA	MIPF0000021; mir-124	0	0	1	1	0,00	0,00	64,11	84,37	32,05	42,18
miR-190-5p	AGATATGTTTGATATTCTTG	MIPF0000076; mir-190	0	0	1	0	0,00	0,00	64,11	0,00	32,05	0,00

miRNA Name	Mature sequence	Gene family	Raw read counts				Normalized read counts				Normalized mean read counts	
			Asymp R1	Symp R1	Asymp R2	Symp R2	Asymp R1	Symp R1	Asymp R1	Symp R1	Asymp	Symp
miR-210-3p	CTTGTGCGTGTGACAGCGG	MIPF0000086; mir-210	0	1	1	0	0,00	224,42	64,11	0,00	32,05	112,21
miR-449-3p	TGGCAGTGTGGTTAGCTGGTT	MIPF0000133; mir-449	0	0	1	0	0,00	0,00	64,11	0,00	32,05	0,00
miR-449-5p	TGGCAGTGTGGTTAGCTGGTTA	MIPF0000133; mir-449	0	1	1	0	0,00	224,42	64,11	0,00	32,05	112,21
miR-989-3p	TGTGATGTGACGTAGTGGAAC	MIPF0000885; mir-989	0	0	1	0	0,00	0,00	64,11	0,00	32,05	0,00
miR-994-5p	CTAAGGAAATAGTAGCCGTGAT	MIPF0001045; mir-994	0	1	1	3	0,00	224,42	64,11	253,10	32,05	238,76
miR-315-5p	GGCTTTCTGAGCAACAATCAAAA	MIPF0000141; mir-315	1	0	1	0	218,96	0,00	64,11	0,00	141,53	0,00
let-7-3p	CTATACAACGTGCTAGCTTTCT	MIPF0000002; let-7	3	2	3	4	656,89	448,83	192,32	337,47	424,60	393,15
let-7-5p	CGAGGTAGTAGGTTGTATAGTA	MIPF0000002; let-7	2	5	4	3	437,92	1122,08	256,43	253,10	347,18	687,59
miR-1000-5p	ATATTGTCCTGTCACAGCAG	MIPF0000701; mir-1000	1	0	2	0	218,96	0,00	128,21	0,00	173,59	0,00
miR-252-5p	CTAAGTACTAGTGCCGCAGGAGA	MIPF0000285; mir-252	0	1	0	0	0,00	224,42	0,00	0,00	0,00	112,21
miR-137-3p	TATTGCTTGAGAATACACGTAG	MIPF0000106; mir-137	0	0	0	1	0,00	0,00	0,00	84,37	0,00	42,18
miR-133-3p	TTGGTCCCCTTCAACCAGCTG	MIPF0000029; mir-133	0	0	0	2	0,00	0,00	0,00	168,73	0,00	84,37
miR-13-3p	TATCACAGCCATTTTGACGAGT	MIPF0000049; mir-2	2	1	1	3	437,92	224,42	64,11	253,10	251,02	238,76
miR-375-3p	TTTGTTCGTTTGGCTTAAGTT	MIPF0000114; mir-375	0	1	0	2	0,00	224,42	0,00	168,73	0,00	196,58
miR-14-3p	TCAGTCTTTTTCTCCCTCCTATC	MIPF0000182; mir-14	3	1	8	5	656,89	224,42	512,85	421,83	584,87	323,13
miR-11-3p	CATCACAGTCTGCGTTCTTGCA	MIPF0000252; mir-11	0	2	3	6	0,00	448,83	192,32	506,20	96,16	477,52
miR-67-3p	TCACAACCTCTTTGAGTGAGCTA	MIPF0000293; mir-67	0	0	2	1	0,00	0,00	128,21	84,37	64,11	42,18
miR-927-3p	TGGTAAAGCGTAGGAATTCTAAA	MIPF0000452; mir-927	0	1	3	2	0,00	224,42	192,32	168,73	96,16	196,58
miR-927-5p	TTTAGAATTCCTCCGCTTTACCA	MIPF0000452; mir-927	0	0	0	3	0,00	0,00	0,00	253,10	0,00	126,55
miR-275-3p	TCCGGTACCTGAAGTAGCGCGCGA	MIPF0000187; mir-275	0	0	0	1	0,00	0,00	0,00	84,37	0,00	42,18

Notably, six of the 57 identified host miRNAs (i.e. miR-1-3p, miR-184-3p, miR-263-5p, miR-277-3p, miR-283-5p, and miR-8-3p) were found to be conserved in five insect species (*Ae. aegypti*, *Anopheles gambiae*, *Bombyx mori*, *Drosophila melanogaster* and *Apis mellifera*) from the miRBase (data not shown).

Based on the criteria for pre-miRNA prediction (i.e. the presence of mature miRNA and mfe < -20 kcal/mol) and the secondary structure analysis, six putative GpSGHV-encoded miRNAs were identified from the 60 sRNA reads that mapped to the GpSGHV-Uga genome from the symptomatic library (**Table 2**). The identified GpSGHV-encoded miRNAs were named according to their position and orientation on the GpSGHV-Uga genome. These six putative GpSGHV-encoded miRNAs were mir-GpSGHV_164791F, mir-GpSGHV_170050R, mir-GpSGHV_165482F, mir-GpSGHV_165479R, mir-GpSGHV_151557R, and mir-GpSGHV_165975R (**Table 4**).

Table 4: VMir predicted and sequenced GpSGHV pre-miRNA and their characteristics

MiRNA name	Pre-miRNA Length	Mfe (kcal /mol)	Strand	Position GpSGHV-Uga	VMir score	Mature miRNA Length	Mature miRNA Location	Mature miRNA sequence (5'-3')
Mir-GpSGHV_170050R	145	-41.60	Rev	170050-170194	254.9	22	3p	CTACTTGGAGA TATAATAGAAG
Mir-GpSGHV_165479R	87	-22.70	Rev	165479-165565	217.9	22	3p	AAATGGATCGC TGTAGTTTTAA
Mir-GpSGHV_165481F	83	-21.0	For	165481-165564	214.1	22	3p	TGGATTACTCT GGTTTTAACTT
Mir-GpSGHV_165975R	87	-23.60	Rev	165975-166062	180.7	22	3p	ATGGTTGAGAT TCTTCAGATCG
Mir-GpSGHV_164791F	66	-19.0	For	164791-164857	137.4	22	3p	TGGATCAATGT ATTTCCATCTC
Mir-GpSGHV_151557R	55	-9.30	Rev	151532-151583	130	22	3p	GGACGTGTCAT TATATAATCGG

The secondary structures of the GpSGHV-encoded pre-miRNAs were analysed and revealed the 3' overhangs associated with the Dicer/Drosha-mediated processing and the mature miRNA sequences (**Figure 2A**). Since most virus-encoded miRNAs appeared to be localized antisense to their viral transcripts, which could be the

obvious potential targets (Sullivan *et al.*, 2005), the possible GpSGHV-encoded miRNAs interacting with the respective GpSGHV transcripts were identified (**Figure 2B**). From the three sRNA reads that mapped to the GpSGHV-Uga genome from the asymptomatic library, only one read was identified as mir-GpSGHV_170050R; this putative viral miRNA was also present in the symptomatic sRNA library.

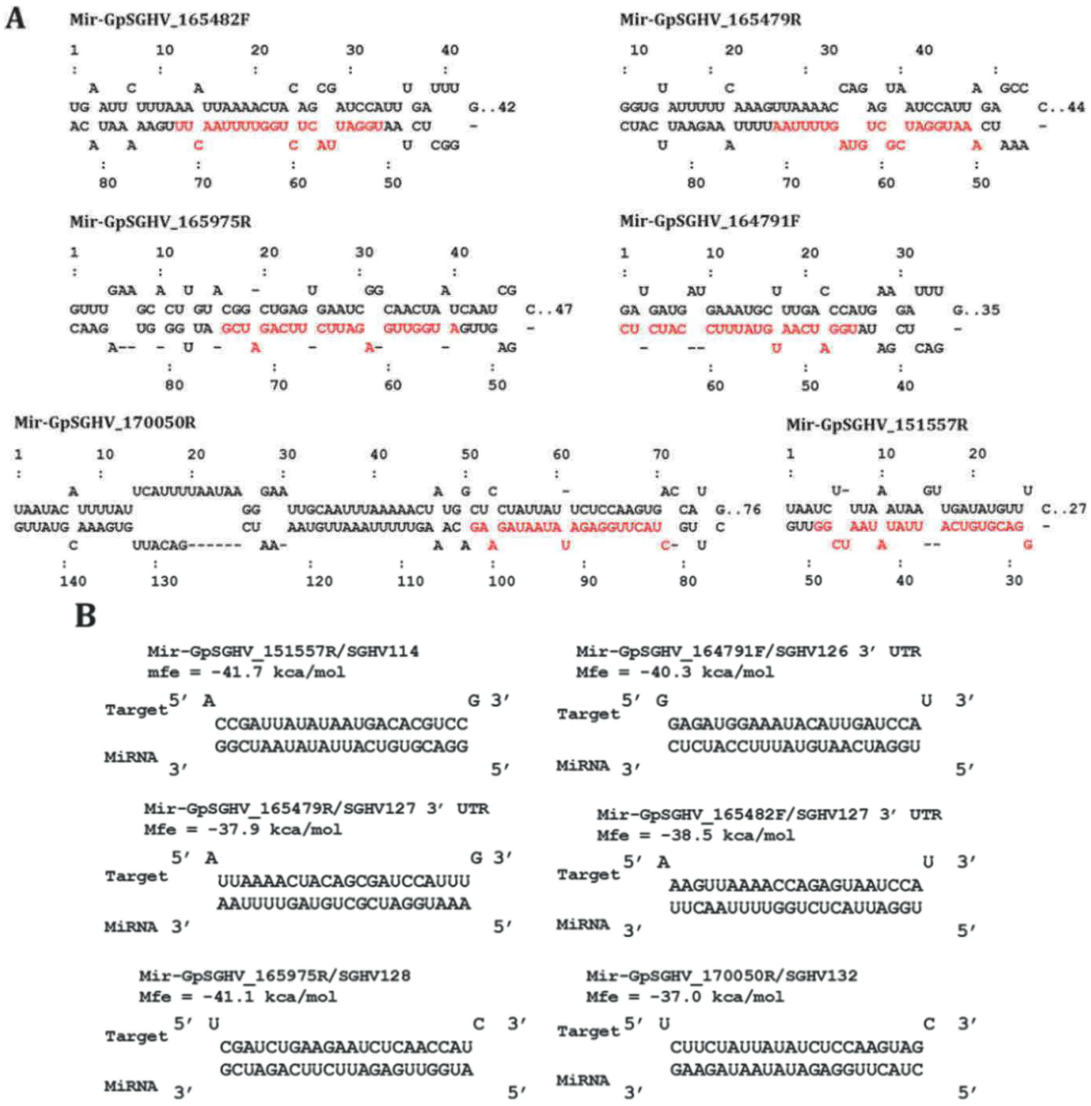


Figure 2: GpSGHV-encoded miRNAs identified from the NGS data. (A) Secondary structures of the GpSGHV pre-miRNAs named according to their position and orientation on the GpSGHV-Uga genome (Accession number: EF568108). Mature miRNA sequences are presented in red. (B) Interaction of mature GpSGHV-encoded miRNA sequences and their possible viral mRNA targets.

The identified viral miRNAs from the sRNA libraries were compared to the VMir predicted pre-miRNA hairpins, whereby the GpSGHV-Uga genome was supplied to the VMir program, and the initial pre-miRNA hairpin search (without filtering) detected a total of 2,328 main hairpins (MHPs). After filtering with settings of 115 and 35 as the values for minimum scores and window counts, respectively, 167 pre-miRNA hairpins were selected. Notably, four of the six GpSGHV-encoded miRNAs identified from the Illumina sequencing library were among the pre-miRNA hairpins that were predicted with high scores (**Figure 3**).

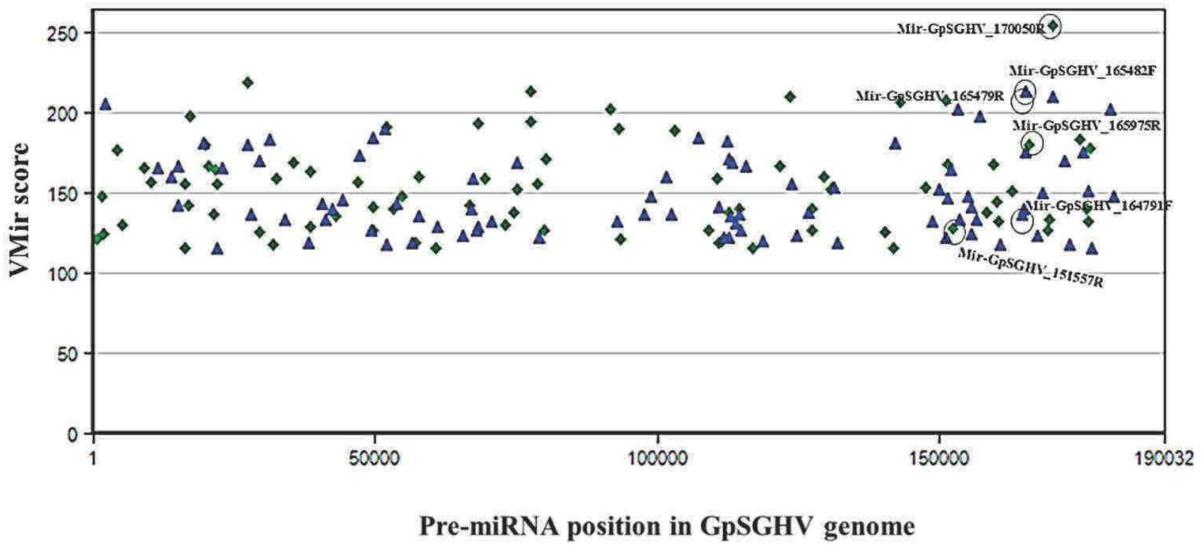


Figure 3: Distribution of predicted pre-miRNA hairpins on the GpSGHV-Uga genome. The hairpins are plotted according to their location on GpSGHV-Uga genome and their VMir score. Only the hairpins with VMir score of above 115 and that can fold in 35 or more windows are plotted. The green diamonds indicate the hairpins on the reverse orientation while the blue triangles indicate the hairpins on the forward orientation. The hairpins corresponding to the miRNAs identified by NGS are circled in black.

The locations of the predicted pre-miRNA hairpins on the GpSGHV genome and their VMir scores are also presented in **Figure 3**, showing that the high-scoring pre-miRNA hairpins are located between nucleotides 143,000 and 180,000 of the GpSGHV genome. This is similar to the locations of the experimentally obtained Illumina sequenced viral miRNAs, which are circled in green in **Figure 3**.

Differential expression of miRNAs in asymptomatic and symptomatic flies

Heat mapping of the above-mentioned 38 host miRNAs identified in both libraries of *G. pallidipes* revealed different expression patterns between the asymptomatic and symptomatic flies (**Figure 4A**). After excluding the host miRNAs with less than 10 raw reads from the asymptomatic and symptomatic libraries, analyses were done on differential expression of the 17 remaining host miRNAs. Based on the log_2FC (> 0.1 or < -0.1), 15 miRNAs were considered as differentially expressed, of which 10 were down-regulated and five were up-regulated in the symptomatic flies.

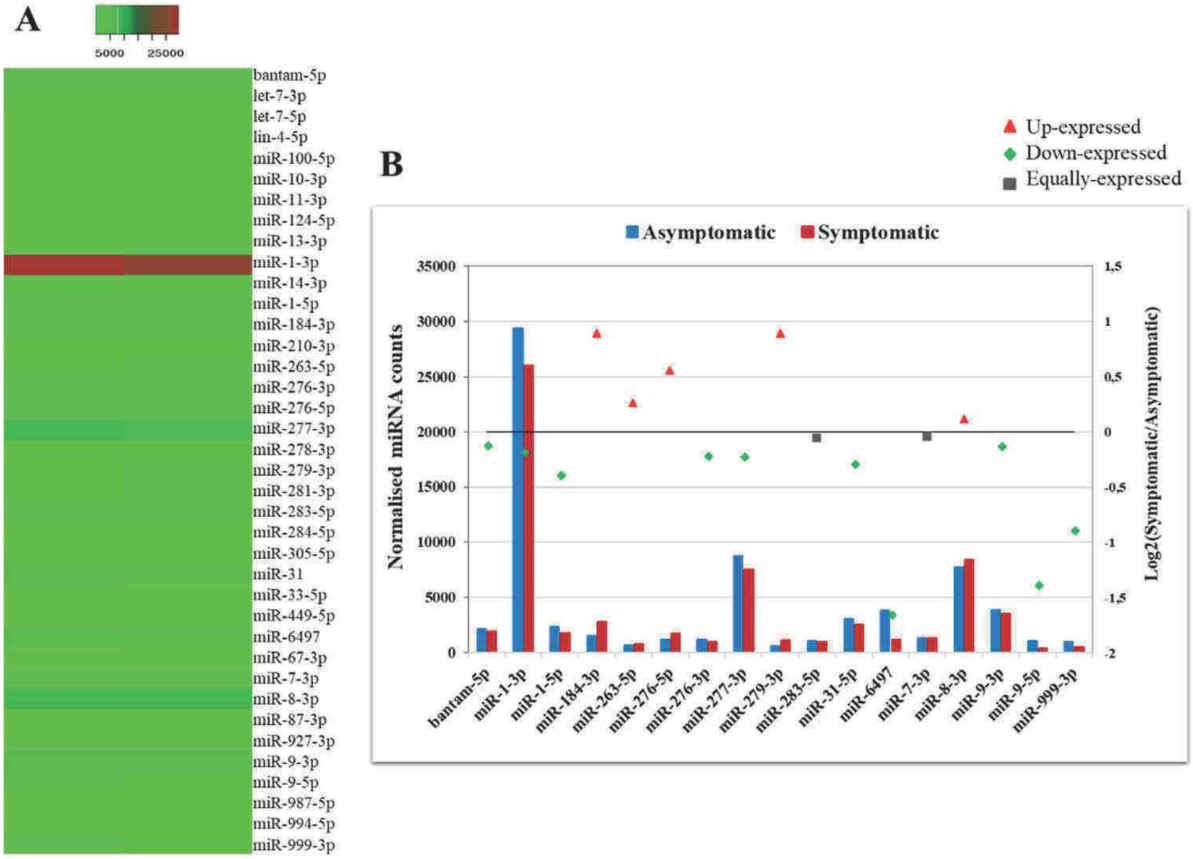


Figure 4: Differential expression of miRNAs in asymptomatic and symptomatic flies. (A) Heat map of the co-expressed host miRNAs. The highly expressed miRNAs are shown in red and the low expressed in green with absolute signal intensity ranging from 1 to 25000. (B) Expression analysis of the host miRNAs with more than 10 raw reads. The horizontal axis, the left vertical axis and the right vertical axis indicate the miRNA, the normalized miRNA expression values and the log2 fold change of the miRNA respectively. The log2 fold change (Log2FC) of the up-regulated miRNAs are shown in red, down-regulated in green and equally expressed in black.

The up-regulated (enriched in symptomatic flies compared to their asymptomatic counterparts) miRNAs were miR-184-3p, miR-279-3p, miR-276-5p, miR-263-5p, and miR-8-3p with log2FC values of 0.9, 0.9, 0.6, 0.3 and 0.1, respectively. Among the down-regulated (depleted in symptomatic flies) miRNAs with maximum log2FC were miR-6497, miR-9-5p, miR-999-3p, miR-1-5p and miR-31-5p with log2FC of -1.6, -1.4, -0.9, -0.4 and -0.3, respectively. There was no change in the expression (0.0 $\log_2$FC) of miR-283-5p and miR-7-3p, and hence these were considered equally expressed in both asymptomatic and symptomatic fly libraries (**Figure 4B**).

Validation of differentially expressed miRNAs in G. pallidipes by RT-qPCR

To validate the expression levels of some of the miRNAs that were differentially expressed according to the NGS data, RT-qPCR analysis was performed on a separate pool of asymptomatic and symptomatic flies to further quantify these miRNAs. The RT-qPCR results showed that the differences in the expression levels of the analysed miRNAs when comparing asymptomatic and symptomatic flies was mostly consistent with the earlier observed differences in the NGS data. For instance, the up-regulated miRNAs, miR-184-3p, miR-276-5p, miR-263-5p, and miR-8-3p, according to the NGS analysis, were up-regulated during symptomatic infection by RT-qPCR approach. The miRNAs that were considered down-regulated (e.g., miR-6497, miR-1-3p, miR-277-3p and miR-999-3p) based on the NGS data, showed no significant change in their expression when using the RT-qPCR approach (**Figure 5A**). The only inconsistency was observed for the equally expressed miRNAs (miR-283-5p and miR-7-3p) and miR-9-3p (slighlty down-regulated) in symptomatic flies according to the NGS data, but showed up-regulation with the RT-qPCR analysis approach (**Figure 5B**).

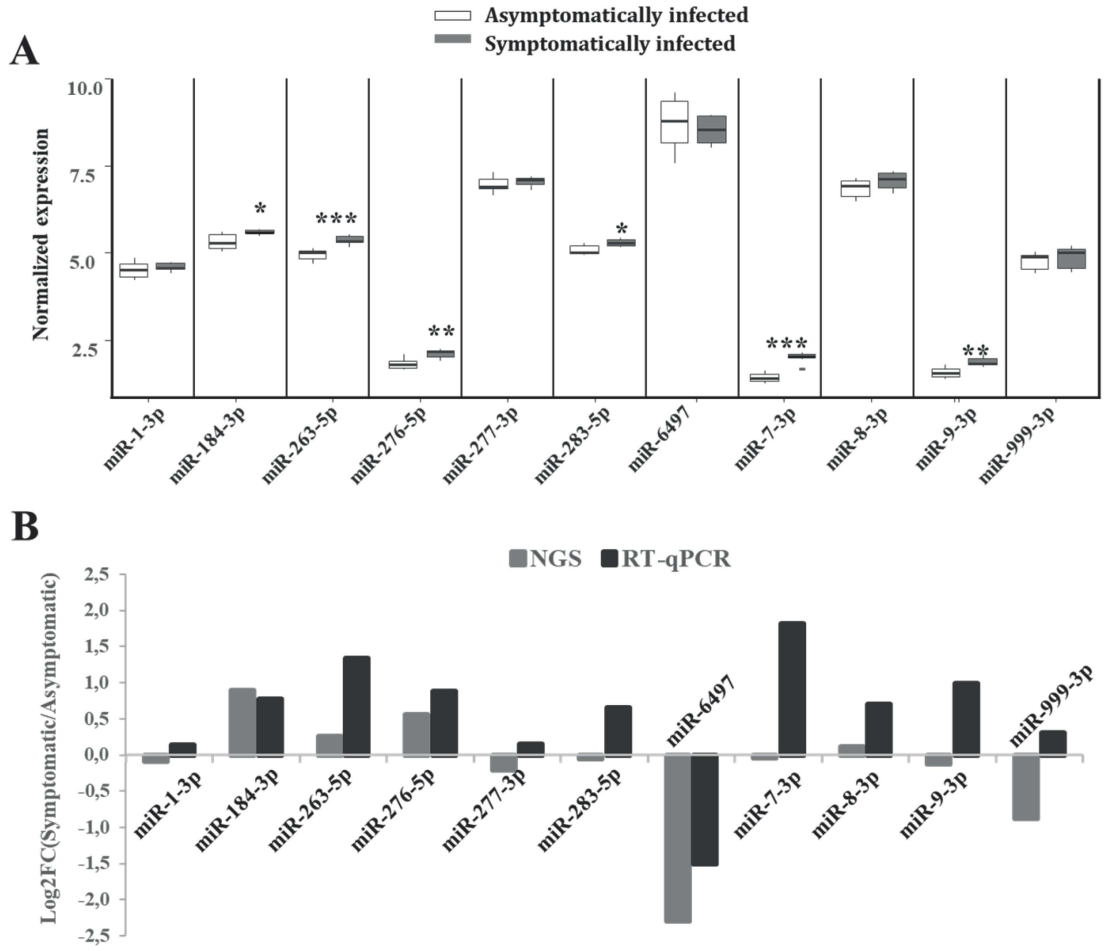


Figure 5: Differential expression of *G. pallidipes* miRNAs upon symptomatic infection. (A) Reverse transcription qPCR data analysis of selected host miRNAs: Expression profiles of miRNAs in asymptomatic and symptomatic *G. pallidipes*. Error bars show the standard deviation from the replicates. Asterisks indicate the statistical significance: ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$. (B) The graph shows a comparison of Log2 fold changes of *G. pallidipes* miRNAs based on NGS and Reverse transcription qPCR (RT-qPCR) analysis.

Prediction of mRNA targets for the differentially expressed host miRNAs

The analysis for potential mRNA targets of the 15 differentially expressed miRNAs as identified by NGS revealed that the miRNAs up-regulated in symptomatic flies potentially targeted about 715 putative host mRNAs, compared to 757 genes that potentially targeted by the down-regulated miRNAs (**Supplementary material Table 4A and 4B; DOI: 10.3389/fmicb.2018.02037**). Of the 715 putative genes targeted by the miRNAs up-regulated in symptomatic flies, 154 were immune-related such as Ras-related protein-27 (*Rab27*), homeodomain interacting protein

kinase (*Hipk*) and apolipoprotein lipid transfer particle (*Apoltp*) (**Supplementary material Table 4C; DOI: 10.3389/fmicb.2018.02037**). The blast search using the IIID software revealed that the immune-related genes could be involved in various immune signalling pathways, mostly in immune deficiency (Imd), Humoral response and Toll pathways, and some of these genes were involved in multiple immune pathways. Approxiamately, 70% (108/154) of these immune genes targeted by the up-regulated miRNAs were targeted by miR-184-3p, which was among the most highly up-regulated miRNA (0.9 log2FC) in symptomatic flies (**Figure 6**) and one of the miRNAs confirmed to be up-regulated with the RT-qPCR experiment (**Figure 5A** and **B**).

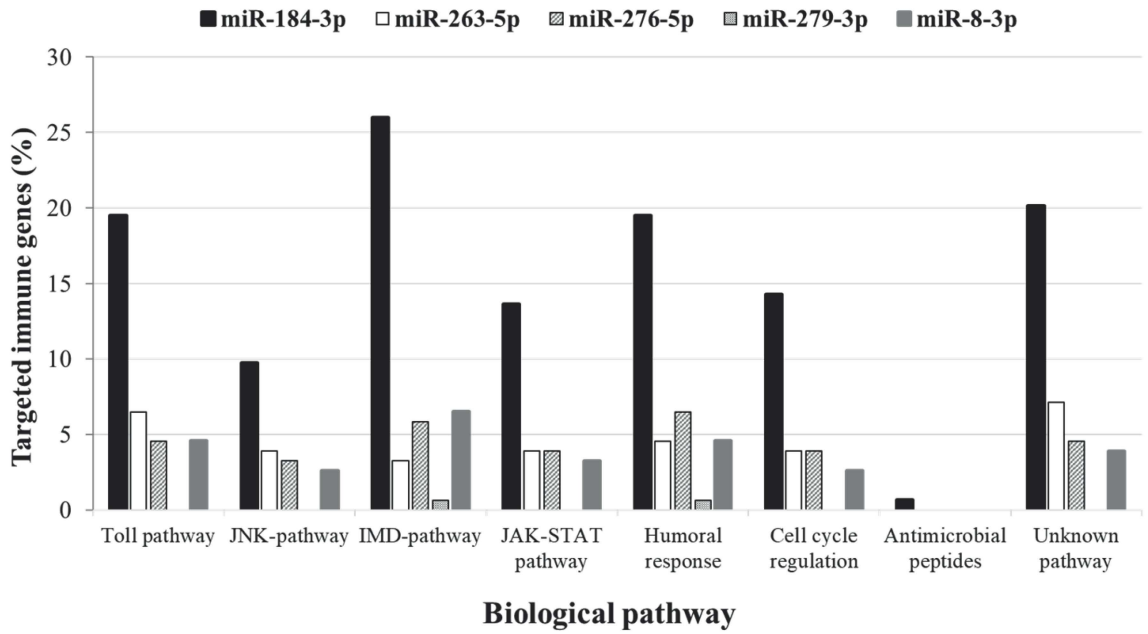


Figure 6: Biological immune pathways of the immune related genes targeted by the up-regulated miRNAs; miR-184-3p (black bars) miR-263-5p (white bars), miR-276-5p (striped bars), miR-279-3p (dotted bars) and miR-8-3p (grey bars). The percentages of the immune genes regulated by each miRNA are shown in different colors for each immune pathway.

