IDENTIFICATION AND CHARACTERIZATION OF STAGE-SPECIFIC AND SEX-SPECIFIC MICRO-RNAS OF *Glossina pallidipes* TSETSE FLY, VECTOR OF AFRICAN TRYPANOSOMIASIS

CAREEN NAITORE IKUNYUA

MASTER OF SCIENCE

(Molecular Biology & Bioinformatics)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2022

Identification and characterization of stage-specific and sex-specific micro-rnas of *Glossina pallidipes* tsetse fly, vector of African trypanosomiasis

Careen Naitore Ikunyua

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Molecular Biology and Bioinformatics of the Jomo Kenyatta University of Agriculture and Technology

2022

DECLARATION

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

Signature …… Date … ……

Careen Naitore Ikunyua

This thesis has been submitted for examination with our approval as university supervisors.

Signature …… Date … ……

Dr. Joel L. Bargul, PhD
JKUAT, Kenya

Signature …… Date … ……

Dr.JandouweVillinger, PhD
ICIPE, Kenya

DEDICATION

I dedicate this thesis to my loving parents

Elias Mutugwa Ikunyua and Charity Gakii Mugania

Who laid the foundation for my accomplishment?

ACKNOWLEDGMENTS

Firstly, I would like to thank God, without whom I would never have come so far in life. I would like to express my sincere gratitude to my supervisors, Dr. Joel L. Bargul, Dr. Jandouwe Villinger, and Dr. Daniel Masiga, who took me as their Master's student, always showing understanding, provided the mentorship, inspiration and vast knowledge in tsetse genetics. I would like to thank Mr. James Kabii for providing me with guidance and assistance in the laboratory. I would like to thank Jomo Kenyatta University of Agriculture and Technology where I am registered and International Centre of Insect Physiology and Ecology (*icipe*, Kenya), where I carried out my research studies. I would also like to give a special thanks to Dr. Caleb Kibet, one of my supervisors for his tremendous support and guidance during my bioinformatics analysis. I would like to acknowledge my colleagues and bioinformatics team at *icipe* who there were constantly to consult. Special thanks go to Prof Allan Christoffels of the South African National Bioinformatics Institute for financial support and guidance. Finally, to my family, special thanks for their understanding and being the cornerstone of my support to the completion of this thesis.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AAT	African Animal Trypanosomasis
AGO2	Argonuate 2 Protein
CP4H	Collagen Prolyl - 4 - Hydroxyls
DGR8	Di George Syndrome Critical Region 8
DNA	Deoxyribonucleic acid
EPK	Eukaryotic Protein Kinase
GO	Gene Ontology
ITS	Internal Transcribed Spacer
KALRO	Kenya Agricultural and Livestock Research Organization
KEGG	Kyoto Encyclopedia of Genes and Genomes
miRNA	Micro Ribonucleic Acid
mRNA	Messenger Ribonucleic Acid
ncRNA	Non-Coding RNA
NGS:	Next Generation Sequencing
nt	Nucleotides
PATTEC	Pan-African Tsetse and Trypanosomiasis Eradication Campaign
RNA	Ribonucleic Acid
RPM	Reads per Million
***r*RNA**	Ribosomal Ribonucleic Acid
***s*RNA**	Small (non-coding) RNA
TRBP	TAR – RNA Binding Protein
***t*RNA**:	Transfer Ribonucleic Acid
WHO	World Health Organisation

ABSTRACT

The field of miRNA biology is tremendously growing, moreover, their roles of regulating a wide array of cell function is extensive. The importance of miRNAs in development has become nearly ubiquitous, with miRNA contributing to development of most cells and organs. Although miRNAs are clearly interwoven into known regulatory networks that control cell development, the specific modalities by which they intersect are often quite distinct in different organism. In insect species miRNA involved in development has been well study in *Drosphila spp* and *Anopheles spp.* However, information concerning possible developmental role of miRNAs in tsetse is limited. Hence, the aim of the study is to identify and characterize miRNA genes in the developmental cycle of *G. pallidipes*. Firstly, the small RNAs from the five developmental stages (larvae, pupae, teneral adults, non-teneral adults and gravid females) were sequenced using next generation sequencing technology. A total of 157 miRNAs were identified, which included 99 known tsetse miRNAs, 46 miRNAs conserved in other insects, and 12 novel miRNAs that had not been reported in any species. Moreover, we identified 93 miRNA genes that were differentially expressed by sex and/or in specific developmental stages. Hence, the 5550 target genes for the differentially expressed genes were identified using miRanda and RNAhyrid. Functional annotations using Blast2GO yielded KEGG pathways: Purine metabolism (n = 183), Thiamine metabolism (n = 163), Biosynthesis of