Notably, only one of these 154 immune-related genes was a potential target of miR-279-3p, the other most up-regulated miRNA (0.9 log2FC) in symptomatic flies. Additionally, about 30% (35/108) of the miR-184-3p immune targeted genes were also targeted by the identified GpSGHV-encoded miRNAs (**Supplementary material Table 4D; DOI: 10.3389/fmicb.2018.02037**). The possible interactions

between the host modulated miRNAs, the GpSGHV-encoded miRNAs and a selection of their targeted genes were determined, and network produced using Cytoscape (**Figure 7; Supplementary materials Table 4E available at; DOI: 10.3389/fmicb.2018.02037**)

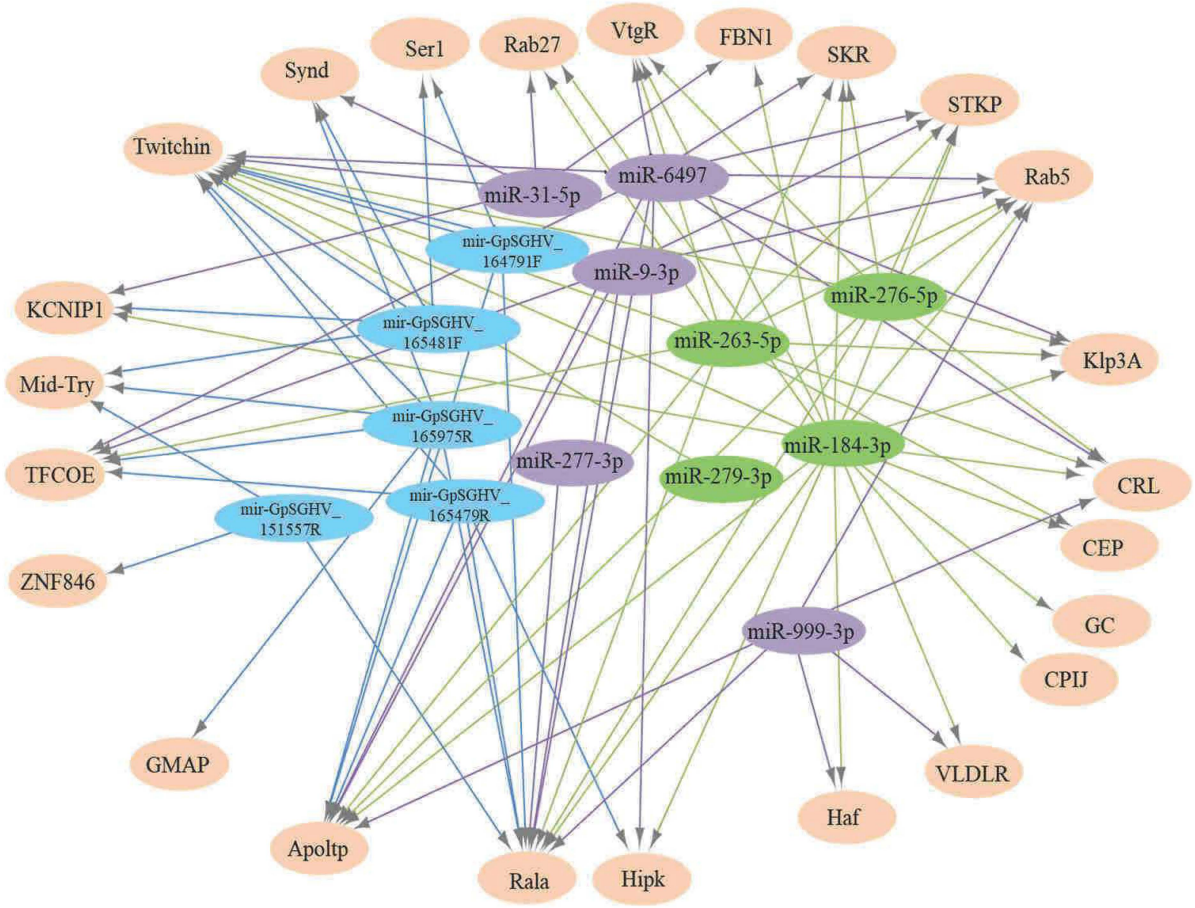


Figure 7: A network of interactions between up-regulated (in green) and down-regulated (in purple) host miRNAs, GpSGHV-encoded miRNAs (in blue) and some selected target genes (light orange). The network was visualized using Cytoscape with miRNAs as the source nodes and the mRNA as the targets. Abbreviations: *KCNIP1* (Kv channel-interacting protein 1), *ZNF846* (Zinc finger protein 846), *CEP* (Centrosomal protein), *GC* (Guanylyl cyclase), *FBN1* (fibrillin-1), *Klp3A* (Kinesin-like protein at 3A), *GMAP* (Golgi microtubule-associated protein), *Rala* (Ras-related protein Ral-a), *VLDLR* (Very low-density lipoprotein receptor domain class A), *CPIJ* (Cuticular protein 50Cb), *SKR* (Serine/threonine-protein kinase receptor), *Hipk* (homeodomain interacting protein kinase), *TFCOE* (Transcription factor collier), *Rab27* (Ras-related protein Rab-27A), *CRL* (cytokine receptor-like), *Synd* (Syndapin), *Ser1* (Serine protease 1), *Apoltp* (Apolipoprotein lipid transfer particle), *Mid-Try* (Midgut trypsin), *STKP* (Protein-serine/threonine kinase), *VtgR* (Vitellogenin receptor), *haf* (hattifattener).

To explore the impact of the regulated miRNAs on host immunity, RT-qPCR expression analysis of 10 potentially targeted immune genes (Centrosomal protein (*CEP*), fibrillin-1 (*FBN1*), Ras-related protein (*Ral-a*), *Hipk*, *Rab27*, *Apoltp*, Transcription factor collier (*TFCOE*), Protein-serine/threonine kinase (*STKP), Twitchin* and Vitellogenin receptor (*VtgR*)) was performed on asymptomatically and symptomatically infected flies (**Table 5**).

Table 5: The top 10 predicted targets of the virus modulated host miRNAs and GpSGHV-encoded miRNAs

Target gene (VectoBase ID)	Gene name	miRNA	Molecular function/biological process
GPAI025158	Homeodomain interacting protein kinase (Hipk)	miR-184-3p	ATP binding, protein kinase activity. Plays a role in cell proliferation and development.
		Mir-GpSGHV_165479R	
		miR-6497	
GPAI030501	Ras-related protein Rab-27	miR-184-3p	GTPase activity Source, GTP binding. Involved in exocytosis and phagocytosis
GPAI014544	Fibrillin-1 (FBN1)	miR-184-3p	Calcium ion binding. Involved in cell communication
		miR-31-5p	
GPAI038987	Protein-serine/threonine kinase (STKP)	miR-6497	ATP binding, G-protein coupled receptor kinase activity. Involved in regulation of innate immune response and oogenesis.
		miR-9-3p	
		miR-263-5p	
		miR-184-3p	
		miR-276-5p	
GPAI034557	Apolipoprotein lipid transfer particle (Apoltp)	miR-184-3p	Lipid transporter activity, lipoprotein particle receptor binding. Provides the major yolk proteins during vitellogenesis
		miR-999-3p	
		miR-6497	
		miR-276-5p	
		miR-263-5p	
		miR-9-3p	
		Mir-GpSGHV_165479R	
		Mir-GpSGHV_165975R	
		Mir-GpSGHV_164791F	
GPAI025990	Transcription factor collier (TFCOE)	miR-6497	DNA binding, metal ion binding. Involved in development
		miR-9-3p	
		miR-263-5p	
		Mir-GpSGHV_165975R	
		Mir-GpSGHV_165479R	
GPAI001218	Twitchin	miR-6497	ATP binding, protein kinase activity. Mesoderm development
		miR-276-5p	
		miR-263-5p	

		miR-184-3p	
		miR-9-3p	
		miR-31-5p	
		miR-279-3p	
		Mir-GpSGHV_165975R	
		Mir-GpSGHV_164791F	
		Mir-GpSGHV_165482F	
		Mir-GpSGHV_165479R	
GPAI015640	Ras-related protein Ral-a (Rala)	miR-184-3p	GTPase activity, GTP binding. Innate immune response and signal transduction.
		miR-277-3p	
		miR-999-3p	
		miR-6497	
		miR-263-5p	
		miR-276-5p	
		miR-9-3p	
		Mir-GpSGHV_151557R	
		Mir-GpSGHV_165975R	
		Mir-GpSGHV_164791F	
		Mir-GpSGHV_165479R	
GPAI042543	Vitellogenin receptor (VtgR)	miR-6497	Calcium ion binding. Involved in uptake of vitellogen by endocytosis during oogenesis
		miR-184-3p	
		miR-276-5p	
		miR-263-5p	
GPAI007448	Centrosomal protein (CEP)	miR-184-3p	Centriole-centriole cohesion Source, centriole replication. Involved in spermatogenesis
		miR-263-5p	

The analysis showed that five of these immune genes (*Fibrillin-1, Rab27*, *VtgR*, *TFCOE* and *Apoltp*) were down-regulated, while only the *CEP* was up-regulated in symptomatic flies. There was no significant difference in expression levels of the genes encoding *Hipk*, *Twitchin*, *STKP* and *Ral-a* (**Figure 8**).

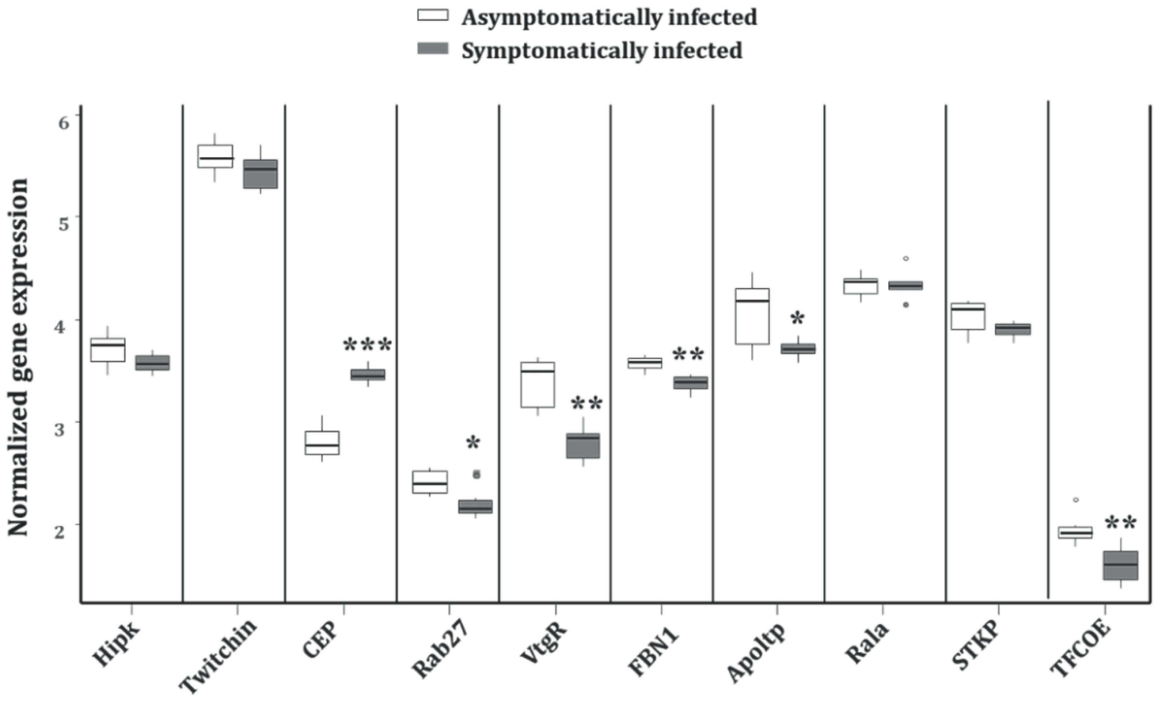


Figure 8: Reverse transcription qPCR (RT-qPCR) analysis of selected immune target genes in asymptomatically and symptomatically infected *G. pallidipes*. Error bars show the standard deviation from the replicates. Asterisks indicate the statistical significance: ***P < 0.001, **P < 0.01, *P < 0.05. Abbreviation: *CEP* (Centrosomal protein), *FBN1* (fibrillin-1), *Rala* (Ras-related protein Ral-a), *Hipk* (homeodomain interacting protein kinase), *Rab27* (Ras-related protein Rab-27A), *Apoltp* (Apolipoprotein lipid transfer particle), *TFCOE* (Transcription factor collier), *STKP* (Protein-serine/threonine kinase) and *VtgR* (Vitellogenin receptor).

Role of miR-184-3p during GpSGHV infection

To investigate the role of miR-184-3p, the most up-regulated miRNA in symptomatic flies (0.9 $\log_2$FC) that potentially targets most of the immune genes, during GpSGHV infection, miR-184-3p inhibitor and mimic sequences were designed and injected together with GpSGHV into *G. pallidipes*. A significant up-regulation and down-regulation of the expression of miR-184-3p was observed in the flies injected with the miR-184-3p mimic and inhibitor, respectively (**Figure 9A**). Injection of the miR-184-3p mimic showed a significant increase in GpSGHV *odv-e66* transcript levels, signalling up-regulated GpSGHV expression, while miR-184-3p inhibition had no impact on GpSGHV *odv-e66* transcript levels in both *G. pallidipes* females and males (**Figure 9B and 9C**).

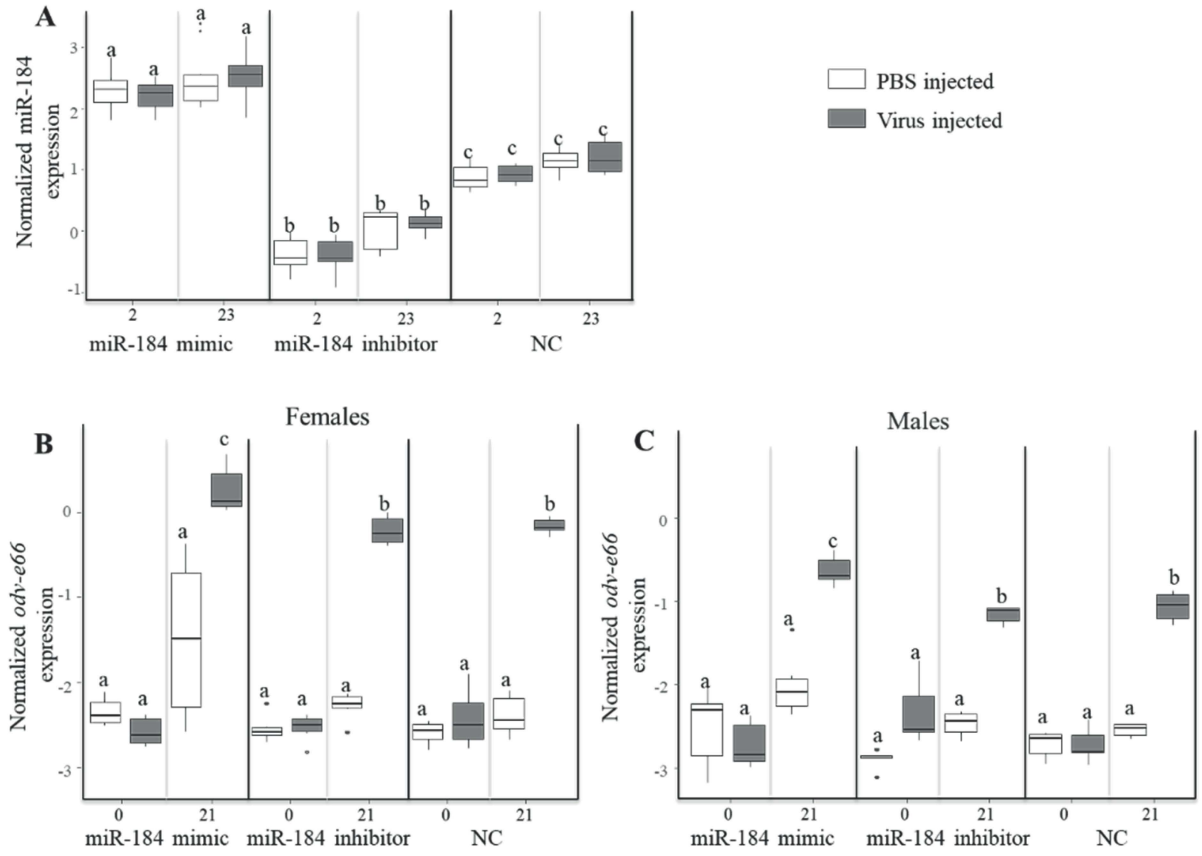


Figure 9: Impact of miR-184-3p inhibitor or mimic injection on GpSGHV infection. (A) Expression analysis of miR-184-3p 2 and 23 days post injection of mimic or inhibitor in *G. pallidipes*. (B) Expression levels of GpSGHV *odv-e66* gene following miR-184-3p inhibitor or mimic injection in *G. pallidipes* females and (C) males, compared to the RNase-free water injected flies (NC). Error bars show the standard deviation from the replicates. The expression results marked with the same lower-case letter do not differ at the 0.05 level.

Discussion

In insects, both host and virus-encoded miRNAs have been shown to regulate cellular responses, including immune responses, during virus infections (Asgari, 2014). Virus-encoded miRNAs may function just as cellular miRNAs and inhibit the translation of cellular mRNAs through direct interaction with its target mRNA, mediated by partial complementarity (Cullen, 2009; Kincaid and Sullivan, 2012). In addition, virus-encoded miRNAs can regulate virus encoded genes, especially genes that are involved in regulation of virus replication, and as such may control the latent and lytic infection stages by fully or partially aligning antisense to a target viral

mRNA (Asgari, 2015; He *et al.,* 2014; Kincaid and Sullivan, 2012). The molecular mechanisms of interactions between tsetse flies and GpSGHV are poorly understood, making it difficult to define the factors that determine the switch from asymptomatic to symptomatic infection. This study aimed to identify the host and virus-encoded miRNAs that are expressed during GpSGHV infection in *G. pallidipes* by the Illumina sRNA sequencing approach. The elucidation of the role of miRNAs in GpSGHV-tsetse interactions may open ways to manage virus replication or transmission in tsetse-mass rearing facilities, which would facilitate the implementation of the SIT as part of area-wide integrated pest management (AW-IPM) programs in the fight against tsetse and trypanosomosis.

Approximately 6.3 million sRNA reads were obtained from asymptomatically infected flies versus 4.8 million reads from symptomatically infected individuals, which may imply a depletion of sRNA synthesis and perhaps a reduced efficiency of the miRNA pathway during symptomatic infections. About 80% of the obtained total sRNA clean reads were mapped onto the genome of *G. pallidipes*. Most of the identified host miRNAs were also found to be conserved among other insects such as *Ae. aegypti* and *D. melanogaster* as predicted in the insect miRBase database. Among these miRNAs, miR-1-3p, miR-277-3p and miR-8-3p were highly expressed in both asymptomatic and symptomatic flies suggesting that they may have potential roles in regulating gene expression in *G. pallidipes* in general. For instance, miR-1-3p is known to regulate muscle cell differentiation and development in *Drosophila* (Kwon *et al.,* 2005). It was also found that 15 host miRNAs were differentially expressed in asymptomatically and symptomatically infected flies. According to the NGS data sets, miR-184-3p and miR-6497 were the most up- and down-regulated miRNAs, respectively, in symptomatic flies. MiR-184-3p has been reported to play a role during virus infections in mosquitoes (Maharaj *et al.,* 2015) where it was up-regulated in *Ae. albopictus* mosquitoes infected with the chikungunya virus (CHIKV) and dengue virus (DENV) (Liu *et al.,* 2015). MiR-184-3p has also been reported to be up-regulated in baculovirus-infected *Spodoptera frugiperda* cells (Mehrabadi *et al.,* 2013). It is therefore possible that miR-184-3p is an immune-associated miRNA in *G. pallidipes*; miR-184-3p has been reported to regulate the phagocytosis and phenoloxidase pathways in shrimps (Yang *et al.,* 2012). In

addition, miR-184-3p has also been implicated in multiple roles in *Drosophila* where it regulates oogenesis and early embryogenesis (Iovino *et al.,* 2009). Thus, the up-regulation of miR-184-3p in symptomatically infected *G. pallidipes* tsetse flies may hint at clues to the mechanism of GpSGHV-induced sterility.

The validation of the NGS differentially expressed miRNAs using RT-qPCR confirmed the expression of the up-regulated miRNAs (miR-184-3p, miR-276-5p, miR-263-5p, and miR-8-3p) in symptomatic flies. Some inconsistencies were noted in the down-regulated and equally expressed miRNAs according to our NGS data, which showed no significant change in their expression or up-regulated according to RT-qPCR respectively. Although different pools of asymptomatic and symptomatic flies were used for the analyses in the two approaches, these inconsistencies were unexpected as both approaches are considered highly sensitive and accurate. Nevertheless, such inconsistencies between RT-qPCR and NGS approaches for miRNA expression analyses are not uncommon as they have been observed in previous studies (Liu *et al.,* 2015; Saldaña *et al.,* 2017). Based on the *P* values ($P < 0.001$) of the RT-qPCR analysis, miR-263-5p was the most up-regulated miRNA whose modulation by virus infections has been reported in CHIKV and DENV-infected *Ae. albopictus* (Liu *et al.,* 2015; Maharaj *et al.,* 2015). In CHIKV-infected *Ae. aegypti,* miR-263-5p was not only highly expressed but also up-regulated (Caljon *et al.,* 2009). This finding implies that miR-263-5p (in addition to miR-184-3p) may play certain roles during GpSGHV infections in *G. pallidipes*.

In addition to the identification of host miRNAs, six GpSGHV-encoded miRNAs were identified that may potentially be involved in regulating GpSGHV infections. Four of these GpSGHV-encoded miRNAs were among the viral miRNA hairpins with high scores as predicted by VMir, a program that applies a low stringency prediction method by sliding a 500-nt window and utilizes RNAfold to analyse the RNA structure (Grundhoff, 2011). Additionally, these GpSGHV-encoded miRNAs have also been identified previously using sRNAloop (Garcia-Maruniak *et al.,* 2009), a program that predicts pre-miRNAs based on sequence structure and thermodynamic analyses (Grad *et al.,* 2003). These virus encoded-miRNAs may have potential roles during GpSGHV symptomatic infections, possibly to prolong the lifespan of the infected cells by targeting and suppressing pro-apoptotic host

genes as reported, for instance, for an Epstein Barr virus (EBV)-encoded miRNA (miR-BART5) (Marquitz *et al.,* 2011). The virus-encoded miRNAs can also assist in immune response evasion by negatively regulating early viral gene expression as in the case of Simian Vacuolating Virus40 (SV40)-encoded miRNA (Sullivan *et al.,* 2005). Mir-GpSGHV_170050R was identified in both asymptomatic and symptomatic *G. pallidipes* by the NGS approach. Although the direct viral target of this virus-encoded miRNA was a hypothetical GpSGHV protein (GpSGHV ORF132), the identification of a GpSGHV-encoded miRNA during asymptomatic infections may suggest a role in maintaining latent infection in *G. pallidipes.* Similar observations were made for *Heliothis zea* nudivirus-1 (HzNV-1), another large, rod-shaped, DNA insect virus that encodes miRNAs to promote latent infections by inhibiting viral gene expression (Wu *et al.,* 2011).

Prediction of putative target transcripts of the differentially expressed host and GpSGHV-encoded miRNAs may help in understanding the transcriptional regulation of genes depending on whether the GpSGHV infection becomes symptomatic or remains asymptomatic. The targets of the 14 differentially expressed miRNAs were predicted from the 3'-UTR's of the 6,071 available *G. pallidipes* transcripts in the VectorBase (Giraldo-Calderón *et al.,* 2015). The potential target transcripts were classified into different categories according to GO annotations. The current study focused on the immune related genes targeted by the regulated host miRNAs and the virus-encoded miRNAs for the GO enrichment (i.e. biological process, molecular function or cellular component) and pathway analyses. A single miRNA might regulate multiple target genes and even regulate the same target gene at multiple sites (Skalsky and Cullen, 2010). In the current study one of the most up-regulated host miRNA (i.e. miR-184-3p) according to the NGS data was found to target most of the immune genes predicted to be targeted by the complete set of up-regulated miRNAs. Most of these targeted immune genes appear to be involved in the Imd and Toll pathways, which are known to play a role in antiviral immunity in insects (Kingsolver *et al.,* 2013), for example against arbovirus infections (Avadhanula *et al.,* 2009; Xi *et al.,* 2008). Some of the targeted immune genes include the *FBN1* (glycoprotein involved in cell communication), *Rab27* and *Apoltp* (involved in positive regulation of lipid transport) and these were indeed down-

regulated in the symptomatic flies. In baculovirus-infected *S. frugiperda* cells, where miR-184-3p was up-regulated, target prediction and transcript level analysis showed that this miRNA may either positively or negatively regulate particular target gene transcripts (Mehrabadi *et al.,* 2013). Notably, the *Apoltp* and *VtgR* were predicted to be involved in vitellogenesis, the main process in oogenesis and egg production. The down-regulation of these genes in symptomatic flies may explain the ovarian abnormalities and reduced reproductive fitness observed in symptomatically infected female tsetse flies. Similar observations have been reported in the housefly, *Musca domestica*, whereby MdSGHV infections which causes similar SGH syndrome in their host, were found to suppress vitellogenesis by blocking the transcription of hexamerin and yolk proteins and cause shut down of oogenesis and hence reduce reproduction (Kariithi *et al.,* 2017b).

Multiple miRNAs might co-regulate one target gene at the same time (Skalsky and Cullen, 2010). For instance, in this study the *CEP* gene, which was up-regulated in symptomatic *G. pallidipes*, contained target sites for two up-regulated miRNAs (miR-184-3p and miR-263-5p). This also indicates that although most reports show that miRNA-target interaction lead to negative regulation of the target gene (Asgari, 2011), a positive regulatory effect may also occur by promoting transcript stabilization or translation as previously reported (Conrad *et al.,* 2013; Hussain *et al.,* 2011). Approximately, 30 % of the immune genes targeted by the up-regulated host miRNAs were also found to be targets of the GpSGHV-encoded miRNAs, with most of these genes involved in Toll pathway signalling. The Toll pathway is known to direct antiviral defence in DENV-infected *Ae. albopictus* following down-regulation of the host miRNA mir-375 (Liu *et al.,* 2015). The GpSGHV-encoded miRNAs targeted transcripts of *Rab27*, *FBN1* and the *Apoltp* genes, transcripts that can also be targeted by host miR-184-3p. In addition, these viral miRNAs also specifically targeted the thyroid receptor–interacting protein (*TRIPB*), serine protease 1 *(Ser1)* and midgut trypsin *(Mid-Try)*. Trypsin is involved in serine endopeptidase activity and has been reported to cleave the well conserved baculovirus P74, a viral attachment protein. This p74 cleavage is crucial for infection and necessary for the baculovirus to establish a primary infection in midgut cells (Slack *et al.,* 2008). It should be stressed here that GpSGHV encodes a homolog of

the baculovirus P74 protein (Abd-Alla *et al.,* 2008). Why the virus up-regulates trypsin in already infected insects is not clear.

In this study, attempts to artificially up-regulate miR-184-3p by injecting its mimic led to increased expression levels of the GpSGHV *odv-e66* gene. However, the corresponding miR-184-3p inhibitor did not cause any significant difference in GpSGHV *odv-e66* expression. Since miR-184-3p may regulate transcripts of several host genes, how this affects GpSGHV infection requires further studies. Notably, miR-184-3p has been reported to be induced by Interleukin-22 (IL-22), an inflammatory cytokine, by down-regulating the expression of Argonaute-2 (AGO-2), a key protein of the RNAi pathway (Roberts *et al.,* 2013). As RNAi is an important immune defence pathway against virus infections in most insects (Van Rij, 2008), the up-regulation of miR-184-3p in symptomatically GpSGHV infected flies could modulate *AGO-2* expression and thereby regulate virus replication (See **Chapter 5** of this thesis). This agrees with observations for invertebrate iridovirus (IIV-6) in *Drosophila* (Bronkhorst *et al.,* 2012) and the findings in Chapter 5 where *AGO-2* was found to be down-regulated in symptomatically infected flies compared to asymptomatically infected flies.

Conclusions

The study described in this chapter has identified host and viral-encoded miRNAs and evaluated their expression profiles during GpSGHV asymptomatic and symptomatic infections in *G. pallidipes*. Fifteen differentially expressed host miRNAs were identified and their target predictions suggested that miR-184-3p, the most up-regulated miRNA in symptomatic flies, might be involved in regulating immune responses and oogenesis and hence the reproductive fitness of the flies, since it targeted mostly immune related and vitellogenesis genes. This study further presents the first evidence that GpSGHV alters the host miRNA profile in *Glossina,* a finding that provides a baseline for further investigations to understand the GpSGHV-*Glossina* interactions. Finally, the data from the current study provides insights into the interaction between GpSGHV and *G. pallidipes* miRNAs, and provides potential avenues to further study the mechanisms of immune response during GpSGHV infections in tsetse fly. This information may provide strategies to

control GpSGHV infections in tsetse mass rearing facilities, a prerequisite to SIT implementation, by utilization of particular miRNAs, especially those implicated in anti-viral responses, or by inhibition of pro-viral miRNAs by miR inhibitors. Alternatively, these miRNAs might be overexpressed in symbionts (*Sodalis glossinidius*) via paratransgenesis (De Vooght *et al.,* 2014; 2012).

Acknowledgements

The authors acknowledge Hasim Mohammed, Carmen Marin, Edgardo Lapiz and Henry Adun of the IPCL, for assisting with the fly rearing. This study was sponsored by the Joint FAO/IAEA Insect Pest Control Laboratory Seibersdorf, Austria (Coordinate Research Project No. D42015) and Wageningen University and Research, The Netherlands.

Chapter 7

General discussion

This PhD thesis investigated some of the key factors that may influence the outcome of *Glossina pallidipes* salivary gland hypertrophy (GpSGHV) infections in tsetse fly. In some tsetse species such as *G. pallidipes* the viral infection causes outbreaks of salivary gland hypertrophy (SGH), resulting to reproductive disfunctions and colony collapses. The research described in this thesis aimed to provide answers to several key issues related to GpSGHV infections. Firstly, a comprehensive review was presented on the coevolution of the hytrosaviruses (SGHVs) and their hosts' defence response, which showed that this coevolution is potentially influenced by the ecologies and life histories of their hosts (**Chapter 2**). This virus-host interplay may have impact on the outcomes of SGHVs infection in their hosts. Secondly, a multi-marker tool was developed to correctly identify and distinguish tsetse species in wild populations as a complement to the more conventional morphological species identification (**Chapter 3**). Thirdly, in addition to investigating the prevalence, diversity and ecogeographical spread of GpSGHV in wild and lab-bred tsetse species, the thesis also correlated the different viral strains to species and sub-species of tsetse fly (**Chapter 4**). Fourthly, the role of RNAi machinery during GpSGHV infections in *G. pallidipes* was investigated as part of the host's defence against the virus (**Chapter 5**). Lastly, the thesis provided the first insights into the involvement of miRNAs in tsetse-SGHV interactions in *G. pallidipes* (**Chapter 6**). This last chapter of the thesis is a synopsis on the findings of these topics and the feasibilities of their applications in the management of viral infections in tsetse and other insect mass production. The chapter also highlights some of the remaining gaps in the knowledge of SGHV-host interactions and proposes a way forward to address some of the outstanding questions.

The evolution and genetics of hytrosavirus host shifts

GpSGHV infections hinder production of sufficient numbers of tsetse flies for SIT campaigns in Africa, as well as biological materials (e.g. adults, larvae, pupae) for studies on the biology and genetics of tsetse and their inhabiting microbiota (pathogens and symbionts) (Abd-Alla *et al.,* 2011; 2010b). Virus-host interaction is subjected to a continuous coevolutionary process that involves both host defence system and viral escape mechanisms (Lobo *et al.,* 2009; Taylor *et al.,* 2013). The genetic differences between the Ethiopian and Ugandan GpSGHV strains (e.g. 24

additional and 11 missing ORFs in GpSGHV-Eth compared to the GpSGHV-Uga genomes) may explain the differences in the virus-induced pathogenesis in different *G. pallidipes* populations (Abd-Alla *et al.,* 2008; 2016). Different host and evolutionary factors (e.g. genetic variation, infection history, microbiota and other abiotic components), have been described that may affect a pathogen's ability to infect several species including *Drosophila* and mosquito (Palmer *et al.,* 2018). In addition, the susceptibility of a host to pathogen infection may vary depending on how closely related a novel host species is to the pathogen's natural host, since closely related species may offer a similar environment to a pathogen (Longdon *et al.,* 2014).

In case of tsetse flies, the integration of the nuclear markers (internal transcribed spacers 1; ITS1, and microsatellites), *Wolbachia* diagnosis as well as the place of origin of tsetse species, allowed us to distinctly identify the species. In addition, the mitochondrial gene sequencing was not only able to cluster the tsetse species into the three *Glossina* taxonomic groups (Palpalis, Morsitans and Fusca groups), but they also distinguished populations/haplotypes of the same tsetse species. The phylogenetic relatedness of tsetse species enabled the prediction of the ancestral origin(s)/evolution of GpSGHV and its impact on host shift. The highest GpSGHV prevalence was found in the species belonging to the Morsitans group, with *G. pallidipes* presenting the highest prevalence and multiple virus haplotypes. Unlike other tsetse species belonging to Morsitans group, *G. swynnertoni* was the only species infected with a distinct virus haplotype. The representative samples of species belonging to the Fusca and Palpalis groups were all infected with a single virus haplotype which is similar to the GpSGHV-Uga strain and was found infecting six of the seven analysed tsetse species.

During evolution, viruses can either evolve with their host that diverge from a certain population into a new species or they may jump into a new host species that is related to their previous host (Geoghegan *et al.,* 2017). Several studies have shown that the latter pathway is more common and the cause of majority of emerging infections (Dill *et al.,* 2016; McGeoch and Gatherer, 2005), and this is probably the case for the GpSGHV-tsetse evolution. For example, the distinct specialization of baculovirus lineages to particular insect orders and even genera suggest ancient coevolutionary interactions with their insect hosts (Herniou *et al.,* 2004). In addition,

pathogens need to adapt to a new host to be successful in infecting their hosts, for example by causing mutations in their genomes that are essential for their survival. However, most of the mutations involved in virus-adaptations are lethal and therefore the virus has to optimise the mutation rate in the novel host, especially those involved in host binding/fusion of the virus to the host (Lalić *et al.,* 2011; Longdon *et al.,* 2014; Loverdo and Lloyd-Smith, 2013). In *G. pallidipes,* multiple GpSGHV haplotypes were found unlike in other tsetse species and it is in this tsetse species where SGH outbreaks occur in mass rearing facilities although the other species are also reared (Abd-Alla *et al.,* 2010b). The finding that only one GpSGHV haplotype infected the Palpalis and Fusca groups, which are more closely related to each other than to the Morsitans group, may imply that a species belonging to these two groups is probably the original GpSGHV host(s). Although the first description of SGH symptoms were discovered in *G. pallidipes* (Morsitans group) (Burtt, 1945; Whitnall, 1934), it is possible that the virus had just shifted to this new host. To enhance its fitness in a new host, the virus might have acquired novel and specific mutations for its survival in *G. pallidipes* (**Figure 1**).

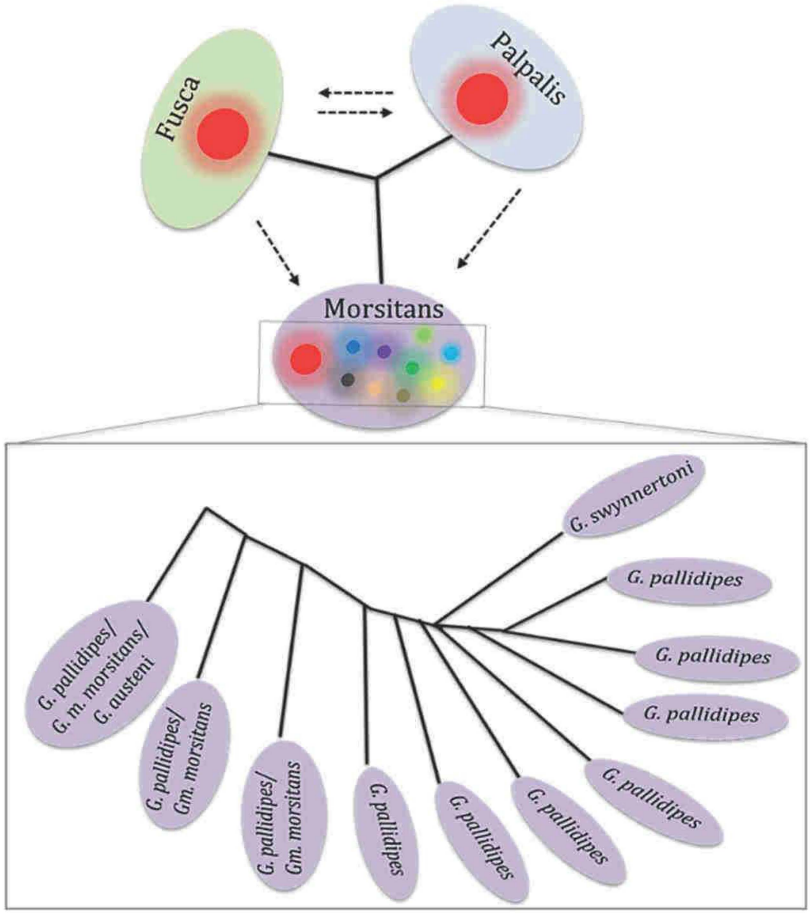


Figure 1: Hypothetical illustration of GpSGHV evolution and host-shift in tsetse species: The ancestral GpSGHV virus haplotype (ball in red) infected tsetse species belonging to the Fuscipes or Palpalis group. Later on, the virus host shift to the Morsitans group could have led to the evolution of novel virus haplotypes (illustrated in the figure by the change of the ball from red to green, yellow or purple. While most of the new virus haplotypes emerged in *G. pallidipes,* similar virus strains emerged from more than two species (see the figure inset).

How the virus was able to spread within different tsetse species in the wild is a puzzle to be solved. However, tsetse fly is a solitary insect and its biology and feeding behaviour in the animal host and habitat may contribute to the horizontal transmission of this virus (Caljon *et al.,* 2010; Spath, 2000; Van Den Abbeele *et al.,* 2010).

The susceptibility of the new host to the pathogen and the subsequent transmission of the pathogen are likely to be determined by the genetics, behaviour and the ecological factors of the host and/or virus. Usually certain host factors such as the nutritional composition, the history to pathogen exposure and the immunocompetence contribute to the host susceptibility to infections (Palmer *et al.,* 2018). In a comparative study of GpSGHV-infected *G. m. morsitans* and symptomatically infected *G. pallidipes* flies, *G. m. morsitans* expressed more antiviral proteins than *G. pallidipes* (Kariithi *et al.,* 2016), which may imply that the genetic architecture of a host may influence the host immune response. Furthermore, there are indications that during GpSGHV-host coevolution, GpSGHV may have recruited some cellular genes from its ancestral host into its genome, which may assist in evading the host defence system. This scenario has indeed been described for other viruses. For instance, the coevolution of herpesvirus and their hosts has created an equilibrium with their immune response, thus allowing persistent viral infections without significant cellular damages during the host lifetime and only become reactivated and pathogenic under certain conditions (Casa-Esper *et al.,* 2012; White *et al.,* 2012). During this coevolution, herpesviruses are thought to have recruited certain host genes and adapted them to assist in evading the immune response (Casa-Esper *et al.,* 2012). Therefore, the host-to-virus horizontal gene transfer (HGT) may contribute to the coevolution between viruses and their hosts, as well as a provide viral mechanism to regulate not only the host defence response, but also provide strategies for the virus proliferation and spread to new related hosts.

The interplay between SGHV infection and antiviral immune pathways

RNAi machinery and SGHV infection

Several conserved signalling pathways and cellular innate responses are known to mediate the antiviral immunity in dipteran insects (Kingsolver *et al.,* 2013; Merkling

and van Rij, 2013). RNA interference (RNAi) is considered the primary antiviral immune pathway in insects against several RNA and DNA viruses (Bronkhorst and van Rij, 2014; Olson and Blair, 2015). However, viruses have evolved strategies to interfere with the RNAi machinery by; (i) use of viral suppressors of RNAi (VSRs), (ii) accumulation of large amounts of specific viral RNA and small RNAs to overflow the RNAi machinery, (iii) regulation of host microRNAs (miRNAs) expression and (iv) by encoding viral miRNAs (Swevers *et al.,* 2013). Two RNAi pathways, that are regulated by different sets of genes, are known to provide antiviral immune response in many insects; the short interfering RNA (siRNA) and the miRNA pathways (van Rij, 2008). The ribonuclease III enzyme, Dicer-1 (DCR1), processes cellular or viral pre-miRNAs (pre-processed by RNase III enzyme Drosha in the nucleus) into mature miRNAs (miRNA pathway) while DCR2 processes exogenous dsRNAs into siRNA (siRNA pathway) (Lee *et al.,* 2004; Lucas and Raikhel, 2013). The miRNAs and siRNAs are then loaded onto the Argonaute (AGO) proteins 1 and 2 respectively, to mediate RNAi by translation repression (AGO1), or cleavage/degradation of target RNAs (AGO2) (Okamura *et al.,* 2004).

In *Drosophila*, mutations of the genes involved in the siRNA pathway made the flies more susceptible to viral infection (Galiana-Arnoux *et al.,* 2006; Wang *et al.,* 2006; Zambon *et al.,* 2006). In *G. pallidipes*, genes regulating the siRNA (*AGO2* and *DCR2*) and miRNA (*AGO1* and *DCR2*) pathways were upregulated following GpSGHV injection (asymptomatic infections), but the expression of some of these key genes (*AGO1*, *AGO2* and *Drosha*) were downregulated in symptomatic flies (**Chapter 5**). Notably, the *G. pallidipes* genome lacks a *DCR1* ortholog, and therefore *DCR2* gene, whose transcripts were not modulated in symptomatic flies, may function in processing both miRNAs and siRNAs (**Figure 2**). However, knockdown of the siRNA pathway key gene (*AGO2*) in *G. pallidipes* did not increase the susceptibility of the flies to virus infection, probably because other innate immune pathways (discussed later) may also participate in elimination of the virus.

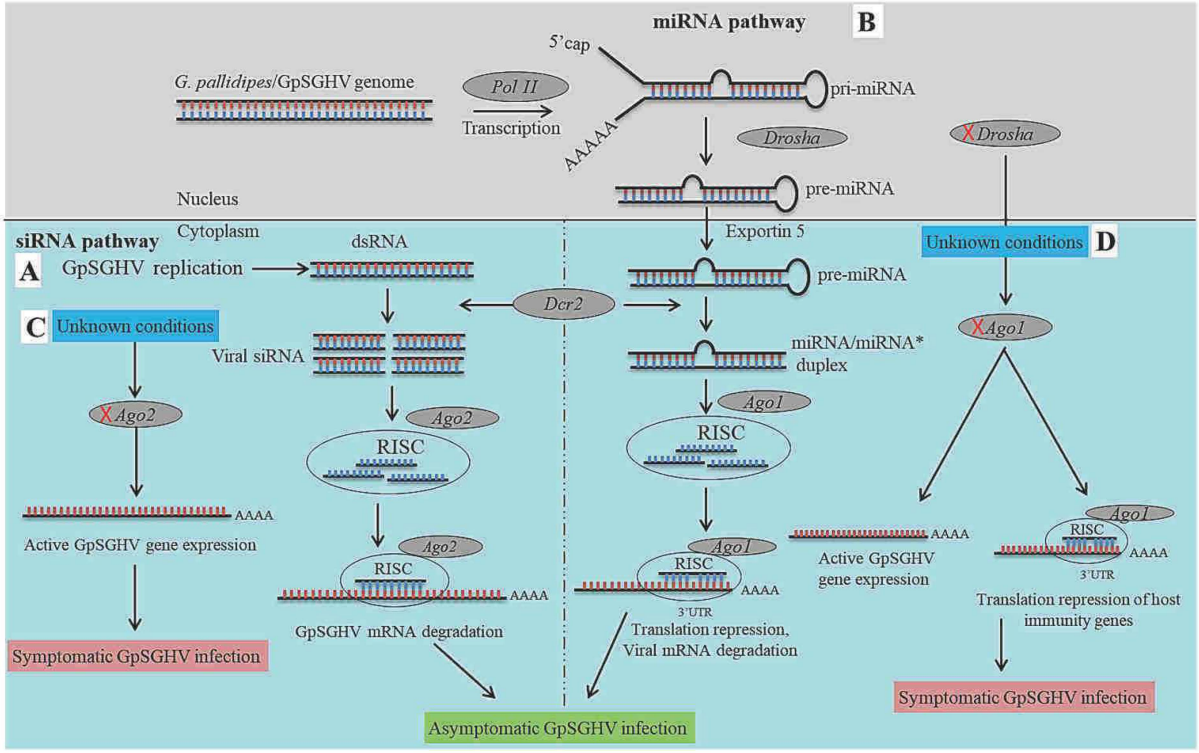


Figure 2: Illustration of the mechanism of RNA interference in GpSGHV infections: During asymptomatic infections in *G. pallidipes,* RNAi is activated, and the key genes involved are highly expressed. In (**A**) GpSGHV produce double-stranded RNAs (dsRNAs) following viral replication, which are processed by the RNAi machinery into siRNAs and activate the RNA-induced silencing complex (RISC) to degrade specific GpSGHV mRNAs. While in (**B**) Host (*G. pallidipes*) and GpSGHV genes are processed in the nucleus into primary pri-miRNAs that are processed further into pre-miRNAs by *Drosha* before they are exported into the cytoplasm. The pre-miRNAs are then cleaved by *DCR-2* into microRNAs, which activate the RISC and suppress the targeted GpSGHV gene expression. Under certain conditions (e.g. environmental triggers), symptomatic infections may be triggered due to (**C**) the decreased *AGO-2* gene expression, that lead to high viral gene expression or viral replication, or (**D**) the decreased *AGO-1* and *Drosha* gene expression levels that regulate miRNA expression profile and utilize the miRNAs that target and suppress expression of host (immunity) genes and thereby increase viral replication. X = indicates the genes with low expression levels in symptomatic infections.

In *Aedes* mosquitoes, a separate RNAi pathway, the Piwi-interacting RNA (piRNA) which is normally involved in epigenetic modifications and regulation of activities of transposons in *Drosophila* (Luo and Lu, 2017), was found to have an antiviral role against arboviruses (Schnettler *et al.,* 2013; Varjak *et al.,* 2017). However, in the case of tsetse flies, the piRNA pathway may not have a functional role during

GpSGHV infection, since the expression level of AGO3, which processes the piRNAs was not modulated in either virus injected or symptomatically infected *G. pallidipes* flies.

To favour their own replication some viruses, manipulate the host's transcriptome via specific degradation or upregulation of specific host miRNAs to escape the hosts' antiviral response and to promote infection (Liu *et al.,* 2015; Maharaj *et al.,* 2015; Saldaña *et al.,* 2017). Overall, several studies have demonstrated that the miRNAs can either positively or negatively modulate the host's immune responses to the advantage of the viral replication. (Lucas and Raikhel, 2013; Scheel *et al.,* 2016). In symptomatically infected *G. pallidipes*, GpSGHV alters the host miRNA profile (**Chapter 6**), which strongly suggests a possible functional importance of host miRNAs during virus infection. Of interest is miR-184-3p, the most up-regulated miRNA in symptomatic flies, which was predicted to target genes involved in immunity and vitellogenesis. This is related to the fact that symptomatic female flies display ovarian abnormalities and reduced reproductive fitness (Abd-Alla *et al.,* 2010b). Thus, the exact role of this specific host miRNA among others that are regulated by GpSGHV infections requires further investigation, in particular since miR-184-3p mimic-injection into *G. pallidipes* led to increased virus replication (**Chapter 6**).

Nuclear replicating DNA viruses such as baculoviruses, ascoviruses and nimaviruses are known to encode miRNAs that act by suppressing both viral and cellular mRNAs to interfere with biological processes such as apoptosis, to evade the host immune response (Grundhoff and Sullivan, 2011; He *et al.,* 2014; Singh *et al.,* 2010a). This action would prolong the life of the infected cells to maximize virus replication, and thereby promotes the establishment of persistent infections (Hussain *et al.,* 2008). One of the key finding in this thesis is the identification of GpSGHV-encoded miRNAs that may contribute in the immune evasion and perhaps establishment of asymptomatic infections in *G. pallidipes* (**Chapter 6**). This would be possible by prolonging the longevity of the infected cells if they target pro-apoptotic host genes as reported for some viruses such as in the Epstein Barr virus (EBV)-encoded miRNA (miR-BART5) (Marquitz *et al.,* 2011), or by negatively regulating early viral gene expression as in the case of Simian Vacuolating Virus 40 (SV40)-encoded

miRNA (Sullivan *et al.,* 2005). The utilization of viral encoded miRNAs to promote persistent infections has been reported in several viruses such as *Heliothis zea* nudivrius-1 (HzNV-1), a large DNA insect virus that encodes miRNAs and promotes latent infection by inhibiting viral gene expression (Wu *et al.,* 2011). Consequently, the role of the GpSGHV-encoded miRNAs in maintaining latent infection and perhaps altering the balance between the asymptomatic and symptomatic infections in *G. pallidipes* requires further investigation.

Other innate immune pathways

While a clear role of RNAi in control of virus infections has been demonstrated in insects, other innate immune pathways such as the Nuclear Factor kB (NF-kB) (i.e., Toll and Imd) and the Jak-STAT pathways, which primarily function in anti-fungal and anti-bacterial defence, may be to some extend have an antiviral function (Costa *et al.,* 2009; Dostert *et al.,* 2005). This has been supported by several studies where knockdown of components of each of these pathways resulted in high viral loads (Kemp *et al.,* 2013; Zambon *et al.,* 2005), although this may depend on the properties of the viruses and the host species. Mostly, viruses trigger a transcriptional response in infected flies depending on the recognition of virus-associated molecular patterns or host manipulation (Lamiable and Imler, 2014; Xu *et al.,* 2012). For example, the NF-kB associated antimicrobial peptides (AMPs) effectors were upregulated after viral challenge in Dengue virus infected *Aedes aegypti* and in alphavirus infected *Drosophila* (Huang *et al.,* 2013). Similar findings were found in *G. pallidipes* flies where the AMPs (attacin, cecropin and defensin) were upregulated in GpSGHV injected compared to the non-infected flies (Unpublished data). The antiviral role of Jak-STAT pathway has been reported in mosquitoes in that this pathway can be activated by *Vago,* an antiviral gene which is normally induced and acts downstream of dsRNA recognition by DCR2 (Deddouche *et al.,* 2008). Increased levels of *Vago* were also found in *Drosophila* following Sindbis virus (SINV) infection (Huang *et al.,* 2013). By using VectorBase (http://www.vectorbase.org), an orthologue of *Drosophila* Vago (*DM*Vago) was found in the *G. pallidipes* genome, a gene which requires further investigation.

Other pathways that contribute to the antiviral defence in dipterans are apoptosis and phagocytosis (Everett and McFadden, 1999). For instance, the apoptosis response in

Drosophila, is induced by viral infection with the aim of reducing the duration of a virus to access the host factors crucial for its replication. In the case of tsetse fly the critical genes, the initiator and effector caspases (apoptosis) and thioester-containing proteins (TEPs) (phagocytosis) have been identified and they may be activated during GpSGHV infections.

The influence of microbial communities on tsetse antiviral immunity

In addition to host genetic makeup, a fly's microbial community also contributes to the susceptibility of insects to virus infection (Hegde *et al.,* 2015; Jupatanakul *et al.,* 2014; Robinson and Pfeiffer, 2014). Tsetse flies harbour three specific bacterial symbionts; obligate *Wigglesworthia glossinidia,* and the facultative *Sodalis glossinidius* and *Wolbachia pipientis,* that are known to directly/indirectly regulate the nutrition, reproduction and immunity of the flies (Moreira *et al.,* 2009; Snyder *et al.,* 2010; Weiss *et al.,* 2012). For instance, *Wigglesworthia* is known to upregulate the host's defence system following trypanosome infections, while *Sodalis* has been reported to influence the host's ability to establish trypanosome infections (Soumana *et al.,* 2014a; Weiss *et al.,* 2012). In addition, antibiotic-mediated removal of both *Wigglesworthia* and *Sodalis* from *G. pallidipes* led to reduced GpSGHV replication and occurrence of SGH symptoms in the progeny (Boucias *et al.,* 2013). Of interest is the *Wolbachia* which is more widespread among insects and is known to induce several phenotypes such as male-killing, male feminization and cytoplasmic incompatibility (CI) (Werren *et al.,* 2008). The CI phenomenon that leads to embryogenesis arrest, when an uninfected female fly mates with a *Wolbachia*-infected male has already been demonstrated in tsetse (Alam *et al.,* 2011; Doudoumis *et al.,* 2012). Increased resistance to virus infection has been reported in *Wolbachia*-infected insects, although the exact mechanism behind the induced antiviral protection remains a puzzle (Hedges *et al.,* 2008; Johnson, 2015; Osborne *et al.,* 2012). Currently, the possible explanations are, the competition between virus and *Wolbachia* for resources, or *Wolbachia*-mediated immune priming (Terradas and McGraw, 2017). *Wolbachia* immune priming is likely since *Wolbachia* infection is reported to be associated with upregulation of antiviral genes such as AGO2 and AMPs, although these effects maybe species specific (Rancès *et al.,* 2012; Terradas and McGraw, 2017). Alternatively, the *Wolbachia* and virus competition for

resources may play a significant role, since alteration of diet in *Wolbachia*-infected *Drosophila* interfered with DCV infection (Caragata *et al.,* 2013).

The absence or low levels of *Wolbachia* in *G. pallidipes* may be linked to the frequent occurrence of SGH in this species as compared to other species such as *G. m. morsitans* that present a high *Wolbachia* prevalence and levels and without any documented evidence for SGH outbreaks (Abd-Alla *et al.,* 2010b; Doudoumis *et al.,* 2012). Although the *Wolbachia* immune priming is possible in *Wolbachia*-infected tsetse flies against GpSGHV infections, nutrients competition between virus and *Wolbachi*a (and other tsetse symbionts) might be likely especially in mass rearing facilities that utilizes a routine blood feeding system compared to diet of tsetse flies in the field (Feldmann, 1994). Notably, *Wolbachia* is undetectable so far in laboratory colonised *G. pallidipes* where symptomatic GpSGHV infections frequently occur compared to the wild populations where low levels of *Wolbachia* are present (Doudoumis *et al.,* 2012). In addition, since *Wolbachia* is maternally transmitted, there are high chances of driving the infection in some populations and not others and therefore the reason for the selected virus resistance or susceptibility in some fly populations (Martinez *et al.,* 2015). If this were the case, then the *Wolbachia*-infected tsetse populations would remain resistant to GpSGHV or asymptomatically infected while the *Wolbachia* free species such as *G. pallidipes* remain susceptible. *Wolbachia* can as well offer protection against virus infections, by modulating miRNA expression profile in insects and thereby influence the dynamics of virus infections (Hussain *et al.,* 2011). For instance, *Wolbachia*-infected mosquito *Ae. Aegypti*, upregulates the expression of mir-2940 that targets and suppresses the host methyltransferase gene and hence block DENV-2 replication (Zhang *et al.,* 2013a). Therefore, the differences in *Wolbachia* prevalence in tsetse species may alter the miRNAs expression and thereby influence the occurrence and outcome of GpSGHV infections.

In addition to the symbionts, the gut microbiota community may as well influence the outcome of virus infections either positively or negatively as reported in other insects (Hegde *et al.,* 2015; Jupatanakul *et al.,* 2014; Robinson and Pfeiffer, 2014). For example, the presence of microbiota in *Spodoptera exigua* led to increase in baculovirus infection while in *D. melanogaster* antibiotic treatment led to increased

Drosophila C virus (DCV) replication (Jakubowska *et al.,* 2013; Sansone *et al.,* 2015). Microbiota may offer antiviral protection by influencing the gut environmental factors that have an antiviral role (via the JAK-STAT and Imd pathways) such as the AMPs, which were in fact upregulated in presence of microbiota in mosquitoes (Buchon *et al.,* 2009). It is also likely that the species mixture and level of gut microbiota may define the fly's immune response. For example, an isolate of *Chromabacterium* was found to inhibit DENV replication in vertebrate cells, while a *Talaromyces* isolate increased DENV infection in mosquitoes (Angleró-Rodríguez *et al.,* 2017; Ramirez *et al.,* 2014).

Wild populations of tsetse flies have been reported to have a low diversity of gut microbiota as compared to other insects (Aksoy *et al.,* 2014). Several bacteria species have been isolated in laboratory colonised tsetse species, including the *Serratia glossinae* from *G. p. gambiensis* (Geiger *et al.,* 2010). From wild tsetse populations bacterial species belonging to genera such as *Enterobacter*, *Enterococcus*, and *Acinetobacter*, *Providencia*, *Sphingobacterium*, *Chryseobacterium*, *Lactococcus*, *Staphylococcus*, and *Pseudomonas* among others have been identified (Geiger *et al.,* 2009; 2011; Lindh and Lehane, 2011). However, the origin of microbiota with such diversity in tsetse flies is unclear compared to other hematophagous insects such as mosquitoes, although it is possible that tsetse flies have other sources of food other than blood in the field (Solano *et al.,* 2015). Alternatively, tsetse flies may ingest bacteria found on the skin surface of the host animal when biting to feed blood (Poinar *et al.,* 1979). The question is whether these diverse microbial communities influence tsetse fly immunity and therefore affect GpSGHV infections outcome as well as the fly's vector competence to trypanosome transmission in wild tsetse populations.

Novel control strategies for GpSGHV infections in *G. pallidipes* colonies

Application of antiviral drugs (e.g. valacyclovir) and the clean feeding system (sanitary measure), either singly or in combination, have been successfully used in the management of GpSGHV in tsetse mass reading facilities (Abd-Alla *et al.,* 2012; 2014). However, resistance development of viruses against antiviral drugs is a concern, as reported in the case of herpesviruses (Griffiths, 2011). This is especially relevant in the case of *G. pallidipes*, which represents multiple GpSGHV strains.

One of the key findings in this thesis is that the RNAi machinery (siRNA and miRNA pathways), which is the primary antiviral defence, is compromised in symptomatically infected *G. pallidipes*. Therefore, one could postulate that the RNAi mechanism can be exploited to treat viral infections in tsetse flies and/or prevent SGH outbreaks. The remaining section of this chapter discusses and recommends several potential strategies that could be developed on the basis of the data generated in this thesis to manage GpSGHV infections.

Initial screening strategies to avoid introduction of viruses

Field collected material for the establishment of new insect laboratory colonies can be highly diverse in terms of health, immune status and pathogens, which makes the initial screening critically important. Given the existence of covert GpSGHV infections in wild tsetse populations and the mixed virus transmission modes (vertical and horizontal), the risk of introducing the virus to the new colonies is high. It is possible that the virus may preferentially adopt the horizontal transmission once new flies are introduced into the colonies, unlike in the field where the gregarious nature of the insect and environmental conditions are not ideal for horizontal transmission. In laboratory colonies, stressors such as overcrowding may eventually lead to lethal viral outbreaks and evolution of more virulent strains. Therefore, the first recommendation into management of GpSGHV infections in mass production facilities is to prevent introduction of viruses during initial establishment of new tsetse colonies. This can be done using GpSGHV variable genes applied in this thesis to enable identification of different GpSGHV strains and select the appropriate antiviral control strategies (**Chapter 4**).

Control virus spread within the facilities

Even with the strict measures to prevent pathogen (viruses) introduction, some facilities may still encounter disease outbreaks. Therefore, in already established tsetse rearing facilities, in addition to the application of the existing virus control strategies, below are the recommended approaches that should be considered as alternatives for virus management.

a) Silencing of viral infection by RNAi

The existence of the RNAi machinery in the tsetse fly makes it a potential tool to target and induce silencing of genes of interest. The success of RNAi and its long-term expression depends on effective siRNA design and method of delivery into the target cells (Swevers *et al.,* 2013). RNAi has been applied in control/management of virus infections by targeting viral genes or alternatively, cellular genes required for virus replication (Dykxhoorn and Lieberman, 2006). Control of virus infection using RNAi has been demonstrated in several insects including, management of Israeli Acute Paralysis Virus (IAPV) in honeybees, as well as in mosquitoes and *Drosophila* against several viruses (Burand and Hunter, 2013). Although the strategy can be challenging or be impossible to apply in wild insect populations, it can be promising for laboratory colonised insects. For example, the application of antiviral dsRNA in *An. gambiae* against DENV-2 was a very promising approach in 13 generations, although its protective effect was reduced in later generations (Idrees and Ashfaq, 2013). Although the RNAi silencing applied in this thesis was by injection of dsRNA, feeding of dsRNA specific for the gene of interest to tsetse flies has been demonstrated successful (Walshe *et al.,* 2009). The RNAi based approach for management of GpSGHV infections can be promising in tsetse mass rearing facilities where dsRNAs specific for viral genes essential for viral entry and replication can be targeted (**Figure 3A**). Some of the potential candidate gene targets for silencing to prevent viral entry would be the *per os* infectivity factors (PIFs), while to suppress virus transcription, replication and translation, genes such as *thymidylate synthase* (ORF35), *dihydrofolate reductase thymidylate synthase* (ORF36), *HSP90-like ATPase* (ORF39), *p53-trancription factor-like* (ORF62) and ABC-ATPase (ORF64) can be targeted (Kariithi *et al.,* 2011). To enhance GpSGHV suppression, targeting multiple viral genes, is recommended to reduce the chances of viral escape. In the case of baculovirus, this approach has been successful in controlling / mitigating the replication of *Bombyx mori* NPV in silkworm by downregulating immediate early gene 1 (ie-1) (Zhang *et al.,* 2013b).

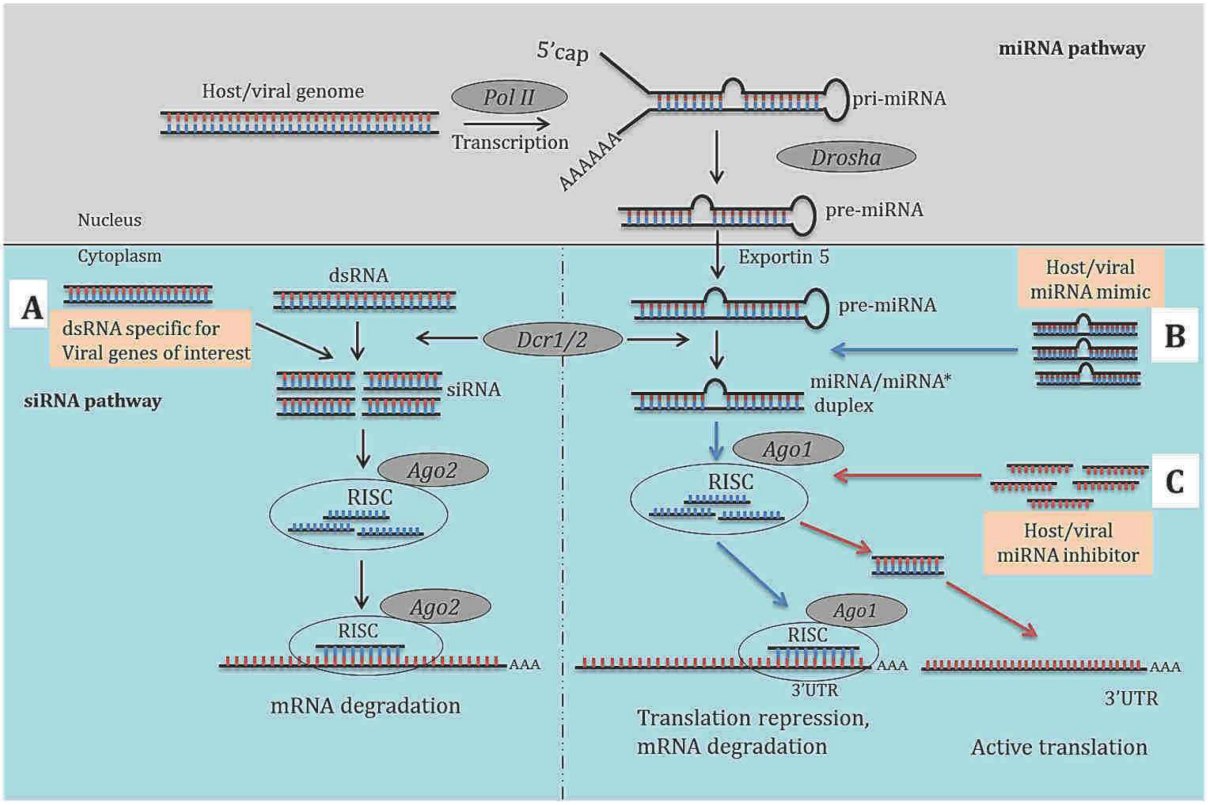


Figure 3: Proposed mechanism on the application of RNAi to control GpSGHV infections. A) DsRNAs specific for genes essential for virus replication can be injected/fed to the tsetse flies. Following the siRNA pathway, the siRNAs will bind to the targeted viral gene mRNA and induce its degradation. B) Mimics for the selected host or viral-encoded miRNAs can be injected to target and repress viral mRNAs while in C), introduction of miRNA inhibitor will complementarily bind to a mature miRNA and block it from repressing translation of its targeted mRNA.

b) Application of miRNA mimics/inhibitors to control viral infections

The expression levels of several host miRNAs changes following virus infections (Liu *et al.,* 2015; Saldaña *et al.,* 2017). Several host miRNAs expressed in *G. pallidipes* were identified and some were in deed modulated in symptomatic infections, for example the upregulated miR-184-3p (**Chapter 6**). Some studies have already utilized miRNA mimics and inhibitors to regulate miRNAs that target genes essential for virus replication or host immune genes (Stenvang *et al.,* 2012). For instance, the application of mimic or inhibitor of miRNA bantam in *Spodoptera litura*, led to a decrease and increase in virus replication respectively (Shi *et al.,* 2016). Usually, miRNA bantam expression levels increase after AcMNPV infection

in *S. litura.* This approach can as well be applied on the host and GpSGHV-encoded miRNAs identified in *G. pallidipes.* In this case, miRNA mimics can be applied to upregulate miRNAs that target viral genes and therefore suppress virus infection (**Figure 3B**). On the other hand, inhibitors of miRNAs for example those that target tsetse immunity genes can be applied and consequently restore their protein translation (**Figure 3C**). The host miRNAs modulated in symptomatic flies such as miR-184-3p, miR-283-5p, miR-263-5p and miR-276-5p as well as the GpSGHV-encoded miRNAs (especially those detected in both asymptomatic and symptomatic flies) can be potential candidates for this approach to manage GpSGHV infections.

c) Paratransgenesis for GpSGHV control

Paratransgenesis is a technique involving use of genetically modified symbiotic bacteria to express effector molecules to control a disease of interest (Wilke and Marrelli, 2015). This technique requires better understanding of symbiotic bacteria and the desired effector molecules. In mosquitoes, some bacteria species producing anti-*plasmodium* properties have been applied in *Anopheles* mosquito to control malaria (Ren *et al.,* 2008). The approach can have different outcomes on the vector once introduced; the genetically modified bacteria can become pathogenic to the host rather than a symbiont, interfere with the host's reproduction during oogenesis and embryogenesis, or reduce the vector competence (Wilke and Marrelli, 2015). However, paratransgenesis is considered more adaptable than the genetic modification of the vector because the effector molecules or the symbiotic bacteria can be replaced easily if required (Wilke and Marrelli, 2015). In tsetse flies, the paratransgenesis strategy has been considered promising by modifying *Sodalis* to express anti-trypanosome products, which were confirmed to reduce trypanosome infections (De Vooght *et al.,* 2014; 2012; Medlock *et al.,* 2013). Among the tsetse symbionts, *Sodalis* was selected mainly because it can be cultured and genetically modified *in vitro*. In addition, *Sodalis* is localised in different tsetse tissues where GpSGHV resides (i.e., hemolymph, milk glands, midguts and ovaries), which makes it a potential symbiont to express so-called nanobodies (antibody complexes) against viral proteins. Genes involved in viral entry (attachment) onto the host cells and signalling, such as the GpSGHV PIF genes, (key initiators of virus infection) (Kariithi *et al.,* 2011), would be potential targets for the effector molecules to control GpSGHV infections.

Alternatively, *Wolbachia* can be a potential symbiont to deliver the effector molecules, since it infects insect gonads and it is maternally transmitted and therefore molecules can be co-inherited with *Wolbachia*. However, a mechanism has to be developed for the effector molecules to reach other tissues to be effective such as salivary glands and hemolymph where the virus is localised. On the other hand, artificial introduction of *Wolbachia* in tsetse flies could induce antiviral protection (*Wolbachia*-immune priming) as shown in the case of *Ae. aegypti*, where decreased levels of DENV and chikungunya virus infection were observed in the mosquito following introduction of *Wolbachia* (Moreira *et al.,* 2009).

Concluding remarks

In mass rearing facilities of *G. pallidipes*, SGH outbreaks due to GpSGHV infections cause reproductive dysfunctionality of infected flies and negatively impact the sterile insect technique (SIT) application for this species, which requires production of large numbers of flies. The findings in this PhD thesis describe possible mechanism of GpSGHV evolution in tsetse species as well as the factors that may contribute to SGH outbreaks. First, the evolution of the two SGHVs (*Glossinavirus* and *Muscavirus)* with their respective hosts (tsetse fly and housefly) appears to have resulted in major differences in the number of host immunity genes. This clearly explains the differences in the induced pathologies of the these SGHVs. Further, this thesis has described multi-marker tools that can be used to distinguish tsetse species, not only to enable tracing of GpSGHV ancestral host but because SIT is a species-specific approach that requires accurate species identification during the initial establishment of a mass rearing facility. This is a great improvement for the identification of morphologically similar species and sub-species of tsetse flies, and similar approaches can be applied to develop tools that can distinctly identify other insects. Although GpSGHV infects at least seven *Glossina* species, the distribution, diversity and prevalence of the virus variants was highest in species belonging to the Morsitans group especially in *G. pallidipes*. The representative species of the Palpalis or Fusca group were infected with one virus haplotype. It therefore appears that GpSGHV's original host was a species belonging to one of these two *Glossina* groups and the virus has evolutionarily reached a stable but dynamic equilibrium state with these species but not species belonging to the Morsitans group. This indicates that species belonging to the Morsitans group are more recent hosts for

GpSGHV and that the virus is still adapting into its new host probably by accumulating mutations essential for its survival, which may account for the emerging virus strains and SGH outbreaks in *G. pallidipes* and not in other tsetse species.

Symptomatically infected *G. pallidipes* flies showed an impaired RNAi machinery, which is the primary antiviral immune response in insects. In addition, as reported for other insect viruses, GpSGHV infections do not only alter the host miRNA expression profile in *G. pallidipes* but the virus encodes miRNAs as well. These miRNAs may target and regulate host immunity genes or viral genes expression and therefore influence the outcome of GpSGHV infections. These findings indicate that the virus is capable of evading the host immune system to enhance its replication and dissemination by establishing symptomatic infections and consequently SGH outbreaks in tsetse.

Altogether, this thesis has provided some insights and alternatives to the precautions and control methods that can be considered during the establishment of new tsetse colonies to predict/prevent SGH outbreaks in tsetse rearing facilities, particularly where multiple *Glossina* species are mass produced. It could be recommended that the initial materials required for establishment of new insect facilities be screened for the presence of SGHV (pathogen). In addition to the already existing virus management strategies (antiviral drugs, e.g. valacyclovir, and clean feeding system) in already existing tsetse facilities, several other alternative recommendations include the following; (i) the application of RNAi silencing to target viral genes, (ii) utilization of mimics/inhibitors of GpSGHV modulated host and viral encoded miRNAs to suppress virus infection or restore expression of targeted immunity genes and (iii) paratransgenesis approach by genetic modification of symbiotic bacteria (*Sodalis*/*Wolbachia*) to express effector molecules (nanobodies) that target genes essential for virus entry into the host cell or viral replication.

The data provided in this thesis may not only contribute to the further understanding of virus-host interaction in invertebrates, more specifically with this relatively novel large double stranded DNA virus in tsetse flies. It also gives further support and alternatives in the management of mass rearing facilities of tsetse flies essential for the wider implementation of SIT-mediated control of tsetse and trypanosomiasis in sub-Saharan Africa.

References

Abd-Alla, A., Bossin, H., Cousserans, F., Parker, A., Bergoin, M., Robinson, A., 2007. Development of a non-destructive PCR method for detection of the salivary gland hypertrophy virus (SGHV) in tsetse flies. Journal of Virological Methods 139, 143-149.

Abd-Alla, A., Parker, A., Vreysen, M., Bergoin, M., 2011. Tsetse salivary gland hypertrophy virus: hope or hindrance for tsetse control? PloS Neglected Tropical Diseases 5, e1220.

Abd-Alla, A.M., Vlak, J.M., Bergoin, M., Maruniak, J.E., Parker, A., Burand, J.P., Jehle, J.A., Boucias, D.G., 2009a. *Hytrosaviridae*: a proposal for classification and nomenclature of a new insect virus family. Archives of Virology 154, 909-918.

Abd-Alla, A.M.M., Adun, H., Parker, A.G., Vreysen, M.J.B., Bergoin, M., 2012. The antiviral drug valacyclovir successfully suppresses salivary gland hypertrophy virus (SGHV) in laboratory colonies of *Glossina pallidipes*. PloS One 7, e38417.

Abd-Alla, A.M.M., Boucias, D.G., Bergoin, M., 2010a. Hytrosaviruses: structure and genomic properties., in: Asgari, S.J., K.N. (Ed.), Insect Virology. caister Academic Press: Norfolk, pp. 103-121.

Abd-Alla, A.M.M., Cousserans, F., Parker, A.G., Jehle, J.A., Parker, N.J., Vlak, J.M., Robinson, A.S., Bergoin, M., 2008. Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus reveals a novel, large, double-stranded circular DNA virus. Journal of Virology 82, 4595-4611.

Abd-Alla, A.M.M., Cousserans, F., Parker, A.G., Jridi, C., Bergoin, M., Robinson, A.S., 2009b. Quantitative PCR analysis of the salivary gland hypertrophy virus (GpSGHV) in a laboratory colony of *Glossina pallidipes*. Virus Research 139, 48-53.

Abd-Alla, A.M.M., Kariithi, H.M., Cousserans, F., Parker, N.J., İnce, İ.A., Scully, E.D., Boeren, S., Geib, S.M., Mekonnen, S., Vlak, J.M., Parker, A.G., Vreysen, M.J.B., Bergoin, M., 2016. Comprehensive annotation of *Glossina pallidipes* salivary gland hypertrophy virus from Ethiopian tsetse flies: a proteogenomics approach. Journal of General Virology 97, 1010-1031.

Abd-Alla, A.M.M., Kariithi, H.M., Mohamed, A.H., Lapiz, E., Parker, A.G., Vreysen, M.J.B., 2013. Managing hytrosavirus infections in *Glossina pallidipes* colonies: Feeding regime affects the prevalence of salivary gland hypertrophy syndrome. PloS One 8, e61875.

Abd-Alla, A.M.M., Kariithi, H.M., Parker, A.G., Robinson, A.S., Kiflom, M., Bergoin, M., Vreysen, M.J.B., 2010b. Dynamics of the salivary gland hypertrophy virus in laboratory colonies of *Glossina pallidipes* (Diptera: *Glossinidae*). Virus Research 150, 103-110.

Abd-Alla, A.M.M., Marin, C., Parker, A.G., Vreysen, M.J.B., 2014. Antiviral drug valacyclovir treatment combined with a clean feeding system enhances the suppression of salivary gland hypertrophy in laboratory colonies of *Glossina pallidipes*. Parasites & Vectors 7, 214-214.

Abila, P.P., Kiendrebeogo, M., Mutika, G.N., Parker, A.G., Robinson, A.S., 2003. The effect of age on the mating competitiveness of male *Glossina fuscipes fuscipes* and *Glossina palpalis palpalis*. Journal of Insect Science 3, 13.

Aksoy, E., Telleria, E.L., Echodu, R., Wu, Y., Okedi, L.M., Weiss, B.L., Aksoy, S., Caccone, A., 2014. Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. Applied and Environmental Microbiology 80, 4301-4312.

Aksoy, S., 2011. Sleeping sickness elimination in sight: time to celebrate and reflect, but not relax. PloS Neglected Tropical Diseases 5, e1008.

Alam, U., Hyseni, C., Symula, R.E., Brelsfoard, C., Wu, Y., Kruglov, O., Wang, J., Echodu, R., Alioni, V., Okedi, L.M., Caccone, A., Aksoy, S., 2012. Implications of microfauna-host interactions for trypanosome transmission dynamics in *Glossina fuscipes fuscipes* in Uganda. Applied and Environmental Microbiology 78, 4627-4637.

Alam, U., Medlock, J., Brelsfoard, C., Pais, R., Lohs, C., Balmand, S., Carnogursky, J., Heddi, A., Takac, P., Galvani, A., Aksoy, S., 2011. *Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans*. PloS Pathogens 7, e1002415.

Alemu, T., Kapitano, B., Mekonnen, S., Aboset, G., Kiflom, M., Bancha, B., Woldeyes, G., Bekele, K., Feldmann, U., 2007. Area-wide control of tsetse and trypanosomosis: Ethiopian experience in the Southern Rift Valley, in: Vreysen, M.J.B., Robinson, A.S., Hendrichs, J. (Eds.), Area-wide control of insect pests: From research to field implementation. Springer Netherlands, Dordrecht, pp. 325-335.

Allsopp, R., 2001. Options for vector control against trypanosomiasis in Africa. Trends in Parasitology 17, 15-19.

Allsopp, R., 2009. Control of tsetse flies (Diptera: *Glossinidae*) using insecticides: a review and future prospects. Bulletin of Entomological Research 74, 1-23.

Alsan, M., 2015. The effect of the tsetse fly on African development. American Economic Review 105, 382-410.

Amargier, A., Lyon, J.P., Vago, C., Meynadier, G., Veyrunes, J.C., 1979. Discovery and purification of a virus in gland hyperplasia of insects. Study of *Merodon equistris* F. (Diptera, *Syrphidae*). Comptes Rendus des Seances de l'Academie des Sciences. Serie D, Sciences Naturelles 289, 481-484.

An, C., Jiang, H., Kanost, M.R., 2010. Proteolytic activation and function of the cytokine Spätzle in innate immune response of a lepidopteran insect, *Manduca sexta*. The FEBS Journal 277, 148-162.

Angleró-Rodríguez, Y.I., Talyuli, O.A.C., Blumberg, B.J., Kang, S., Demby, C., Shields, A., Carlson, J., Jupatanakul, N., Dimopoulos, G., 2017. An *Aedes aegypti*-associated fungus increases susceptibility to Dengue virus by modulating gut trypsin activity. eLife 6, e28844.

Antonovics, J., Wilson, A.J., Forbes, M.R., Hauffe, H.C., Kallio, E.R., Leggett, H.C., Longdon, B., Okamura, B., Sait, S.M., Webster, J.P., 2017. The evolution of transmission

mode. Philosophical Transactions of the Royal Society Series B: Biological Sciences 372, 20160083.

Arif, B., Pavlik, L., 2013. Insect cell culture: virus replication and applications in biotechnology. Journal of Invertebrate Pathology 112, S138-S141.

Asgari, S., 2011. Role of microRNAs in insect host–microorganism interactions. Frontiers in Physiology 2, 48.

Asgari, S., 2013. MicroRNA functions in insects. Insect Biochemistry and Molecular Biology 43, 388-397.

Asgari, S., 2014. Role of microRNAs in arbovirus/vector interactions. Viruses 6, 3514-3534.

Asgari, S., 2015. Regulatory role of cellular and viral microRNAs in insect–virus interactions. Current Opinion in Insect Science 8, 104-110.

Avadhanula, V., Weasner, B.P., Hardy, G.G., Kumar, J.P., Hardy, R.W., 2009. A novel system for the launch of alphavirus RNA synthesis reveals a role for the Imd pathway in arthropod antiviral response. PloS Pathogens 5, e1000582.

Bah, T., 2011. Inkscape: guide to a vector drawing program. Prentice Hall Press.

Bahrndorff, S., de Jonge, N., Skovgård, H., Nielsen, J.L., 2017. Bacterial communities associated with houseflies (*Musca domestica* L.) sampled within and between farms. PloS One 12, e0169753.

Baker, M.D., Krafsur, E.S., 2001. Identification and properties of microsatellite markers in tsetse flies *Glossina morsitans* sensu lato (Diptera: *Glossinidae*). Molecular Ecology Notes 1, 234-236.

Baker, N., de Koning, H.P., Mäser, P., Horn, D., 2013. Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. Trends in Parasitology 29, 110-118.

Bekele, E., 2015. The current situation and diagnostic approach of nagana in Africa: a review. Journal of Natural Sciences Research 5, 2224-3186.

Benedict, C.A., Norris, P.S., Ware, C.F., 2002. To kill or be killed: viral evasion of apoptosis. Nature Immunology 3, 1013-1018.

Benoit, J.B., Attardo, G.M., Baumann, A.A., Michalkova, V., Aksoy, S., 2015. Adenotrophic viviparity in tsetse flies: potential for population control and as an insect model for lactation. Annual Review of Entomology 60, 351-371.

Best, S.M., 2008. Viral subversion of apoptotic enzymes: escape from death row. Annual Review of Microbiology 62, 171-192.

Bian, G., Xu, Y., Lu, P., Xie, Y., Xi, Z., 2010. The endosymbiotic bacterium *Wolbachi*a induces resistance to Dengue virus in *Aedes aegypti*. PloS Pathogens 6, e1000833.

Bloem, S., Carpenter, J.E., Dorn, S., 2006. Mobility of mass-reared diapaused and nondiapaused *Cydia pomonella* (Lepidoptera: *Tortricidae*): effect of different constant temperatures and lengths of cold storage. Journal of Economic Entomology 99, 707-713.

Boldogh, I., Albrecht, T., Porter, D.D., 1996. Persistent viral infections. In: Baron S, editor. Medical Microbiology 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston.

Boss, I.W., Renne, R., 2010. Viral miRNAs: tools for immune evasion. Current Opinion in Immunology 13, 540-545.

Boucias, D.G., Kariithi, H.M., Bourtzis, K., Schneider, D.I., Kelley, K., Miller, W.J., Parker, A.G., Abd-Alla, A.M.M., 2013. Transgenerational transmission of the *Glossina pallidipes* hytrosavirus depends on the presence of a functional symbiome. PloS One 8, e61150.

Bouyer, J., Seck, M.T., Sall, B., 2013. Misleading guidance for decision making on tsetse eradication: response to Shaw *et al.* (2013). Preventive Veterinary Medicine 112, 443-446.

Brelsfoard, C., Tsiamis, G., Falchetto, M., Gomulski, L.M., Telleria, E., Alam, U., Doudoumis, V., Scolari, F., Benoit, J.B., Swain, M., Takac, P., Malacrida, A.R., Bourtzis, K., Aksoy, S., 2014. Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. PloS Neglected Tropical Diseases 8, e2728.

Bronkhorst, A.W., van Cleef, K.W.R., Venselaar, H., van Rij, R.P., 2014. A dsRNA-binding protein of a complex invertebrate DNA virus suppresses the *Drosophila* RNAi response. Nucleic Acids Research 42, 12237-12248.

Bronkhorst, A.W., van Cleef, K.W.R., Vodovar, N., İnce, İ.A., Blanc, H., Vlak, J.M., Saleh, M.-C., van Rij, R.P., 2012. The DNA virus invertebrate iridescent virus 6 is a target of the *Drosophila* RNAi machinery. Proceedings of the National Academy of Sciences of the United States of America 109, 3604-3613.

Bronkhorst, A.W., van Rij, R.P., 2014. The long and short of antiviral defense: small RNA-based immunity in insects. Current Opinion in Virology 7, 19-28.

Brown, J.E., Komatsu, K.J., Abila, P.P., Robinson, A.S., Okedi, L.M.A., Dyer, N., Donnelly, M.J., Slotman, M.A., Caccone, A., 2008. Polymorphic microsatellite markers for the tsetse fly *Glossina fuscipes fuscipes* (Diptera: *Glossinidae*), a vector of human African trypanosomiasis. Molecular Ecology Resources 8, 1506-1508.

Brucker, R.M., Funkhouser, L.J., Setia, S., Pauly, R., Bordenstein, S.R., 2012. Insect innate immunity database (IIID): an annotation tool for identifying immune genes in insect genomes. PloS One 7, e45125.

Buchon, N., Broderick, N.A., Chakrabarti, S., Lemaitre, B., 2009. Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in *Drosophila*. Genes & Development 23, 2333-2344.

Burand, J.P., Hunter, W.B., 2013. RNAi: future in insect management. Journal of Invertebrate Pathology 112, S68-S74.

Burtt, E., 1945. Hypertrophied salivary glands in *Glossina*: evidence that *G. pallidipes* with this abnormality is peculiarly suited to trypanosome infection. Annals of Tropical Medicine & Parasitology 39, 11-13.

Büscher, P., Cecchi, G., Jamonneau, V., Priotto, G., 2017. Human African trypanosomiasis. The Lancet 390, 2397-2409.

Caljon, G., Broos, K., De Goeyse, I., De Ridder, K., Sternberg, J.M., Coosemans, M., De Baetselier, P., Guisez, Y., Van Den Abbeele, J., 2009. Identification of a functional Antigen5-related allergen in the saliva of a blood feeding insect, the tsetse fly. Insect Biochemistry and Molecular Biology 39, 332-341.

Caljon, G., De Ridder, K., De Baetselier, P., Coosemans, M., Van Den Abbeele, J., 2010. Identification of a tsetse fly salivary protein with dual inhibitory action on human platelet aggregation. PloS One 5, e9671.

Caragata, E.P., Rancès, E., Hedges, L.M., Gofton, A.W., Johnson, K.N., O'Neill, S.L., McGraw, E.A., 2013. Dietary cholesterol modulates pathogen blocking by *Wolbachia*. PloS Pathogens 9, e1003459.

Carmell, M.A., Hannon, G.J., 2004. RNase III enzymes and the initiation of gene silencing. Nature Structural & Molecular Biology 11, 214-218.

Carmell, M.A., Xuan, Z., Zhang, M.Q., Hannon, G.J., 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes & Development 16, 2733-2742.

Casa-Esper, d.l., #xf3, n, E., 2012. Horizontal transfer and the evolution of host-pathogen interactions. International Journal of Evolutionary Biology 2012, 9.

Cecchi, G., Mattioli, R.C., Slingenbergh, J., De La Rocque, S., 2008. Land cover and tsetse fly distributions in sub-Saharan Africa. Medical and Veterinary Entomology 22, 364-373.

Chaston, T.B., Lidbury, B.A., 2001. Genetic 'budget' of viruses and the cost to the infected host: a theory on the relationship between the genetic capacity of viruses, immune evasion, persistence and disease. Immunology and Cell Biology 79, 62-66.

Chen, X., Li, S., Li, C., Zhao, S., Aksoy, S., 1999. Phylogeny of genus *Glossina* (Diptera: *Glossinidae*) according to ITS2 sequences. Science China Life Sciences 42, 249-258.

Christensen, B.M., Li, J., Chen, C.-C., Nappi, A.J., 2005. Melanization immune responses in mosquito vectors. Trends in Parasitology 21, 192-199.

Cirimotich, C.M., Ramirez, J.L., Dimopoulos, G., 2011. Native microbiota shape insect vector competence for human pathogens. Cell Host & Microbe 10, 307-310.

Clarke, T.E., Clem, R.J., 2003. Insect defenses against virus infection: the role of apoptosis. International Reviews of Immunology 22, 401-424.

Clavijo, G., Williams, T., Muñoz, D., Caballero, P., López-Ferber, M., 2010. Mixed genotype transmission bodies and virions contribute to the maintenance of diversity in an insect virus. Proceedings of the Royal Society of London. Series B, Biological Sciences 277, 943-951.

Clem, R.J., 2001. Baculoviruses and apoptosis: the good, the bad, and the ugly. Cell Death and Differentiation 8, 137-143.

Coler, R.R., Boucias, D.G., Frank, J.H., Maruniak, J.E., Garcia-Canedo, A., Pendland, J.C., 1993. Characterization and description of a virus causing salivary gland hyperplasia in the housefly, *Musca domestica*. Medical and Veterinary Entomology 7, 275-282.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

Conrad, K.D., Giering, F., Erfurth, C., Neumann, A., Fehr, C., Meister, G., Niepmann, M., 2013. MicroRNA-122 dependent binding of Ago2 protein to Hepatitis C virus RNA is associated with enhanced RNA stability and translation stimulation. PloS One 8, e56272.

Cooper, D.M., Granville, D.J., Lowenberger, C., 2009. The insect caspases. Apoptosis 14, 247-256.

Cordon-Obras, C., Cano, J., Knapp, J., Nebreda, P., Ndong-Mabale, N., Ncogo-Ada, P.R., Ndongo-Asumu, P., Navarro, M., Pinto, J., Benito, A., Bart, J.-M., 2014. *Glossina palpalis palpalis* populations from Equatorial Guinea belong to distinct allopatric clades. Parasites & Vectors 7, 1-10.

Costa, A., Jan, E., Sarnow, P., Schneider, D., 2009. The IMD pathway is involved in antiviral immune responses in *Drosophila*. PloS One 4, e7436.

Cullen, B.R., 2009. Viral and cellular messenger RNA targets of viral microRNAs. Nature 457, 421-425.

Dale, C., Moran, N.A., 2006. Molecular interactions between bacterial symbionts and their hosts. Cell 126, 453-465.

Dayrat, B., 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society 85, 407-415.

De Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Research 34, W362-W365.

De Vooght, L., Caljon, G., De Ridder, K., Van Den Abbeele, J., 2014. Delivery of a functional anti-trypanosome nanobody in different tsetse fly tissues via a bacterial symbiont, *Sodalis glossinidius*. Microbial Cell Factories 13, 156.

De Vooght, L., Caljon, G., Stijlemans, B., De Baetselier, P., Coosemans, M., Van Den Abbeele, J., 2012. Expression and extracellular release of a functional anti-trypanosome nanobody® in *Sodalis glossinidius*, a bacterial symbiont of the tsetse fly. Microbial Cell Factories 11, 23.

De Vooght, L., Caljon, G., Van Hees, J., Van Den Abbeele, J., 2015. Paternal transmission of a secondary symbiont during mating in the viviparous tsetse fly. Molecular Biology and Evolution 32, 1977-1980.

Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D., Dostert, C., Antoniewski, C., Hoffmann, J.A., Imler, J.-L., 2008. The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *Drosophila*. Nature Immunology 9, 1425-1432.

Demirbas-Uzel, G., Kariithi, H.M., Parker, A.G., Vreysen, M.J.B., Mach, R.L., Abd-Alla, A.M.M., 2018. Susceptibility of tsetse species to *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV). Frontiers in Microbiology 9, 701.

Dill, J.A., Camus, A.C., Leary, J.H., Di Giallonardo, F., Holmes, E.C., Ng, T.F.F., 2016. Distinct viral lineages from fish and amphibians reveal the complex evolutionary history of hepadnaviruses. Journal of Virology 90, 7920-7933.

Ding, S.-W., 2010. RNA-based antiviral immunity. Nature Reviews Immunology 10, 632-644.

Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J.A., Imler, J.-L., 2005. The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. Nature Immunology 6, 946-953.

Doudoumis, V., Alam, U., Aksoy, E., Abd-Alla, A.M.M., Tsiamis, G., Brelsfoard, C., Aksoy, S., Bourtzis, K., 2013. Tsetse-*Wolbachia* symbiosis: comes of age and has great potential for pest and disease control. Journal of Invertebrate Pathology 112, S94-S103.

Doudoumis, V., Blow, F., Saridaki, A., Augustinos, A., Dyer, N.A., Goodhead, I., Solano, P., Rayaisse, J.B., Takac, P., Mekonnen, S., Parker, A.G., Abd-Alla, A.M.M., Darby, A., Bourtzis, K., Tsiamis, G., 2017. Challenging the *Wigglesworthia*, *Sodalis*, *Wolbachia* symbiosis dogma in tsetse flies: *Spiroplasma* is present in both laboratory and natural populations. Scientific Reports 7, 4699.

Doudoumis, V., Tsiamis, G., Wamwiri, F., Brelsfoard, C., Alam, U., Aksoy, E., Dalaperas, S., Abd-Alla, A., Ouma, J., Takac, P., Aksoy, S., Bourtzis, K., 2012. Detection and characterization of *Wolbachia* infections in laboratory and natural populations of different species of tsetse flies (genus *Glossina*). BMC Microbiology 12, S3-S3.

Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nature Reviews Genetics 9, 267-276.

Dyer, N.A., Furtado, A., Cano, J., Ferreira, F., Odete Afonso, M., Ndong-Mabale, N., Ndong-Asumu, P., Centeno-Lima, S., Benito, A., Weetman, D., 2009. Evidence for a discrete evolutionary lineage within Equatorial Guinea suggests that the tsetse fly *Glossina palpalis palpalis* exists as a species complex. Molecular Ecology 18, 3268-3282.

Dyer, N.A., Lawton, S.P., Ravel, S., Choi, K.S., Lehane, M.J., Robinson, A.S., Okedi, L.M., Hall, M.J.R., Solano, P., Donnelly, M.J., 2008. Molecular phylogenetics of tsetse flies (Diptera: *Glossinidae*) based on mitochondrial (COI, 16S, ND2) and nuclear ribosomal DNA sequences, with an emphasis on the Palpalis group. Molecular Phylogenetics and Evolution 49, 227-239.

Dyer, N.A., Ravel, S., Choi, K.-S., Darby, A.C., Causse, S., Kapitano, B., Hall, M.J.R., Steen, K., Lutumba, P., Madinga, J., Torr, S.J., Okedi, L.M., Lehane, M.J., Donnelly, M.J., 2011.

Cryptic diversity within the major trypanosomiasis vector *Glossina fuscipes* revealed by molecular markers. PloS Neglected Tropical Diseases 5, e1266.

Dykxhoorn, D.M., Lieberman, J., 2006. Silencing viral infection. PloS Medicine 3, e242.

Echodu, R., Sistrom, M., Hyseni, C., Enyaru, J., Okedi, L., Aksoy, S., Caccone, A., 2013. Genetically distinct *Glossina fuscipes fuscipes* populations in the lake Kyoga region of Uganda and its relevance for human African trypanosomiasis. BioMed Research International 2013, 12.

Eleftherianos, I., Atri, J., Accetta, J., Castillo, J.C., 2013. Endosymbiotic bacteria in insects: guardians of the immune system? Frontiers in Physiology 4, 46.

Ellis, D.S., Maudlin, I., 1987. Salivary gland hyperplasia in wild caught tsetse from Zimbabwe. Entomologia Experimentalis et Applicata 45, 167-173.

Elsen, P., A Amoudi, M., Leclercq, M., 1991. First record of *Glossina fuscipes fuscipes* Newstead, 1910 and *Glossina morsitans submorsitans* Newstead, 1910 in SouthWestern Saudi Arabia. Annales de la Société Belge de Médecine Tropicale 70, 281-287.

Enkerlin, W., Gutiérrez-Ruelas, J.M., Cortes, A.V., Roldan, E.C., Midgarden, D., Lira, E., López, J.L.Z., Hendrichs, J., Liedo, P., Arriaga, F.J.T., 2015. Area freedom in Mexico from Mediterranean fruit fly (Diptera: *Tephritidae*): a review of over 30 years of a successful containment program using an integrated area-wide SIT approach. Florida Entomologist 98, 665-681.

Everett, H., McFadden, G., 1999. Apoptosis: an innate immune response to virus infection. Trends in Microbiology 7, 160-165.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47-50.

FAO, 2014. Senegal nears first victory in war on tsetse fly. [Internet].

Feldmann, U., 1994. Guidelines for the rearing of tsetse flies using the membrane feeding technique. In: Ochieng'-Odero, J.P.R. editor. Techniques of insect rearing for the development of integrated pest and vector management strategies. Nairobi, Kenya: ICIPE Science Press, pp. 449-471.

Feldmann, U., Dyck, V.A., Mattioli, R.C., Jannin, J., 2005. Potential impact of tsetse fly control involving the sterile insect technique, in: Dyck, V.A., Hendrichs, J., Robinson, A.S. (Eds.), Sterile Insect Technique. Springer Netherlands, pp. 701-723.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research 44, D279-D285.

Foley, E., O'Farrell, P.H., 2004. Functional dissection of an innate immune response by a genome-wide RNAi screen. PloS Biology 2, e203.

Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Mattioli, R.C., Argaw, D., 2017. Monitoring the elimination of human African trypanosomiasis: update to 2014. PloS Neglected Tropical Diseases 11, e0005585.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology & Evolution 27, 480-488.

Fullaondo, A., Lee, S.Y., 2012. Regulation of *Drosophila*-virus interaction. Developmental & Comparative Immunology 36, 262-266.

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J.A., Imler, J.-L., 2006. Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. Nature Immunology 7, 590-597.

Gammon, D.B., Mello, C.C., 2015. RNA interference-mediated antiviral defense in insects. Current Opinion in Insect Science 8, 111-120.

Garcia-Maruniak, A., Abd-Alla, A.M.M., Salem, T.Z., Parker, A.G., Lietze, V.-U., van Oers, M.M., Maruniak, J.E., Kim, W., Burand, J.P., Cousserans, F., Robinson, A.S., Vlak, J.M., Bergoin, M., Boucias, D.G., 2009. Two viruses that cause salivary gland hypertrophy in *Glossina pallidipes* and *Musca domestica* are related and form a distinct phylogenetic clade. Journal of General Virology 90, 334-346.

Garcia-Maruniak, A., Maruniak, J.E., Farmerie, W., Boucias, D.G., 2008. Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV. Virology 377, 184-196.

Gariou-Papalexiou, A., Yannopoulos, G., Robinson, A.S., Zacharopoulou, A., 2007. Polytene chromosome maps in four species of tsetse flies *Glossina austeni*, *G. pallidipes*, *G. morsitans morsitans* and *G. m. submorsitans* (Diptera: *Glossinidae*): a comparative analysis. Genetica 129, 243-251.

Geden, C., Garcia-Maruniak, A., Lietze, V.U., Maruniak, J., Boucias, D.G., 2011a. Impact of house fly salivary gland hypertrophy virus (MdSGHV) on a heterologous host, *Stomoxys calcitrans*. Journal of Medical Entomology 48, 1128-1135.

Geden, C.J., Lietze, V.-U., Boucias, D.G., 2008. Seasonal prevalence and transmission of salivary gland hypertrophy virus of house flies (Diptera: *Muscidae*). Journal of Medical Entomology 45, 42-51.

Geden, C.J., Steenberg, T., Lietze, V.U., Boucias, D.G., 2011b. Salivary gland hypertrophy virus of house flies in Denmark: prevalence, host range, and comparison with a Florida isolate. Journal of Vector Ecology 36, 231-238.

Geerts, S., Holmes, P.H., Eisler, M.C., Diall, O., 2001. African bovine trypanosomiasis: the problem of drug resistance. Trends in Parasitology 17, 25-28.

Geiger, A., Fardeau, M.-L., Falsen, E., Ollivier, B., Cuny, G., 2010. *Serratia glossinae* sp. nov., isolated from the midgut of the tsetse fly *Glossina palpalis gambiensis*. International Journal of Systematic and Evolutionary Microbiology 60, 1261-1265.

Geiger, A., Fardeau, M.-L., Grebaut, P., Vatunga, G., Josénando, T., Herder, S., Cuny, G., Truc, P., Ollivier, B., 2009. First isolation of *Enterobacter*, *Enterococcus*, and *Acinetobacter* spp. as inhabitants of the tsetse fly (*Glossina palpalis palpalis*) midgut. Infection, Genetics and Evolution 9, 1364-1370.

Geiger, A., Fardeau, M.-L., Njiokou, F., Joseph, M., Asonganyi, T., Ollivier, B., Cuny, G., 2011. Bacterial diversity associated with populations of *Glossina* spp. from Cameroon and distribution within the campo sleeping sickness focus. Microbial Ecology 62, 632-643.

Gemayel, R., Cho, J., Boeynaems, S., Verstrepen, K.J., 2012. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. Genes 3, 461-480.

Geoghegan, J.L., Duchêne, S., Holmes, E.C., 2017. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PloS Pathogens 13, e1006215.

Ghazal, P., García-Ramírez, J., González-Armas, J.C., Kurz, S., Angulo, A., 2000a. Principles of homeostasis in governing virus activation and latency. Immunologic Research 21, 219-223.

Ghazal, P., González Armas, J.C., García-Ramírez, J.J., Kurz, S., Angulo, A., 2000b. Viruses: hostages to the Cell. Virology 275, 233-237.

Giraldo-Calderón, G.I., Emrich, S.J., MacCallum, R.M., Maslen, G., Dialynas, E., Topalis, P., Ho, N., Gesing, S., the VectorBase, C., Madey, G., Collins, F.H., Lawson, D., 2015. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Research 43, D707-D713.

Glaser, R.L., Meola, M.A., 2010. The native *Wolbachia* endosymbionts of *Drosophila melanogaster* and *Culex quinquefasciatus* increase host resistance to West Nile virus infection. PloS One 5, e11977.

Goic, B., Saleh, M.-C., 2012. Living with the enemy: viral persistent infections from a friendly viewpoint. Current Opinion in Microbiology 15, 531-537.

Goic, B., Vodovar, N., Mondotte, J.A., Monot, C., Frangeul, L., Blanc, H., Gausson, V., Vera-Otarola, J., Cristofari, G., Saleh, M.-C., 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. Nature Immunology 14, 396-403.

Gooding, R.H., Krafsur, E.S., 2005. Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. Annual Review of Entomology 50, 101-123.

Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W.H., Chi, J.-T.A., Braich, R., Manoharan, M., Soutschek, J., Ohler, U., Cullen, B.R., 2007. A viral microRNA functions as an ortholog of cellular miR-155. Nature 450, 1096-1099.

Gouli, V., Gouli, S., Marcelino, J., 2011. Principal group of infectious insect diseases, Common infectious diseases of insects in culture: diagnostic and prophylactic methods. Springer Netherlands, Dordrecht, pp. 7-42.

Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G., Kim, J., 2003. Computational and experimental identification of *C. elegans* microRNAs. Molecular Cell 11, 1253-1263.

Graham, R.I., Grzywacz, D., Mushobozi, W.L., Wilson, K., Ebert, D., 2012. *Wolbachia* in a major African crop pest increases susceptibility to viral disease rather than protects. Ecology Letters 15, 993-1000.

Grant, I.F., 2001. Insecticides for tsetse and trypanosomiasis control: is the environmental risk acceptable? Trends in Parasitology 17, 10-14.

Green, C.H., 1994. Bait methods for tsetse fly control, in: Baker, J.R., Muller, R., Rollinson, D. (Eds.), Advances in Parasitology. Academic Press, pp. 229-291.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2007. miRBase: tools for microRNA genomics. Nucleic Acids Research 36, D154-D158.

Griffiths, A., 2011. Slipping and sliding: frameshift mutations in herpes simplex virus thymidine kinase and drug-resistance. Drug resistance updates: reviews and commentaries in antimicrobial and anticancer chemotherapy 14, 251-259.

Grubaugh, N.D., Weger-Lucarelli, J., Murrieta, R.A., Fauver, J.R., Garcia-Luna, S.M., Prasad, A.N., Black Iv, W.C., Ebel, G.D., 2016. Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. Cell Host & Microbe 19, 481-492.

Grundhoff, A., 2011. Computational prediction of viral miRNAs, in: van Rij, R.P. (Ed.), Antiviral RNAi: concepts, methods, and applications. Humana Press, Totowa, NJ, pp. 143-152.

Grundhoff, A., Sullivan, C.S., 2011. Virus-encoded microRNAs. Virology 411, 325-343.

Gu, S., Jin, L., Huang, Y., Zhang, F., Kay, M.A., 2012. Slicing-independent RISC activation requires the argonaute PAZ domain. Current Biology 22, 1536-1542.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyML 3.0. Systematic Biology 59, 307-321.

Guo, Y.E., Steitz, J.A., 2014. Virus meets host microRNA: the destroyer, the booster, the hijacker. Molecular and Cellular Biology 34, 3780-3787.

Haac, M.E., Anderson, M.A.E., Eggleston, H., Myles, K.M., Adelman, Z.N., 2015. The hub protein loquacious connects the microRNA and short interfering RNA pathways in mosquitoes. Nucleic Acids Research 43, 3688-3700.

Haines, L.R., Lehane, S.M., Pearson, T.W., Lehane, M.J., 2010. Tsetse EP protein protects the fly midgut from trypanosome establishment. PloS Pathogens 6, e1000793.

Hall, T., 2011. BioEdit: an important software for molecular biology. GERF Bulletin of Biosciences 2, 30-61.

Hammond, S.M., 2005. Dicing and slicing. FEBS Letters 579, 5822-5829.

Hay, B.A., Guo, M., 2006. Caspase-dependent cell death in *Drosophila*. Annual Review of Cell and Developmental Biology 22, 623-650.

He, Y., Yang, K., Zhang, X., 2014. Viral microRNAs targeting virus genes promote virus infection in shrimp in vivo. Journal of Virology 88, 1104-1112.

Hedges, L.M., Brownlie, J.C., O'Neill, S.L., Johnson, K.N., 2008. *Wolbachia* and virus protection in insects. Science 322, 702-702.

Hegde, S., Rasgon, J.L., Hughes, G.L., 2015. The microbiome modulates arbovirus transmission in mosquitoes. Current Opinion in Virology 15, 97-102.

Herniou, E.A., Olszewski, J.A., O'Reilly, D.R., Cory, J.S., 2004. Ancient coevolution of baculoviruses and their insect hosts. Journal of Virology 78, 3244-3251.

Huang, Q., Deveraux, Q.L., Maeda, S., Salvesen, G.S., Stennicke, H.R., Hammock, B.D., Reed, J.C., 2000. Evolutionary conservation of apoptosis mechanisms: lepidopteran and baculoviral inhibitor of apoptosis proteins are inhibitors of mammalian caspase-9. Proceedings of the National Academy of Sciences of the United States of America 97, 1427-1432.

Huang, Z., Kingsolver, M.B., Avadhanula, V., Hardy, R.W., 2013. An antiviral role for antimicrobial peptides during the arthropod response to alphavirus replication. Journal of Virology 87, 4272-4280.

Hudson, R.R., 1990. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7, 1-44.

Hussain, M., Asgari, S., 2010. Functional analysis of a cellular microRNA in insect host-ascovirus interaction. Journal of Virology 84, 612-620.

Hussain, M., Frentiu, F.D., Moreira, L.A., O'Neill, S.L., Asgari, S., 2011. *Wolbachia* uses host microRNAs to manipulate host gene expression and facilitate colonization of the Dengue vector *Aedes aegypti*. Proceedings of the National Academy of Sciences 108, 9250-9255.

Hussain, M., Taft, R.J., Asgari, S., 2008. An insect virus-encoded microRNA regulates viral replication. Journal of Virology 82, 9164-9170.

Huvenne, H., Smagghe, G., 2010. Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review. Journal of Insect Physiology 56, 227-235.

Idrees, S., Ashfaq, U.A., 2013. RNAi: antiviral therapy against Dengue virus. Asian Pacific Journal of Tropical Biomedicine 3, 232-236.

International Glossina Genome Initiative, 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. Science 344, 380-386.

Iovino, N., Pane, A., Gaul, U., 2009. MiR-184 has multiple roles in *Drosophila* female germline development. Developmental Cell 17, 123-133.

Ishimwe, E., Hodgson, J.J., Clem, R.J., Passarelli, A.L., 2015. Reaching the melting point: degradative enzymes and protease inhibitors involved in baculovirus infection and dissemination. Virology 479-480, 637-649.

Jaenson, T.G.T., 1978. Virus-like rods associated with salivary gland hyperplasia in tsetse, *Glossina pallidipes*. Transactions of The Royal Society of Tropical Medicine and Hygiene 72, 234-238.

Jaenson, T.G.T., 1986. Sex ratio distortion and reduced lifespan of *Glossina pallidipes* infected with the virus causing salivary gland hyperplasia. Entomologia Experimentalis et Applicata 41, 265-271.

Jakubowska, A.K., Vogel, H., Herrero, S., 2013. Increase in gut microbiota after immune suppression in baculovirus-infected larvae. PloS Pathogens 9, e1003379.

Janeway, C.A., 1989. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harbor Symposia on Quantitative Biology 54, 1-13.

Jayachandran, B., Hussain, M., Asgari, S., 2012. RNA interference as a cellular defense mechanism against the DNA virus baculovirus. Journal of Virology 86, 13729-13734.

Jehle, J.A., Abd-Alla, A.M.M., Wang, Y., 2013. Phylogeny and evolution of *Hytrosaviridae*. Journal of Invertebrate Pathology 112, S62-S67.

Jenni, L., 1973. Virus-like particles in a strain of *G. morsitans centralis*, Machado 1970. Transactions of The Royal Society of Tropical Medicine and Hygiene 67, 295-295.

Jenni, L., Steiger, R., 1974a. Virus like particles of *Glossina fuscipes fuscipes* Newst. 1910. Acta Tropica 31, 177-180.

Jenni, L., Steiger, R.F., 1974b. Virus like particles in the tsetse fly, *Glossina morsitans* sspp. Preliminary results. Revue Suisse de Zoologie 81, 663-666.

Johnson, K.N., 2015. The impact of *Wolbachia* on virus infection in mosquitoes. Viruses 7, 5705-5717.

Jupatanakul, N., Sim, S., Dimopoulos, G., 2014. The insect microbiome modulates vector competence for arboviruses. Viruses 6, 4294-4313.

Jura, W.G.Z.O., Odhiambo, T.R., Otieno, L.H., Tabu, N.O., 1988. Gonadal lesions in virus-infected male and female tsetse, *Glossina pallidipes* (Diptera: *Glossinidae*). Journal of Invertebrate Pathology 52, 1-8.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., Petrov, A.I., 2017. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Research, 1038-1038.

Kambris, Z., Blagborough, A.M., Pinto, S.B., Blagrove, M.S.C., Godfray, H.C.J., Sinden, R.E., Sinkins, S.P., 2010. *Wolbachia* stimulates immune gene expression and inhibits *Plasmodium* development in *Anopheles gambiae*. PloS Pathogens 6, e1001143.

Kanost, M.R., Jiang, H., 2015. Clip-domain serine proteases as immune factors in insect hemolymph. Current Opinion in Insect Science 11, 47-55.

Kapun, M., Nolte, V., Flatt, T., Schlötterer, C., 2010. Host range and specificity of the *Drosophila* C virus. PloS One 5, e12421.

Kariithi, H., van Oers, M., Vlak, J., Vreysen, M., Parker, A., Abd-Alla, A., 2013a. Virology, epidemiology and pathology of *Glossina* hytrosavirus, and its control prospects in laboratory colonies of the tsetse fly, *Glossina pallidipes* (Diptera; *Glossinidae*). Insects 4, 287-319.

Kariithi, H.M., Ahmadi, M., Parker, A.G., Franz, G., Ros, V.I.D., Haq, I., Elashry, A.M., Vlak, J.M., Bergoin, M., Vreysen, M.J.B., Abd-Alla, A.M.M., 2013b. Prevalence and genetic variation of salivary gland hypertrophy virus in wild populations of the tsetse fly *Glossina pallidipes* from southern and eastern Africa. Journal of Invertebrate Pathology 112, S123-S132.

Kariithi, H.M., Ince, I.A., Boeren, S., Abd-Alla, A.M.M., Parker, A.G., Aksoy, S., Vlak, J.M., van Oers, M.M., 2011. The salivary secretome of the tsetse fly *Glossina pallidipes* (Diptera: *Glossinidae*) infected by salivary gland hypertrophy virus. PloS Neglected Tropical Diseases 5, e1371.

Kariithi, H.M., İnce, İ.A., Boeren, S., Murungi, E.K., Meki, I.K., Otieno, E.A., Nyanjom, S.R.G., van Oers, M.M., Vlak, J.M., Abd-Alla, A.M.M., 2016. Comparative analysis of salivary gland proteomes of two *Glossina* species that exhibit differential hytrosavirus pathologies. Frontiers in Microbiology 7, 89.

Kariithi, H.M., Meki, I.K., Boucias, D.G., Abd-Alla, A.M.M., 2017a. Hytrosaviruses: current status and perspective. Current Opinion in Insect Science 22, 71-78.

Kariithi, H.M., van Lent, J.W.M., Boeren, S., Abd-Alla, A.M.M., İnce, İ.A., van Oers, M.M., Vlak, J.M., 2013c. Correlation between structure, protein composition, morphogenesis and cytopathology of *Glossina pallidipes* salivary gland hypertrophy virus. Journal of General Virology 94, 193-208.

Kariithi, H.M., Yao, X., Yu, F., Teal, P.E., Verhoeven, C.P., Boucias, D.G., 2017b. Responses of the housefly, *Musca domestica*, to the hytrosavirus replication: impacts on host's vitellogenesis and immunity. Frontiers in Microbiology 8, 583.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30, 772-780.

Keene, K.M., Foy, B.D., Sanchez-Vargas, I., Beaty, B.J., Blair, C.D., Olson, K.E., 2004. RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; *Togaviridae*) infection of *Anopheles gambiae*. Proceedings of the National Academy of Sciences of the United States of America 101, 17240-17245.

Kemp, C., Mueller, S., Goto, A., Barbier, V., Paro, S., Bonnay, F., Dostert, C., Troxler, L., Hetru, C., Meignin, C., Pfeffer, S., Hoffmann, J.A., Imler, J.-L., 2013. Broad RNA interference-mediated antiviral immunity and virus-specific inducible responses in *Drosophila*. Journal of Immunology 190, 650-658.

Kennedy, P.G.E., 2013. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). The Lancet Neurology 12, 186-194.

Kennerdell, J.R., Yamaguchi, S., Carthew, R.W., 2002. RNAi is activated during *Drosophila* oocyte maturation in a manner dependent on aubergine and spindle-E. Genes & Development 16, 1884-1889.

Kgori, P.M., Modo, S., Torr, S.J., 2006. The use of aerial spraying to eliminate tsetse from the Okavango Delta of Botswana. Acta Tropica 99, 184-199.

Kim, C.H., Shin, Y.P., Noh, M.Y., Jo, Y.H., Han, Y.S., Seong, Y.S., Lee, I.H., 2010. An insect multiligand recognition protein functions as an opsonin for the phagocytosis of microorganisms. The Journal of Biological Chemistry 285, 25243-25250.

Kincaid, R.P., Sullivan, C.S., 2012. Virus-encoded microRNAs: an overview and a look to the future. PloS Pathogens 8, e1003018.

Kingsolver, M.B., Huang, Z., Hardy, R.W., 2013. Insect antiviral innate immunity: pathways, effectors, and connections. Journal of Molecular Biology 425, 4921-4936.

Knipling, E.F., 1959. Sterile-male method of population control: successful with some insects, the method may also be effective when applied to other noxious animals. Science 130, 902-904.

Kodandaramaiah, U., Simonsen, T.J., Bromilow, S., Wahlberg, N., Sperling, F., 2013. Deceptive single-locus taxonomy and phylogeography: *Wolbachia*-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. Ecology and Evolution 3, 5167-5176.

Kotzamanis, K., Angulo, A., Ghazal, P., 2015. Infection homeostasis: implications for therapeutic and immune programming of metabolism in controlling infection. Medical Microbiology and Immunology 204, 395-407.

Krafsur, E.S., 2009. Tsetse flies: genetics, evolution, and role as vectors. Infection, Genetics and Evolution 9, 124-141.

Krafsur, E.S., Endsley, M.A., 2002. Microsatellite diversities and gene flow in the tsetse fly, *Glossina morsitans* s.l. Medical and Veterinary Entomology 16, 292-300.

Kremer, N., Charif, D., Henri, H., Gavory, F., Wincker, P., Mavingui, P., Vavre, F., 2012. Influence of *Wolbachia* on host gene expression in an obligatory symbiosis. BMC Microbiology 12, S7-S7.

Krüger, J., Rehmsmeier, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Research 34, W451-W454.

Kwon, C., Han, Z., Olson, E.N., Srivastava, D., 2005. MicroRNA1 influences cardiac differentiation in *Drosophila* and regulates Notch signaling. Proceedings of the National Academy of Sciences of the United States of America 102, 18986-18991.

Lalić, J., Cuevas, J.M., Elena, S.F., 2011. Effect of host species on the distribution of mutational fitness effects for an RNA virus. PloS Genetics 7, e1002378.

Lamiable, O., Imler, J.-L., 2014. Induced antiviral innate immunity in *Drosophila*. Current Opinion in Immunology 20, 62-68.

Lamkanfi, M., Declercq, W., Kalai, M., Saelens, X., Van Den Abeele, P., 2002. Alice in caspase land. A phylogenetic analysis of caspases from worm to man. Cell Death and Differentiation 9, 358-361.

Lan, H., Chen, H., Liu, Y., Jiang, C., Mao, Q., Jia, D., Chen, Q., Wei, T., 2016. Small interfering RNA pathway modulates initial viral infection in midgut epithelium of insect after ingestion of Virus. Journal of Virology 90, 917-929.

Langley, P.A., Maly, H., 1969. Membrane feeding technique for tsetse flies (*Glossina* spp.). Nature 221, 855-856.

Lazzaro, B.P., Little, T.J., 2009. Immunity in a variable world. Philosophical Transactions of the Royal Society Series B: Biological Sciences 364, 15-26.

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, Wallingford.

Lee, J.-C., Chen, H.-H., Chao, Y.-C., 1998. Persistent baculovirus infection results from deletion of the apoptotic suppressor gene p35. Journal of Virology 72, 9157-9165.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., Carthew, R.W., 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell 117, 69-81.

Lemaitre, B., Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*. Annual Review of Immunology 25, 697-743.

Li, F., Ding, S.-W., 2006. Virus counterdefense: diverse strategies for evading the rna-silencing immunity. Annual Review of Microbiology 60, 503-531.

Li, J., Yuan, J., 2008. Caspases in apoptosis and beyond. Oncogene 27, 6194-6206.

Lieberman, P.M., 2016. Epigenetics and genetics of viral latency. Cell Host & Microbe 19, 619-628.

Lietze, V.-U., Abd-Alla, A.M.M., Vreysen, M.J.B., Geden, C.J., Boucias, D.G., 2011a. Salivary gland hypertrophy viruses: a novel group of insect pathogenic viruses. Annual Review of Entomology 56, 63-80.

Lietze, V.-U., Geden, C.J., Blackburn, P., Boucias, D.G., 2007. Effects of salivary gland hypertrophy virus on the reproductive behavior of the housefly, *Musca domestica*. Applied and Environmental Microbiology 73, 6811-6818.

Lietze, V.-U., Geden, C.J., Doyle, M.A., Boucias, D.G., 2012. Disease dynamics and persistence of *Musca domestica* salivary gland hypertrophy virus infections in laboratory house fly (*Musca domestica*) populations. Applied and Environmental Microbiology 78, 311-317.

Lietze, V.-U., Keesling, J.E., Lee, J.A., Vallejo, C.R., Geden, C.J., Boucias, D.G., 2013. *Muscavirus* (MdSGHV) disease dynamics in house fly populations – how is this virus transmitted and has it potential as a biological control agent? Journal of Invertebrate Pathology 112, S40-S43.

Lietze, V.-U., Salem, T.Z., Prompiboon, P., Boucias, D.G., 2011b. Tissue tropism of the *Musca domestica* salivary gland hypertrophy virus. Virus Research 155, 20-27.

Lietze, V.-U., Sims, K.R., Salem, T.Z., Geden, C.J., Boucias, D.G., 2009. Transmission of MdSGHV among adult house flies, *Musca domestica* (Diptera: *Muscidae*), occurs via oral secretions and excreta. Journal of Invertebrate Pathology 101, 49-55.

Lindh, J.M., Lehane, M.J., 2011. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: *Glossina*) harbours a surprising diversity of bacteria other than symbionts. Antonie van Leeuwenhoek 99, 711-720.

Liu, Y., Zhou, Y., Wu, J., Zheng, P., Li, Y., Zheng, X., Puthiyakunnon, S., Tu, Z., Chen, X.-G., 2015. The expression profile of *Aedes albopictus* miRNAs is altered by dengue virus serotype-2 infection. Cell & Bioscience 5, 16.

Lobo, F.P., Mota, B.E.F., Pena, S.D.J., Azevedo, V., Macedo, A.M., Tauch, A., Machado, C.R., Franco, G.R., 2009. Virus-host coevolution: common patterns of nucleotide motif usage in *Flaviviridae* and their hosts. PloS One 4, e6282.

Longdon, B., Brockhurst, M.A., Russell, C.A., Welch, J.J., Jiggins, F.M., 2014. The evolution and genetics of virus host shifts. PloS Pathogens 10, e1004395.

López-Uribe, M.M., Sconiers, W.B., Frank, S.D., Dunn, R.R., Tarpy, D.R., 2016. Reduced cellular immune response in social insect lineages. Biology Letters 12, 20150984.

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms for Molecular Biology 6, 26-26.

Loverdo, C., Lloyd-Smith, J.O., 2013. Evolutionary invasion and escape in the presence of deleterious mutations. PloS One 8, e68179.

Lu, F., Stedman, W., Yousef, M., Renne, R., Lieberman, P.M., 2010. Epigenetic regulation of Kaposi's sarcoma-associated herpesvirus latency by virus-encoded microRNAs that target RTA and the cellular Rbl2-DNMT pathway. Journal of Virology 84, 2697-2706.

Lucas, K., Raikhel, A.S., 2013. Insect microRNAs: biogenesis, expression profiling and biological functions. Insect Biochemistry and Molecular Biology 43, 24-38.

Luna, C., Bonizzoni, M., Cheng, Q., Robinson, A.S., Aksoy, S., Zheng, L., 2001. Microsatellite polymorphism in tsetse flies (Diptera: *Glossinidae*). Journal of Medical Entomology 38, 376-381.

Luo, S., Lu, J., 2017. Silencing of transposable elements by piRNAs in *Drosophila*: an evolutionary perspective. Genomics, Proteomics & Bioinformatics 15, 164-176.

Maharaj, P.D., Widen, S.G., Huang, J., Wood, T.G., Thangamani, S., 2015. Discovery of mosquito saliva microRNAs during CHIKV infection. PloS Neglected Tropical Diseases 9, e0003386.

Manji, G.A., Hozak, R.R., LaCount, D.J., Friesen, P.D., 1997. Baculovirus inhibitor of apoptosis functions at or upstream of the apoptotic suppressor P35 to prevent programmed cell death. Journal of Virology 71, 4509-4516.

Marquez, J.G., Vreysen, M.J.B., Robinson, A.S., Bado, S., Krafsur, E.S., 2004. Mitochondrial diversity analysis of *Glossina palpalis gambiensis* from Mali and Senegal. Medical and Veterinary Entomology 18, 288-295.

Marquitz, A.R., Mathur, A., Nam, C.S., Raab-Traub, N., 2011. The Epstein-Barr virus BART microRNAs target the pro-apoptotic protein Bim. Virology 412, 392-400.

Martinez, J., Ok, S., Smith, S., Snoeck, K., Day, J.P., Jiggins, F.M., 2015. Should symbionts be nice or selfish? Antiviral effects of *Wolbachia* are costly but reproductive parasitism is not. PloS Pathogens 11, e1005021.

Masotti, A., Preckel, T., 2006. Analysis of small RNAs with the Agilent 2100 Bioanalyzer. Nature Methods 3, 658.

Matovu, E., Seebeck, T., Enyaru, J.C.K., Kaminsky, R., 2001. Drug resistance in *Trypanosoma brucei* spp., the causative agents of sleeping sickness in man and nagana in cattle. Microbes and Infection 3, 763-770.

Mauck, K., Bosque-Pérez, N.A., Eigenbrode, S.D., De Moraes, C.M., Mescher, M.C., 2012. Transmission mechanisms shape pathogen effects on host–vector interactions: evidence from plant viruses. Functional Ecology 26, 1162-1175.

McGeoch, D.J., Gatherer, D., 2005. Integrating reptilian herpesviruses into the family *Herpesviridae*. Journal of Virology 79, 725-731.

Medlock, J., Atkins, K.E., Thomas, D.N., Aksoy, S., Galvani, A.P., 2013. Evaluating paratransgenesis as a potential control strategy for African trypanosomiasis. PloS Neglected Tropical Diseases 7, e2374.

Mehrabadi, M., Hussain, M., Asgari, S., 2013. MicroRNAome of *Spodoptera frugiperda* cells (Sf9) and its alteration following baculovirus infection. Journal of General Virology 94, 1385-1397.

Mehrabadi, M., Hussain, M., Matindoost, L., Asgari, S., 2015. The baculovirus antiapoptotic p35 protein functions as an inhibitor of the host RNA interference antiviral response. Journal of Virology 89, 8182-8192.

Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55, 715-728.

Meister, G., 2013. Argonaute proteins: functional insights and emerging roles. Nature Reviews Genetics 14, 447-459.

Meister, G., Tuschl, T., 2004. Mechanisms of gene silencing by double-stranded RNA. Nature 431, 343-349.

Merkling, S.H., van Rij, R.P., 2013. Beyond RNAi: antiviral defense strategies in *Drosophila* and mosquito. Journal of Insect Physiology 59, 159-170.

Meyer, A., Holt, H.R., Selby, R., Guitian, J., 2016. Past and ongoing tsetse and animal trypanosomiasis control operations in five African countries: a systematic review. PloS Neglected Tropical Diseases 10, e0005247.

Michel, T., Reichhart, J.-M., Hoffmann, J.A., Royet, J., 2001. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature 414, 756-759.

Miesen, P., Joosten, J., van Rij, R.P., 2016. PIWIs go viral: arbovirus-derived piRNAs in vector mosquitoes. PloS Pathogens 12, e1006017.

Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A.M., Lim, B., Rigoutsos, I., 2006. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 126, 1203-1217.

Mocaski, E.S., 2007. Introduction: definition and classification of the human herpesviruses, comparative analysis of herpesvirus-common proteins. Biology, Therapy and Immunoprophylaxis Cambridge University Press, Cambridge, pp. 44–58.

Moreira, L.A., Iturbe-Ormaetxe, I., Jeffery, J.A., Lu, G., Pyke, A.T., Hedges, L.M., Rocha, B.C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L.E., Johnson, K.N., Kay, B.H., McGraw, E.A., van den Hurk, A.F., Ryan, P.A., O'Neill, S.L., 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *plasmodium*. Cell 139, 1268-1278.

Mutika, G.N., Marin, C., Parker, A.G., Boucias, D.G., Vreysen, M.J.B., Abd-Alla, A.M.M., 2012. Impact of salivary gland hypertrophy virus infection on the mating success of male *Glossina pallidipes*: consequences for the sterile insect technique. PloS One 7, e42188.

Nainu, F., Tanaka, Y., Shiratsuchi, A., Nakanishi, Y., 2015. Protection of insects against viral infection by apoptosis-dependent phagocytosis. Journal of Immunology 195, 5696-5706.

Naito, Y., Ui-Tei, K., 2012. siRNA design software for a target gene-specific RNA interference. Frontiers in Genetics 3, 102.

Nandety, R.S., Kuo, Y.-W., Nouri, S., Falk, B.W., 2015. Emerging strategies for RNA interference (RNAi) applications in insects. Bioengineered 6, 8-19.

Nayduch, D., Burrus, R.G., 2017. Flourishing in filth: house fly–microbe interactions across life history. Annals of the Entomological Society of America 110, 6-18.

O'Neill, S.L., Gooding, R.H., Aksoy, S., 1993. Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. Medical and Veterinary Entomology 7, 377-383.

O'Neill, S.L., Karr, T.L., 1990. Bidirectional incompatibility between conspecific populations of *Drosophila simulans*. Nature 348, 178-180.

Obbard, D.J., Dudas, G., 2014. The genetics of host–virus coevolution in invertebrates. Current Opinion in Virology 8, 73-78.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., Jiggins, F.M., 2009. The evolution of RNAi as a defence against viruses and transposable elements. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 99-115.

Obbard, D.J., Jiggins, F.M., Bradshaw, N.J., Little, T.J., 2011. Recent and recurrent selective sweeps of the antiviral RNAi gene argonaute-2 in three species of *Drosophila*. Molecular Biology and Evolution 28, 1043-1056.

Odindo, M.O., 1982. Incidence of salivary gland hypertrophy in field populations of the tsetse *Glossina pallidipes* on the South Kenyan coast. Insect Science and its Application 3, 59-64.

Odindo, M.O., Payne, C.C., Crook, N.E., Jarrett, P., 1986. Properties of a novel DNA Virus from the tsetse fly, *Glossina pallidipes*. Journal of General Virology 67, 527-536.

Okamura, K., Ishizuka, A., Siomi, H., Siomi, M.C., 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes & Development 18, 1655-1666.

Olson, K.E., Blair, C.D., 2015. Arbovirus-mosquito interactions: RNAi pathway. Current Opinion in Virology 15, 119-126.

Ormerod, W.E., 1986. Trypanosomiasis control and African rural development. Parasitology Today 2, 362.

Osborne, S.E., Iturbe-Ormaetxe, I., Brownlie, J.C., O'Neill, S.L., Johnson, K.N., 2012. Antiviral protection and the importance of *Wolbachia* density and tissue tropism in *Drosophila simulans*. Applied and Environmental Microbiology 78, 6922-6929.

Otieno, L.H., Kokwaro, E.D., Chimtawi, M., Onyango, P., 1980. Prevalence of enlarged salivary glands in wild populations of *Glossina pallidipes* in Kenya, with a note on the ultrastructure of the affected organ. Journal of Invertebrate Pathology 36, 113-118.

Ouma, J.O., Cummings, M.A., Jones, K.C., Krafsur, E.S., 2003. Characterization of microsatellite markers in the tsetse fly, *Glossina pallidipes* (Diptera: *Glossinidae*). Molecular Ecology Notes 3, 450-453.

Ouma, J.O., Marquez, J.G., Krafsur, E.S., 2006. New polymorphic microsatellites in *Glossina pallidipes* (Diptera: *Glossinidae*) and their cross-amplification in other tsetse fly taxa. Biochemical Genetics 44, 464-470.

Ouma, J.O., Marquez, J.G., Krafsur, E.S., 2007. Patterns of genetic diversity and differentiation in the tsetse fly *Glossina morsitans morsitans* Westwood populations in East and Southern Africa. Genetica 130, 139-151.

Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of taxonomy. Frontiers in Zoology 7, 16-16.

Pais, R., Lohs, C., Wu, Y., Wang, J., Aksoy, S., 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Applied and Environmental Microbiology 74, 5965-5974.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., Elgin, S.C.R., 2004. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. Science 303, 669-672.

Palmer, W.H., Varghese, F.S., van Rij, R.P., 2018. Natural variation in resistance to virus infection in Dipteran insects. Viruses 10, E118.

Pan, X., Zhou, G., Wu, J., Bian, G., Lu, P., Raikhel, A.S., Xi, Z., 2012. *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. Proceedings of the National Academy of Sciences 109, E23-E31.

Parker, J.S., Barford, D., 2006. Argonaute: a scaffold for the function of short regulatory RNAs. Trends in Biochemical Sciences 31, 622-630.

Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.-C., Burke, D.S., Calisher, C.H., Laughlin, C.A., Saif, L.J., Daszak, P., 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews 72, 457-470.

Peacock, L., Ferris, V., Bailey, M., Gibson, W., 2012. The influence of sex and fly species on the development of trypanosomes in tsetse flies. PloS Neglected Tropical Diseases 6, e1515.

Pell, P.E., Southern, D.I., 1976. A cytogenetical study of *Glossina fuscipes fuscipes* including a comparison of the polytene chromosome maps with those of *Glossina austeni* (Diptera, *Glossinidae*). Genetica 46, 511-528.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grässer, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., Russo, J.J., Ju, J., Randall, G., Lindenbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M., Tuschl, T., 2005. Identification of microRNAs of the herpesvirus family. Nature Methods 2, 269-276.

Picozzi, K., Fèvre, E., Odiit, M., Carrington, M., Eisler, M.C., Maudlin, I., Welburn, S.C., 2005. Sleeping sickness in Uganda: a thin line between two fatal diseases. British Medical JournaL 331, 1238-1241.

Poinar, G., Wassink, H.J.M., Leegwater-van der Linden, M.E., van der Geest, L.P., 1979. *Serratia marcescens* as a pathogen of tsetse flies. Acta Tropica 36, 223-227.

Politzar, H., Merot, P., Brandl, F.E., 1984. Experimental aerial release of sterile males of *Glossina palpalis gambiensis* and of *Glossina tachinoides* in a biological control operation. Revue d'elevage et de medecine veterinaire des pays tropicaux 37, 198-202.

Pollock, J.N., 1982. Training manual for tsetse control personnel. 2, 101pp.

Popham, H.J.R., Shelby, K.S., Brandt, S.L., Coudron, T.A., 2004. Potent virucidal activity in larval *Heliothis virescens* plasma against *Helicoverpa zea* single capsid nucleopolyhedrovirus. Journal of General Virology 85, 2255-2261.

Prasad, A.N., Brackney, D.E., Ebel, G.D., 2013. The role of innate immunity in conditioning mosquito susceptibility to West Nile virus. Viruses 5, 3142-3170.

Prompiboon, P., Lietze, V.-U., Denton, J.S.S., Geden, C.J., Steenberg, T., Boucias, D.G., 2010. *Musca domestica* salivary gland hypertrophy virus, a globally distributed insect virus that infects and sterilizes female houseflies. Applied and Environmental Microbiology 76, 994-998.

R-Core, T., 2015. R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria.

Raftery, M., Müller, A., Schõnrich, G., 2000. Herpesvirus homologues of cellular genes. Virus Genes 21, 65-75.

Ramirez, J.L., Short, S.M., Bahia, A.C., Saraiva, R.G., Dong, Y., Kang, S., Tripathi, A., Mlambo, G., Dimopoulos, G., 2014. *Chromobacterium Csp_P* reduces Malaria and Dengue infection in vector mosquitoes and has entomopathogenic and in vitro anti-pathogen activities. PloS Pathogens 10, e1004398.

Rancès, E., Ye, Y.H., Woolfit, M., McGraw, E.A., O'Neill, S.L., 2012. The relative importance of innate immune priming in *Wolbachia*-mediated Dengue interference. PloS Pathogens 8, e1002548.

Rechavi, O., 2014. "Guest list" or "black list"? heritable small RNAs as immunogenic memories. Trends in Cell Biology 24, 212-220.

Ren, X., Hoiczyk, E., Rasgon, J.L., 2008. Viral paratransgenesis in the Malaria vector *Anopheles gambiae*. PloS Pathogens 4, e1000135.

Roberts, J.C., Warren, R.B., Griffiths, C.E.M., Ross, K., 2013. Expression of microRNA-184 in keratinocytes represses argonaute 2. Journal of Cellular Physiology 228, 2314-2323.

Robinson, A.S., 2002. Mutations and their use in insect control. Mutation Research 511, 113-132.

Robinson, A.S., 2005. Genetic basis of the sterile insect technique, in: Dyck, V.A., Hendrichs, J., Robinson, A.S. (Eds.), Sterile Insect Technique. Springer Netherlands, pp. 95-114.

Robinson, C.M., Pfeiffer, J.K., 2014. Viruses and the microbiota. Annual Review of Virology 1, 55-69.

Rodriguez-Andres, J., Rani, S., Varjak, M., Chase-Topping, M.E., Beck, M.H., Ferguson, M.C., Schnettler, E., Fragkoudis, R., Barry, G., Merits, A., Fazakerley, J.K., Strand, M.R., Kohl, A., 2012. Phenoloxidase activity acts as a mosquito innate immune response against infection with Semliki forest virus. PloS Pathogens 8, e1002977.

Rodríguez, F., Oliver, J.L., Marín, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142, 485-501.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61, 539-542.

RStudio, T., 2015. RStudio: integrated development environment for R. RStudio, Inc. Boston, MA.

Sabin, L.R., Hanna, S.L., Cherry, S., 2010. Innate antiviral immunity in *Drosophila*. Current Opinion in Immunology 22, 4-9.

Sabin, L.R., Zheng, Q., Thekkat, P., Yang, J., Hannon, G.J., Gregory, B.D., Tudor, M., Cherry, S., 2013. Dicer-2 processes diverse viral RNA species. PloS One 8, e55458.

Saldaña, M.A., Etebari, K., Hart, C.E., Widen, S.G., Wood, T.G., Thangamani, S., Asgari, S., Hughes, G.L., 2017. Zika virus alters the microRNA expression profile and elicits an RNAi response in *Aedes aegypti* mosquitoes. PloS Neglected Tropical Diseases 11, e0005760.

Sang, R.C., Jura, W.G., Otieno, L.H., Mwangi, R.W., 1998. The effects of a DNA virus infection on the reproductive potential of female tsetse flies, *Glossina morsitans centralis* and *Glossina morsitans morsitans* (Diptera: *Glossinidae*). Memórias do Instituto Oswaldo Cruz 93, 861-864.

Sang, R.C., Jura, W.G.Z.O., Otieno, L.H., Mwangi, R.W., Ogaja, P., 1999. The effects of a tsetse DNA virus infection on the functions of the male accessory reproductive gland in the host fly *Glossina morsitans centralis* (Diptera; *Glossinidae*). Current Microbiology 38, 349-354.

Sang, R.C., Jura, W.G.Z.O., Otieno, L.H., Ogaja, P., 1996. Ultrastructural changes in the milk gland of tsetse *Glossina morsitans centralis* (Diptera; *Glossinidae*) female infected by a DNA virus. Journal of Invertebrate Pathology 68, 253-259.

Sang, R.C., Jura, W.G.Z.O., Otieno, L.H., Tukei, P.M., Mwangi, R.W., 1997. Effects of tsetse DNA virus infection on the survival of a host fly, *Glossina morsitans centralis* (Diptera; *Glossinidae*). Journal of Invertebrate Pathology 69, 253-260.

Sansone, Christine L., Cohen, J., Yasunaga, A., Xu, J., Osborn, G., Subramanian, H., Gold, B., Buchon, N., Cherry, S., 2015. Microbiota-dependent priming of antiviral intestinal immunity in *Drosophila*. Cell Host & Microbe 18, 571-581.

Saridaki, A., Bourtzis, K., 2010. *Wolbachia*: more than just a bug in insects genitals. Current Opinion in Microbiology 13, 67-72.

Sarkar, D., 2008. Lattice: multivariate data visualization with R. Springer, New York.

Scheel, T.K.H., Luna, J.M., Liniger, M., Nishiuchi, E., Rozen-Gagnon, K., Shlomai, A., Auray, G., Gerber, M., Fak, J., Keller, I., Bruggmann, R., Darnell, R.B., Ruggli, N., Rice, C.M., 2016. A broad RNA virus survey reveals both miRNA dependence and functional sequestration. Cell Host & Microbe 19, 409-423.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H., 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annual Review of Entomology 55, 421-438.

Schlötterer, C., 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma 109, 365-371.

Schneider, D.I., Garschall, K.I., Parker, A.G., Abd-Alla, A.M.M., Miller, W.J., 2013. Global *Wolbachia* prevalence, titer fluctuations and their potential of causing cytoplasmic incompatibilities in tsetse flies and hybrids of *Glossina morsitans* subgroup species. Journal of Invertebrate Pathology 112, S104-S115.

Schnettler, E., Donald, C.L., Human, S., Watson, M., Siu, R.W.C., McFarlane, M., Fazakerley, J.K., Kohl, A., Fragkoudis, R., 2013. Knockdown of piRNA pathway proteins results in enhanced Semliki forest virus production in mosquito cells. Journal of General Virology 94, 1680-1689.

Schofield, C.J., Kabayo, J.P., 2008. Trypanosomiasis vector control in Africa and Latin America. Parasites & Vectors 1, 24-24.

Schutze, M.K., Aketarawong, N., Amornsak, W., Armstrong, K.F., Augustinos, A.A., Barr, N.B., Bo, W., Bourtzis, K., Boykin, L., Cáceres, C.E., Cameron, S.L., Chapman, T.A., Chinvinijkul, S., Chomič, A., De Meyer, M., Drosopoulou, E., Englezou, A., Ekesi, S., Gariou-Papalexiou, A., Geib, S.M., Hailstones, D.L., Hasanuzzaman, M., Haymer, D.S., Hee, A.K.W., Hendrichs, J., Jessup, A.J., Ji, Q., Khamis, F.M., Krosch, M.N., Leblanc, L., Mahmood, K., Malacrida, A.R., Mavragani-Tsipidou, P., Mwatawala, M.W., Nishida, R., Ono, H., Reyes, J., Rubinoff, D.Z., San Jose, M., Shelly, T.E., Srikachar, S., Tan, K., Thanaphum, S., Haq, I.U., Vijaysegaran, S., Wee, S., Yesmin, F., Zacharopoulou, A., Clarke, A.R., 2015. Synonymization of key pest species within the *Bactrocera dorsalis* species complex (Diptera: *Tephritidae*): taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data. Systematic Entomology 40, 456-471.

Scott, J.G., Warren, W.C., Beukeboom, L.W., Bopp, D., Clark, A.G., Giers, S.D., Hediger, M., Jones, A.K., Kasai, S., Leichter, C.A., Li, M., Meisel, R.P., Minx, P., Murphy, T.D., Nelson, D.R., Reid, W.R., Rinkevich, F.D., Robertson, H.M., Sackton, T.B., Sattelle, D.B., Thibaud-Nissen, F., Tomlinson, C., van de Zande, L., Walden, K.K., Wilson, R.K., Liu, N., 2014. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. Genome Biology 15, 466.

Shackelton, L.A., Holmes, E.C., 2004. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. Trends in Microbiology 12, 458-465.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13, 2498-2504.

Shaw, M.K., Moloo, S.K., 1993. Virus-like particles in rickettsia within the midgut epithelial cells of *Glossina morsitans centralis* and *Glossina brevipalpis*. Journal of Invertebrate Pathology 61, 162-166.

Shelby, K.S., Popham, H.J.R., 2006. Plasma phenoloxidase of the larval tobacco budworm, *Heliothis virescens*, is virucidal. Journal of Insect Science 6, 13.

Shi, X., Ran, Z., Li, S., Yin, J., Zhong, J., 2016. The effect of microRNA bantam on baculovirus AcMNPV infection in vitro and in vivo. Viruses 8, 136.

Shi, Y., 2004. Caspase activation, inhibition, and reactivation: a mechanistic view. Protein Science : A Publication of the Protein Society 13, 1979-1987.

Shropshire, J.D., Bordenstein, S.R., 2016. Speciation by symbiosis: the microbiome and behavior. mBio 7, e1785-1715.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87, 651-701.

Singh, J., Singh, C.P., Bhavani, A., Nagaraju, J., 2010a. Discovering microRNAs from *Bombyx mori* nucleopolyhedrosis virus. Virology 407, 120-128.

Singh, R., Levitt, A.L., Rajotte, E.G., Holmes, E.C., Ostiguy, N., vanEngelsdorp, D., Lipkin, W.I., dePamphilis, C.W., Toth, A.L., Cox-Foster, D.L., 2010b. RNA viruses in Hymenopteran pollinators: evidence of inter-taxa Virus transmission via pollen and potential impact on non-Apis Hymenopteran species. PloS One 5, e14357.

Siva-Jothy, M.T., Moret, Y., Rolff, J., 2005. Insect immunity: an evolutionary ecology perspective, in: Simpson, S.J. (Ed.), Advances in Insect Physiology. Academic Press, pp. 1-48.

Skalsky, R.L., Cullen, B.R., 2010. Viruses, microRNAs, and host interactions. Annual Review of Microbiology 64, 123-141.

Slack, J., Arif, B.M., 2006. The baculoviruses occlusion-derived virus: virion structure and function, Advances in Virus Research. Academic Press, pp. 99-165.

Snyder, A.K., Deberry, J.W., Runyen-Janecky, L., Rio, R.V.M., 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: *Glossinidae*) symbionts. Proceedings of the Royal Society Series B: Biological Sciences 277, 2389-2397.

Solano, P., 2010. Population genetics as a tool to select tsetse control strategies: suppression or eradication of *Glossina palpalis gambiensis* in the Niayes of Senegal. PloS Neglected Tropical Diseases 4, e692.

Solano, P., Duvallet, G., Dumas, V., Cuisance, D., Cuny, G., 1997. Microsatellite markers for genetic population studies in *Glossina palpalis* (Diptera: *Glossinidae*). Acta Tropica 65, 175-180.

Solano, P., Salou, E., Rayaisse, J.-B., Ravel, S., Gimonneau, G., Traore, I., Bouyer, J., 2015. Do tsetse flies only feed on blood? Infection, Genetics and Evolution 36, 184-189.

Soumana, I.H., Tchicaya, B., Loriod, B., Rihet, P., Geiger, A., 2014a. Identification of overexpressed genes in *Sodalis glossinidius* inhabiting trypanosome-infected self-cured tsetse flies. Frontiers in Microbiology 5, 255.

Soumana, I.H., Tchicaya, B., Simo, G., Geiger, A., 2014b. Comparative gene expression of *Wigglesworthia* inhabiting non-infected and *Trypanosoma brucei gambiense*-infected *Glossina palpalis gambiensis* flies. Frontiers in Microbiology 5, 620.

Southern, D.I., Pell, P.E., 1974. Comparative analysis of the polytene chromosomes of *Glossina austeni* and *Glossina morsitans morsitans*. Chromosoma 47, 213-226.

Spath, J., 2000. Feeding patterns of three sympatric tsetse species (*Glossina* spp.) (Diptera: *Glossinidae*) in the preforest zone of Cote d'Ivoire. Acta Tropica 75, 109-118.

Stenvang, J., Petri, A., Lindow, M., Obad, S., Kauppinen, S., 2012. Inhibition of microRNA function by antimiR oligonucleotides. Silence 3, 1.

Steverding, D., 2008. The history of African trypanosomiasis. Parasites & Vectors 1, 3.

Strand, M.R., 2008. The insect cellular immune response. Insect Science 15, 1-14.

Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M., Ganem, D., 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. Nature 435, 682-686.

Swevers, L., Smagghe, G., 2012. Use of RNAi for control of insect crop pests, in: Smagghe, G., Diaz, I. (Eds.), Arthropod-plant interactions: novel insights and approaches for IPM. Springer Netherlands, Dordrecht, pp. 177-197.

Swevers, L., Vanden Broeck, J., Smagghe, G., 2013. The possible impact of persistent virus infection on the function of the RNAi machinery in insects: a hypothesis. Frontiers in Physiology 4, 319.

Symula, R.E., Alam, U., Brelsfoard, C., Wu, Y., Echodu, R., Okedi, L.M., Aksoy, S., Caccone, A., 2013. *Wolbachia* association with the tsetse fly, *Glossina fuscipes fuscipes*, reveals high levels of genetic diversity and complex evolutionary dynamics. BMC Evolutionary Biology 13, 31.

Tamang, D., Tseng, S.M., Huang, C.Y., Tsao, I.Y., Chou, S.Z., Higgs, S., Christensen, B.M., Chen, C.C., 2004. The use of a double subgenomic Sindbis virus expression system to study mosquito gene function: effects of antisense nucleotide number and duration of viral infection on gene silencing efficiency. Insect Molecular Biology 13, 595-602.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30, 2725-2729.

Tanji, T., Hu, X., Weber, A.N.R., Ip, Y.T., 2007. Toll and imd pathways synergistically activate an innate immune response in *Drosophila melanogaster*. Molecular and Cellular Biology 27, 4578-4588.

Taylor, D.J., Ballinger, M.J., Bowman, S.M., Bruenn, Jeremy A., 2013. Virus-host co-evolution under a modified nuclear genetic code. PeerJ 1, e50.

Teacher, A.G.F., Griffiths, D.J., 2011. HapStar: automated haplotype network layout and visualization. Molecular Ecology Resources 11, 151-153.

Teixeira, L., Ferreira, Á., Ashburner, M., 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PloS Biology 6, e1000002.

Teodoro, J.G., Branton, P.E., 1997. Regulation of apoptosis by viral gene products. Journal of Virology 71, 1739-1746.

Terradas, G., McGraw, E.A., 2017. *Wolbachia*-mediated virus blocking in the mosquito vector *Aedes aegypti*. Current Opinion in Insect Science 22, 37-44.

Thomas, P., Kenny, N., Eyles, D., Moreira, L.A., O'Neill, S.L., Asgari, S., 2011. Infection with the *wMel* and *wMelPop* strains of *Wolbachia* leads to higher levels of melanization in the hemolymph of *Drosophila melanogaster*, *Drosophila simulans* and *Aedes aegypti*. Developmental & Comparative Immunology 35, 360-365.

Tidona, C., Darai, G., 2000. Iridovirus homologues of cellular genes--implications for the molecular evolution of large DNA viruses. Virus Genes 21, 77-81.

Tomoyasu, Y., Miller, S.C., Tomita, S., Schoppmeier, M., Grossmann, D., Bucher, G., 2008. Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. Genome Biology 9, R10.

Tzou, P., De Gregorio, E., Lemaitre, B., 2002. How *Drosophila* combats microbial infection: a model to study innate immunity and host–pathogen interactions. Current Opinion in Microbiology 5, 102-110.

Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J.-M., Lemaitre, B., Hoffmann, J.A., Imler, J.-L., 2000. Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. Immunity 13, 737-748.

Ugelvig, L.V., Cremer, S., 2012. Effects of social immunity and unicoloniality on host–parasite interactions in invasive insect societies. Functional Ecology 26, 1300-1312.

Vallejo, C.R., Lee, J.A., Keesling, J.E., Geden, C.J., Lietze, V.-U., Boucias, D.G., 2013. A mathematic model that describes modes of MdSGHV transmission within house fly populations. Insects 4, 683-693.

van Belkum, A., Scherer, S., van Alphen, L., Verbrugh, H., 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiology and Molecular Biology Reviews 62, 275-293.

Van Den Abbeele, J., Caljon, G., De Ridder, K., De Baetselier, P., Coosemans, M., 2010. *Trypanosoma brucei* modifies the tsetse salivary composition, altering the fly feeding behavior that favors parasite transmission. PloS Pathogens 6, e1000926.

Van den Bossche, P., Rocque, S.d.L., Hendrickx, G., Bouyer, J., 2010. A changing environment and the epidemiology of tsetse-transmitted livestock trypanosomiasis. Trends in Parasitology 26, 236-243.

van Rij, R.P., 2008. Virus meets RNAi. EMBO Reports 9, 725-729.

van Rij, R.P., Saleh, M.C., Berry, B., Foo, C., Houk, A., Antoniewski, C., Andino, R., 2006. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. Genes & Development 20, 2985-2995.

Varjak, M., Maringer, K., Watson, M., Sreenu, V.B., Fredericks, A.C., Pondeville, E., Donald, C.L., Sterk, J., Kean, J., Vazeille, M., Failloux, A.-B., Kohl, A., Schnettler, E., 2017. *Aedes aegypti* Piwi4 is a noncanonical PIWI protein involved in antiviral responses. MSphere 2, e00144-00117.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Springer, New York Fourth Edition.

Vreysen, M., 2001. Principles of area-wide integrated tsetse fly control using the sterile insect technique. Médecine Tropicale : Revue du Corps de Santé Colonial 61, 397-411.

Vreysen, M.J.B., 2005. Monitoring sterile and wild insects in area-wide integrated pest management programmes, in: Dyck, V.A., Hendrichs, J., Robinson, A.S. (Eds.), Sterile insect technique: principles and practice in area-wide integrated pest management. Springer Netherlands, Dordrecht, pp. 325-361.

Vreysen, M.J.B., 2006. Prospects for area-wide integrated control of tsetse flies (Diptera: *Glossinidae*) and trypanosomosis in sub-Saharan Africa. Revista de la Sociedad Entomológica Argentina 65, 1-21.

Vreysen, M.J.B., Saleh, K.M., Ali, M.Y., Abdulla, A.M., Zhu, Z.-R., Juma, K.G., Dyck, V.A., Msangi, A.R., Mkonyi, P.A., Feldmann, H.U., 2000. *Glossina austeni* (Diptera: *Glossinidae*) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. Journal of Economic Entomology 93, 123-135.

Vreysen, M.J.B., Saleh, K.M., Lancelot, R., Bouyer, J., 2011. Factory tsetse flies must behave like wild flies: a prerequisite for the sterile insect technique. PloS Neglected Tropical Diseases 5, e907.

Vreysen, M.J.B., Seck, M.T., Sall, B., Bouyer, J., 2013. Tsetse flies: their biology and control using area-wide integrated pest management approaches. Journal of Invertebrate Pathology 112, S15-S25.

Walshe, D.P., Lehane, S.M., Lehane, M.J., Haines, L.R., 2009. Prolonged gene knockdown in the tsetse fly *Glossina* by feeding double stranded RNA. Insect Molecular Biology 18, 11-19.

Wamwiri, F., Alam, U., Thande, P., Aksoy, E., Ngure, R., Aksoy, S., Ouma, J., Murilla, G., 2013. *Wolbachia*, *Sodalis* and trypanosome co-infections in natural populations of *Glossina austeni* and *Glossina pallidipes*. Parasites & Vectors 6, 232.

Wang, J., Brelsfoard, C., Wu, Y., Aksoy, S., 2013a. Intercommunity effects on microbiome and GpSGHV density regulation in tsetse flies. Journal of invertebrate pathology 112, S32-S39.

Wang, J., Weiss, B.L., Aksoy, S., 2013b. Tsetse fly microbiota: form and function. Frontiers in Cellular and Infection Microbiology 3, 69.

Wang, J., Wu, Y., Yang, G., Aksoy, S., 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. Proceedings of the National Academy of Sciences of the United States of America 106, 12133-12138.

Wang, S.L., Hawkins, C.J., Yoo, S.J., Müller, H.A.J., Hay, B.A., 1999. The *Drosophila* caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID. Cell 98, 453-463.

Wang, X.-H., Aliyari, R., Li, W.-X., Li, H.-W., Kim, K., Carthew, R., Atkinson, P., Ding, S.-W., 2006. RNA interference directs innate immunity against viruses in adult *Drosophila*. Science 312, 452-454.

Wang, Y., Jehle, J.A., 2009. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. Journal of Invertebrate Pathology 101, 187-193.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B.M., Dong, Y., Jiang, H., Kanost, M.R., Koutsos, A.C., Levashina, E.A., Li, J., Ligoxygakis, P., MacCallum, R.M., Mayhew, G.F., Mendes, A., Michel, K., Osta, M.A., Paskewitz, S., Shin, S.W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D.W., Raikhel, A.S., Kafatos, F.C., Dimopoulos, G., Zdobnov, E.M., Christophides, G.K., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science 316, 1738-1743.

Weick, E.-M., Miska, E.A., 2014. piRNAs: from biogenesis to function. Development 141, 3458.

Weiss, B.L., Maltz, M., Aksoy, S., 2012. Obligate symbionts activate immune system development in the tsetse fly. Journal of Immunology (Baltimore, Md. : 1950) 188, 3395-3403.

Weiss, B.L., Wang, J., Aksoy, S., 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PloS Biology 9, e1000619.

Welburn, S.C., Maudlin, I., 1999. Tsetse-trypanosome interactions: rites of passage. Parasitology Today 15, 399-403.

Werren, J.H., Baldo, L., Clark, M.E., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature Reviews Microbiology 6, 741-751.

White, D.W., Beard, R.S., Barton, E.S., 2012. Immune modulation during latent herpesvirus infection. Immunological reviews 245, 189-208.

Whitnall, A.B.M., 1934. The trypanosome infections of *Glossina pallidipes* in the Umfolosi game reserve, Zululand. (Final Report.). The Onderstepoort Journal of Veterinary Science and Animal Industry 2, 7-21.

Wickham, H., 2009. Ggplot2: elegant graphics for data analysis. Springer, New York.

Wilke, A.B.B., Marrelli, M.T., 2015. Paratransgenesis: a promising new strategy for mosquito vector control. Parasites & Vectors 8, 342.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology 54, 844-851.

Wu, Y.-L., Wu, C.P., Liu, C.Y.Y., Hsu, P.W.-C., Wu, E.C., Chao, Y.-C., 2011. A non-coding RNA of insect HzNV-1 virus establishes latent viral infection through microRNA. Scientific Reports 1, 60.

Wyss, J.H., 2000. Screwworm eradication in the Americas. Annals of the New York Academy of Sciences 916, 186-193.

Xi, Z., Ramirez, J.L., Dimopoulos, G., 2008. The *Aedes aegypti* Toll pathway controls Dengue virus infection. PloS Pathogens 4, e1000098.

Xie, J., Vilchez, I., Mateos, M., 2010. *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. PloS One 5, e12149.

Xu, J., Grant, G., Sabin, L.R., Gordesky-Gold, B., Yasunaga, A., Tudor, M., Cherry, S., 2012. Transcriptional pausing controls a rapid antiviral innate immune response in *Drosophila*. Cell Host & Microbe 12, 531-543.

Yadav, S., Frazer, J., Banga, A., Pruitt, K., Harsh, S., Jaenike, J., Eleftherianos, I., 2018. Endosymbiont-based immunity in *Drosophila melanogaster* against parasitic nematode infection. PloS One 13, e0192183.

Yang, G., Yang, L., Zhao, Z., Wang, J., Zhang, X., 2012. Signature miRNAs involved in the innate immunity of invertebrates. PloS One 7, e39015.

Yang, Y.L., Li, X.M., 2000. The IAP family: endogenous caspase inhibitors with multiple biological activities. Cell Research 10, 169-177.

Zambon, R.A., Nandakumar, M., Vakharia, V.N., Wu, L.P., 2005. The Toll pathway is important for an antiviral response in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 102, 7257-7262.

Zambon, R.A., Vakharia, V.N., Wu, L.P., 2006. RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. Cellular Microbiology 8, 880-889.

Zhang, G., Hussain, M., O'Neill, S.L., Asgari, S., 2013a. *Wolbachia* uses a host microRNA to regulate transcripts of a methyltransferase, contributing to Dengue virus inhibition in *Aedes aegypti*. Proceedings of the National Academy of Sciences of the United States of America 110, 10276-10281.

Zhang, P., Wang, J., Lu, Y., Hu, Y., Xue, R., Cao, G., Gong, C., 2013b. Resistance of transgenic silkworm to BmNPV could be improved by silencing *ie-1* and *lef-1* genes. Gene Therapy 21, 81-88.

Zhao, P., Lu, Z., Strand, M.R., Jiang, H., 2011. Antiviral, antiparasitic, and cytotoxic effects of 5,6-dihydroxyindole (DHI), a reactive compound generated by phenoloxidase during insect immune response. Insect Biochemistry and Molecular Biology 41, 645-652.

List of abbreviations

AcMNPV:	*Autographa californica* multiple nucleopolyhedrovirus
Ago:	Argonaute
AMP:	antimicrobial peptides
AAT:	animal African trypanosomosis
ASFV:	African swine fever virus
AW-IPM:	area-wide integrated pest management
BmNPV:	*Bombyx mori* nucleopolyhedrovirus
BusuNPV:	*Buzura suppressaria* nucleopolyhedrovirus
CA/CC:	corpora-allata/cardiaca glands
Caspase:	cysteine-dependent aspartate-specific protease
CHIKV:	chikungunya virus
CI:	cytoplasmic incompatibility
CIV/IIV-6:	Chilo/invertebrate iridescent virus
CLIP:	CLIP-domain Serine Protease
COI:	cytochrome oxidase 1
CpGV:	*Cydia pomonella* granulovirus
CTL:	C-type lectin
DAMM:	death associated molecule related to Mch2
DCP-1:	death caspase protein 1
Dcr:	Dicer
DCV:	*Drosophila* C virus
DECAY:	death executioner caspase-related to Apopain/Yama
DENV:	dengue virus
DIAP1:	*Drosophila* inhibitor of apoptosis protein 1
DREAM:	downstream regulatory element-antagonist modulator
DREDD:	death-related ced-3/Nedd2

DrICE: *Drosophila* interleukin-1β-converting enzyme
DRONC: death regulator Nedd2-like caspase
Dscam1: Down Syndrome cell adhesion molecule-1
Eater: pathogen pattern-recognition receptor Eater
EBV: Epstein-Barr virus
EppoNPV: *Epiphyas postvittana* nucleopolyhedrovirus
FHV: Flock house virus
GO: Gene ontology
GpSGHV: *Glossina pallidipes* salivary gland hypertrophy virus
HAT: human African trypanosomosis
HCV: human cytomegalovirus
HearNPV: *Helicoverpa armigera* single nucleopolyhedrovirus
HGT: horizontal gene transfer
HSV-1: Herpes simplex virus 1
HVS: Herpesvirus saimiri
HzNV-1: *Heliothis zea* nudivirus-1
IAP: inhibitor of apoptosis
IAPV: Israeli Acute Paralysis Virus
IIID: insect innate immunity database
IIV-6: invertebrate iridovirus
Imd: immune deficiency
IPCL: Insect Pest Control Laboratory
ITS1: internal transcribed space 1
KSV: Kaposi's sarcoma-associated herpesvirus
LEF: early/late gene expression factor
Loqs: Loquacious
MdSGHV: *Musca domestica* salivary gland hypertrophy virus
miRNA: microRNA

NGS:	next generating sequencing
NOS:	nitric oxide synthase
OpMNPV:	*Orgyia pseudotsugata* multiple nucleopolyhedrovirus
PIF:	*per os* infectivity factors
piRNA:	Piwi-interacting RNA
PPO:	prophenoloxidase
REL1:	relish protein
RISC:	RNA induced silencing complex
RNAi:	RNA interference
ROS:	reactive oxygen species
RT-qPCR:	reverse transcriptase quantitative PCR
SeMNPV:	*Spodoptera exigua* multiple nucleopolyhedrovirus
Serpin:	serine protease inhibitors
SFV:	Semliki Forest virus
SGH:	salivary gland hypertrophy
SINV:	Sindbis virus
siRNA:	short interfering RNA
SIT:	sterile insect technique
SpexNPV:	*Spodoptera exempta* nucleopolyhedrovirus
Spn-E/hls:	spindle-E/Homeless
Spz1A:	spätzle 1A
sRNA:	small RNA
STRICA:	serine/Threonine-rich caspase-A
TEP:	thioester-containing protein
TPM:	tag per million
vsiRNAs:	virus-derived siRNAs
VNTRs:	variable number tandem repeat regions
WNV:	West Nile virus

Summary

Tsetse vector control via the sterile insect technique (SIT) requires mass production of sterile male insects for subsequent release into target populations of tsetse species. However, SIT campaigns against some tsetse species (e.g., *G. pallidipes*) in regions such as the southern Rift Valley of Ethiopia have been seriously challenged by infections of colonised flies with the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV; *Hytrosaviridae* family). GpSGHV causes salivary gland hypertrophy (SGH), a syndrome that also leads to reproductive dysfunctions and hence colony collapse of tsetse fly colonies. GpSGHV exhibits a dual (vertical/horizontal) transmission mode. The vertical transmission of the virus from infected mothers to the offspring, transovum and via milk glands, allows the virus to invade and spread through tsetse colonies and may ensure the maintenance of (persistent or latent) asymptomatic infections in the colonies. Secondly, the horizontal transmission of the virus from infected to uninfected flies via salivary secretions during membrane feeding of the colonies may assist the virus to rapidly spread through tsetse colonies. The other member of the *Hytrosaviridae* family, *Musca domestica* SGHV (MdSGHV), infects the filth-feeding housefly, exclusively symptomatically. MdSGHV also causes complete shutdown of oogenesis which prevents vertical transmission of this virus. The question of how GpSGHV remains largely asymptomatic in most tsetse species, and the conditions and mechanism(s) that trigger SGH outbreaks in the mass-reared *G. pallidipes* are yet to be resolved. This research was designed to address gaps in the knowledge of the ecology and molecular modes of action of GpSGHV during its infection in tsetse flies.

This PhD thesis has explored how the genetics and ecology of SGHVs and their hosts (housefly and tsetse fly) may have influenced the evolution, pathogenesis and transmission modes of SGHVs. To evade host's immune surveillance, viruses may evolutionary adopt multiple transmission modes to increase chances of persistent infections and facilitate their dispersion. Investigations on the coevolution of the SGHVs and their hosts suggested that, to fight off MdSGHV infection, the housefly has expanded its repertoire of immunity-related genes, such as apoptotic and RNAi genes. However, as a counter-measure, the MdSGHV possesses anti-apoptotic genes

and also appears to have adopted an immediate and rapid replication, thereby causing symptomatic infections. On the other hand, GpSGHV seems to have recruited cellular genes from ancestral host(s) into its genome as a "camouflage" strategy to escape the host's immune surveillance and that perhaps assist the virus in establishing persistent infections in the host. In addition, unlike in the housefly, there also seems to be an interaction between the GpSGHV and tsetse endosymbionts, which may also influence GpSGHV pathogenesis and transmission modes in the tsetse fly. These ecological and other differences between tsetse fly and housefly may influence the outcome of the SGHVs infections, resulting in the occurrence of only symptomatic infections in the housefly and the occasional switch from asymptomatic to symptomatic in particular tsetse fly colonies.

In this thesis, a molecular multi-marker approach was developed for correctly identifying tsetse species and sub-species, and also for determining the genetic diversity and the prevalence of GpSGHV haplotypes in various tsetse fly species. The latter also allowed tracing the ancestral host(s) of these viral strains in natural tsetse populations. These molecular tools are considered a major improvement in the hitherto classical identification of morphologically similar tsetse species, particularly because tsetse is a complex of many closely related species. Using the integration of nuclear markers/microsatellites, *Wolbachia* diagnosis and mitochondrial gene sequencing coupled with data on the geographical origins of tsetse species, it was possible to cluster *Glossina* specimens into respective taxonomic groups (taxons). Further, populations/haplotypes of the same tsetse species could be clearly distinguished. Phylogenetic relatedness of tsetse species allowed prediction of the ancestral host(s) of GpSGHV and its evolution and spread in different tsetse fly species. Together, these new identification tools aid in the development of integrated strategies to manage viral pathogens infecting tsetse species in new and existing colonies.

This thesis has shown that tsetse species belonging to the Morsitans group had the highest relative prevalence of GpSGHV, with *G. pallidipes* presenting the highest prevalence followed by the *G. m. morsitans*, compared to for instance *G. swynnertoni* and *G. austeni* with low prevalence. In addition to the high virus prevalence in *G. pallidipes*, multiple virus haplotypes were noted in this species.

Unlike other tsetse species belonging to the Morsitans group, *G. swynnertoni* appears to be the only *Glossina* species infected with a distinct GpSGHV haplotype. Representatives of the species belonging to the Fusca and Palpalis groups presented low virus prevalence and were all infected with a single virus haplotype. This single GpSGHV haplotype, which is similar to the Ugandan GpSGHV strain, infected most of the other tsetse species that were analysed in the study. In all, these data suggested that the species belonging to the Palpalis and Fusca groups are probably the original host(s) of GpSGHV. It is possible that in these species, the virus has evolutionarily reached a stable and equilibrium infection status as evidenced by their infection by a single GpSGHV haplotype. This is in contrast with the species belonging to the Morsitans group, in which the virus may still be adapting in new host(s), thus leading to evolution of novel virus haplotypes and probably contributing to SGH outbreaks in the colonised *G. pallidipes*.

In this thesis, the hypothesis was tested that certain host-mediated defence mechanisms such as the RNAi machinery may contribute to the maintenance of asymptomatic GpSGHV infections in most of the tsetse species. Although, all *AGO* orthologs were identified in the *G. pallidipes* genome, the absence of an ortholog to a key RNAi pathway gene (*DCR1*), and the presence of a single key ortholog of the *DCR2* gene suggested that the latter (*DCR2*) might function in processing both the short interfering RNA (siRNAs) and the microRNA (miRNAs) in this species. The research also revealed that in *G. pallidipes*, the genes regulating the siRNA and miRNA pathways were upregulated in asymptomatically infected flies but downregulated in symptomatically infected flies. The results of this thesis topic provided the first evidence that the RNAi machinery is a central anti-GpSGHV defence mechanism in tsetse flies. In addition, although the RNAi machinery seems functional during asymptomatically-infected *G. pallidipes*, the pathway is downregulated in symptomatic infections. However, this research did not provide any evidence that SGHVs encode viral suppressors of RNAi (VSR), as a mechanism to impair the host's RNAi machinery.

Certain nuclear-replicating DNA viruses have evolved strategies to exploit the host's miRNA pathway, by for instance encoding their own miRNAs or regulating the host miRNA expression profile. These miRNAs aid in regulating the host's or viral gene

expression. In this thesis, several host and GpSGHV-encoded miRNAs were detected and identified, and their potential host target genes were predicted *in silico*. These observations provided the first evidence that GpSGHV miRNAs may regulate the expression of immunity genes in tsetse or modulate the expression of certain viral genes. They may therefore influence the outcome of virus infections and perhaps promote latent (asymptomatic) infection or assist in the evasion of the host immune system and thereby enhance establishment of symptomatic infections and SGH outbreaks.

In conclusion, the results presented in this thesis support the potential role of small RNAs in the regulation of (a)symptomatic infections in the tsetse flies. The outcomes assist in the understanding of SGH outbreaks in *G. pallidipes* mass rearing facilities. The data presented in this thesis provide a promising basis for future development of novel and/or complementary strategies to control the competence of tsetse fly in transmission of GpSGHV in tsetse fly mass production facilities. The prevention of SGH outbreaks in mass-reared tsetse flies is crucial to improve tsetse colony productivity to allow a wider implementation of SIT-mediated control of several tsetse species in sub-Saharan Africa. This would be a huge immediate benefit especially in the eradication of the *G. pallidipes*, which infests the southern Rift Valley of Ethiopia and has been a target for SIT campaigns in the last decade.

List of publications

1. Kariithi, H.M., İnce, İ.A., Boeren, S., Murungi, E.K., **Meki, I.K**., Otieno, E.A., Nyanjom, S.R.G., van Oers, M.M., Vlak, J.M., Abd-Alla, A.M.M., 2016. Comparative analysis of salivary gland proteomes of two *Glossina* species that exhibit differential hytrosavirus pathologies. Frontiers in Microbiology 7, 89.

2. Kariithi, H.M., **Meki, I.K**., Boucias, D.G., Abd-Alla, A.M.M., 2017. Hytrosaviruses: current status and perspective. Current Opinion in Insect Science 22, 71-78.

3. **Meki, I.K**., Kariithi, H.M., Ahmadi, M., Parker, A.G., Vreysen M.J.B., Vlak, J.M., van Oers, M.M. and Abd-Alla A.M.M. Hytrosavirus genetic diversity and eco-regional spread in *Glossina* species. BMC Microbiology, (*in press*).

4. **Meki, I.K**., Kariithi, H.M., Parker, A.G., Vreysen M.J.B., Ros, V.I.D., Vlak, J.M., van Oers, M.M. and Abd-Alla A.M.M. RNA interference-based antiviral immune response against the salivary gland hypertrophy virus in *Glossina pallidipes*. BMC Microbiology, (*in press*).

5. **Meki, I.K**., İnce, İ., Kariithi, H.M., Boucias, D.G., Orhan O., Parker, A.G., Vlak, J.M., van Oers, M.M. and Abd-Alla A.M.M. Expression profile of *Glossina pallidipes* microRNAs during symptomatic and asymptomatic infection with *Glossina pallidipes* salivary gland hypertrophy virus (hytrosavirus). Frontiers in Microbiology, (*DOI: 10.3389/fmicb.2018.02037*).

6. Augustinos, A.A.*, **Meki, I.K**.*, Demirbas-Uzel, G., Ouédraogo, G.M.S., Saridaki, A., Tsiamis, G., Parker, A., Abd-Alla, A., Bourtzis, K. Nuclear and *Wolbachia*-based multi-marker approach for the rapid and accurate identification of tsetse species. BMC Microbiology, (*in press*).
   *These authors contributed equally to this paper

7. Kariithi, HM, Boucias, D.G., Murungi, E.K., **Meki, I.K**., Demirbas-Uzel, G., van Oers, M.M., Vreysen, M.J.B., Abd-Alla, A.M.M. and Vlak, J.M. Coevolution of hytrosaviruses and host immune responses. BMC Microbiology, (*in press*).

8. Kariithi, H.M.*, **Meki, I.K**.*, Schneider, D.*, De Vooght, L.*, Khamis, F.M.*, Geiger, A.*, Demirbaş-Uzel, G.*, Vlak, J.M., Ince, I.A., Kelm, S., Njiokou, F., Wamwiri, F.N., Malele, I., Weiss, B.L. and Abd-Alla, A.M.M. Enhancing vector refractoriness to trypanosome infection: achievements, challenges and perspectives. BMC Microbiology, (*in press*).
   *These authors contributed equally to this paper

Conference Proceedings

1. **Meki, I.K**., Kariithi, H.M, Vlak, J.M., van Oers, M.M., Parker, A.G., Vreysen M.J.B. and Abd-Alla, A.M.M. Hytrosavirus genetic diversity and eco-regional spread in *Glossina* species. In: Abstract Book of the 50th Annual Meeting of the Society for Invertebrate Pathology, San Diego, California, 13-17 August **2017**, p. 35.
2. Augustinos, A.A., **Meki, I.K**., Saridaki, A., Demirbas-Uzel, G., Tsiamis, G., van Oers, M.M., Vreysen M.J.B., Parker, A., Abd-Alla, A., and Bourtzis, K. Contributing to the resolution of taxonomic puzzles: multiple molecular tools and development of protocols for the accurate identification of tsetse species. In: Abstract Book of the 3rd FAO–IAEA International Conference on Area-wide Management of Insect Pests: Integrating the Sterile Insect and Related Nuclear and Other Techniques, Vienna, Austria, 22-26 May **2017**, p. 271.
3. **Meki, I.K**., Ince, I.A., Kariithi, H.M, Boucias, D.G., Vlak, J.M., van Oers, M.M. and Abd-Alla, A.M.M. Interactions between the salivary gland hypertrophy virus and its host immune system. In: Abstract Book of the 49th Annual Meeting of the Society for Invertebrate Pathology, Tours, France, 24-28 July **2016**, p. 70.
4. **Meki, I.K**., Kariithi, H.M, Vlak, J.M., Abd-Alla, A.M.M. and van Oers, M.M. Role of dsRNA-mediated mechanism in the establishment of "latent" *Glossina* hytrosavirus infections in the tsetse fly. In: Abstract Book of the 48th Annual Meeting of the Society for Invertebrate Pathology, University of British Colombia, Vancouver Canada, 9-13 August **2015**, p. 44.

About the author

Irene Kasindi Meki was born on the 27th July 1989 in Mutomo, Eastern Kenya. She attended Mutomo Girls' High school for her secondary level education and qualified for enrolment at Moi University, Eldoret, Kenya, in 2007 where she obtained a BSc degree in Agricultural Biotechnology in 2011. In September 2012, she started an MSc study in Cellular and Molecular Biotechnology at Wageningen University in the Netherlands, sponsored by the Netherlands Fellowship Program (Nuffic). During her MSc studies, Irene was offered the opportunity to undertake an MSc thesis at the Laboratory of Virology, where she investigated baculovirus-induced behavioural manipulation of the beet armyworm, *Spodoptera exigua*. Later on, in March 2014, she was offered an MSc internship at the FAO/IAEA Biotechnology Laboratories in Seibersdorf, Austria. There she investigated the molecular mechanism(s) that underlay tsetse-*Glossinavirus* interactions in tsetse mass rearing facilities. The preliminary findings of her internship research warranted further investigations, and after she completed her Masters in September 2014 in Wageningen, Irene was offered the opportunity to extend her research into a PhD study. In March 2015, she enrolled for a sandwich PhD program between Wageningen University (supervision by Prof Dr. M. M. van Oers and Prof Dr. J. M. Vlak), the FAO/IAEA Laboratories in Seibersdorf (supervision by Prof Dr. A. M. M. Abd-Alla), and the Kenya Agricultural and Livestock Research Organization (KALRO) in Nairobi (supervision by Dr. H. M. Kariithi). Currently, Irene is pursuing leads to prospective postdoc opportunities in the field of Molecular Biology, especially in research involving the discoveries of novel strategies to mitigate vector-borne diseases.

Acknowledgements

The achievements and completion of this PhD thesis give me immense pleasure, satisfaction and a unique sense of accomplishment, which would not have been possible without the unconditional support of many people and institutions. It would have been a great pleasure to mention all the names of all those who assisted me in this difficult venture. Due to space constraint, this is difficult but does not in any way go unacknowledged by me.

First of all, I would like to thank Wageningen University for offering me the opportunity and academic support to carry out my PhD research at the Laboratory of Virology. Secondly, I acknowledge the overwhelming support by the International Atomic Energy Agency for the unlimited resources and financial support provided to carry out most of my PhD research at the Joint FAO/IAEA Insect Pest Control Laboratories in Seibersdorf, Austria. I am grateful to the Dr Judith Zwartz Foundation, Wageningen, The Netherlands, for sponsoring the printing of this thesis.

The first and foremost among the people whom I express my sincere and humble gratitude are my supervisors for their continuous support of my PhD study and the related research, for their patience, motivation and mentorship. Their guidance was instrumental in shaping my personality and professional attitude during my PhD study. First, thanks to my promotor Prof. Dr. Monique van Oers, the Chair of the Laboratory of Virology in Wageningen for your confidence in my capabilities and nurturing me in many ways throughout my studies. I could not have had a better advisor and mentor for my PhD study. I also thank my second promotor, Prof. Adly Abd-Alla, who always made himself available to discuss my daily research activities despite his busy schedules. You have been a mentor to me both scientifically and socially. Besides my two promotors, I thank the rest of my thesis supervisory team: Prof. Dr. Just Vlak and Dr. Henry Kariithi, for their insightful scholarly comments, consistent encouragement, and poking questions, which were incentives for me to widen my research scope and perspectives.

Dear Prof Just, I count myself lucky to have had you in my thesis supervisory team. At first, your towering height compared to mine was intimidating, to the extent that I was afraid you may not “hear” me, especially when we conversed along the

hallways and corridors of the great Radix building of Wageningen. Upon getting acquainted with you however, your physique equally matched your knowledge and richness in science, which greatly shaped the course of my research and sharpened the findings thereof. Whenever I hit a wall and not sure of how to proceed, you made everything seem possible, easy and achievable, even when I nearly lost hope.

Dear Dr. Kariithi, it was a happy coincidence to have met you for the first time in June 2010 during my short internship at your institute (KALRO). I was impressed by a talk you presented, and it was then that I was inspired to apply for an MSc program at Wageningen University, where you also studied at the time. Since 2010, you have been a mentor and a sincere friend to me, and you motivated me during every step of the way.

This thesis would not have come to a successful completion without the support, encouragements, and the homely environment I received from all my colleagues and other staff at the IAEA laboratories, Seibersdorf. To the IPCL Head, Dr. Marc Vreysen, thank you for your support in technical matters at the IPCL facilities that made my work easier, and your scientific inputs during the lab meetings and seminars that you often arranged. Special thanks to Mr. Andrew Parker for your inspiration and guidance, particularly in the technicalities of experimental designs, data analyses and interpretation of the findings. I sincerely thank Prof. Kostas Bourtzis for his great support and fruitful discussions during my studies. To Hasim Mohammed, Carmen Marin, Henry Adun, Lucia Duran, Edgardo Lapiz and Florence Maxwel, thank you for your unwavering support and technical hand-on training on tsetse rearing. I would like to acknowledge my fellow PhD student at IAEA, Guler Dermibas; thank you for your encouragement and moral support, especially during the late working evenings. To all my friends in Vienna, thanks a lot for the nice moments we shared, which often inspired interesting discussions. The outings, and dinners, we held together were a source of inspiration to me. You made Vienna feel like a home to me.

I also would like to acknowledge all the colleagues at the Laboratory of Virology. In particular, thanks to Dr. Vera Ros who taught me several important technical aspects during my MSc thesis that I later applied during my PhD study. I would like to thank Marleen Henkens for always being there to assist with the office-related issues in the

Virology group. I would like to thank Giel Goertz, for his assistance in the analyses of some of my NGS data through skype discussions and for agreeing to be my paranymph during my defence.

I also acknowledge the Kenya Agricultural and Livestock Research Organisation (KALRO) and South Eastern Kenya University (SEKU) for the support they provided during my studies. I particularly thank Prof Dr. Reuben Muasya, the Deputy Vice Chancellor Finance, Planning and Development at SEKU for supporting me during my studies.

I would like to acknowledge the Kenyan community at Wageningen University. The dinners and the get-togethers that we have had made Wageningen a home away from my home. In particular, I would like to thank Hellen Kimanthi for agreeing to be my second paranymph during my thesis defence.

Last but not the least, to my family, I have no words to express my gratitude for your unceasing support, encouragement and attention. I acknowledge my mum (Patricia Meki), who has always supported and encouraged me at every phase of my personal and academic life, and longed to see this achievement come true. I am much indebted to my brothers (Mutinda, Muimi, Muthoka, and Peter) and my sister (Kanza) who supported me in every possible way to see the completion of my studies. To my dear uncle Robert, I am particularly grateful to you because you supported me all the way during my studies; I would not have come this far without you in my life. Thanks to my aunt, Rhoda and to all my cousins (Vivian, Erica, Eva, Emily, Nicodemus, Mutua, John, Wambua, Kasembi, Jeniffer and Fidelis), for their support and encouragement.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

- Understanding the factors that determine the outcome of *Glossina* virus infections in tsetse fly production facilities

Writing of project proposal (4.5 ECTS)

- Mechanism(s) of GpSGHV infections: prerequisite for *trypanosomosis* control

Post-graduate courses (1.8 ECTS)

- Minke informatics Bio-Linux training; introduction to R for statistical analysis; IAEA/FAO (2016)
- Advanced statistics for food and agriculture: design and implementation of experiments including focus on advanced experiments; IAEA/FAO (2017)

Laboratory training and working visits (2.4 ECTS)

- Impact of anti-microbial peptides on tsetse symbionts and other associated organisms; the Institute national des Sciences Appliquées de Lyon (INSA) (2016)

Competence strengthening / skills courses (1.8 ECTS)

- PhD Competence assessment; WGS (2018)
- Career perspectives; WGS (2018)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.8 ECTS)

- PE&RC Weekend (2015)
- PE&RC Day (2015)
- PE&RC Weekend, last years (2018)

Discussion groups / local seminars / other scientific meetings (4.7 ECTS)

- IAEA –Nuclear sciences and applications workshop (2015)
- Joint FAO /IAEA RCM – annual conference and workshops (2017)
- Work discussions and seminars at the insect pest control laboratories, Seibersdorf, Austria (2015-2017)
- Virology group seminars (2018)
- Insect viruses group seminars (2018)

International symposia, workshops and conferences (10 ECTS)

- 48th Annual Meeting of Society for Invertebrate Pathology, Virus Division, University of British Colombia, Vancouver Canada: Oral presentation "Role of dsRNA-Mediated Mechanism in the Establishment of "Latent" Glossina Hytrosavirus Infections in the Tsetse fly" (2015)
- 49th Annual Meeting of Society for Invertebrate Pathology, virus division yours; oral presentation; France (2016)
- Third FAO–IAEA International Conference on Area-wide Management of Insect Pests: integrating the sterile insect and related nuclear and other techniques; poster presentation; Vienna, Austria (2017)
- 50th Annual Meeting of Society for Invertebrate Pathology, Division of Beneficial Invertebrates; oral presentation; San Diego, California (2017)

The research presented in this thesis was financed by the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Vienna, Austria (CRP No. D42015) and Wageningen University Sandwich PhD Program. The thesis publication was partially funded by the Dr Judith Zwartz Foundation, the Netherlands.

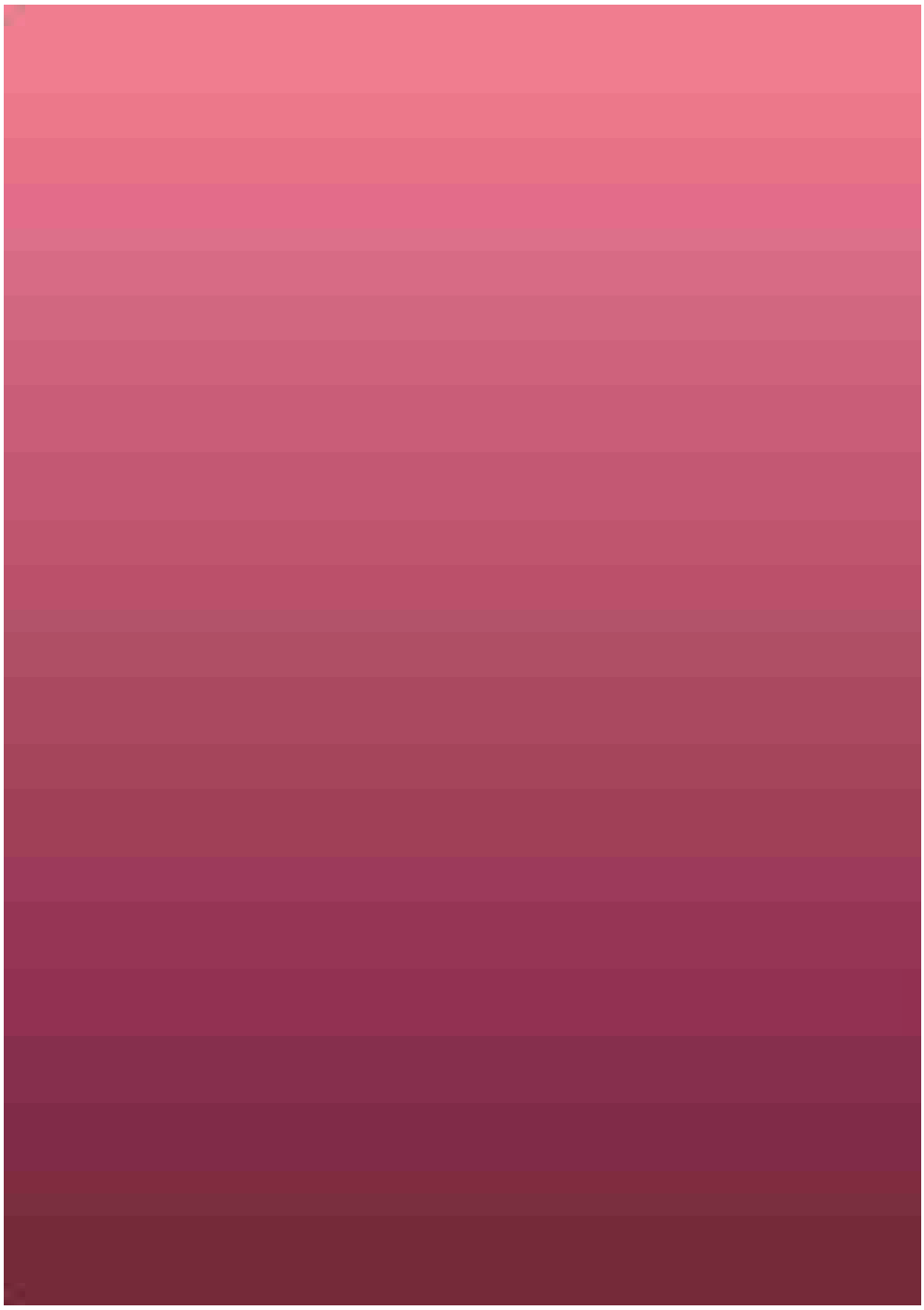
Cover design

Tsetse fly photo: courtesy of Dr. Geoffrey Attardo, University of California, Davis, CA, 95616, USA.

Cover design: A tsetse fly infected by virus (illustrated by the orange particles) sits on the tsetse infested sub-Saharan African countries (unshaded countries).

Cover layout and graphics design: Irene K. Meki.

Printed by: Digiforce ‖ ProefschriftMaken.



Propositions

1. Small RNAs are involved in the regulation of overt versus covert infection of tsetse with salivary gland hypertrophy virus.
   (this thesis)

2. Molecular-based taxonomy of tsetse species is superior to conventional taxonomy.
   (this thesis)

3. The interplay between the cellular (nuclear/mitochondria/chloroplast) and microbiome genomes shaped the biological history of eukaryotes.

4. The apoptotic pathway is crucial for preventing cardiovascular and neurological disorders.

5. Virus eradication is not so much a technical but rather a societal issue.

6. Unless they are triggered, some viruses are as innocent as a gun in the closet.

7. A female PhD holder is overtly admired but often covertly disliked.

Propositions belonging to the PhD thesis

Hytrosavirus in tsetse flies: Phylogeography and molecular mode of action

Irene K. Meki

Wageningen, October 8, 2018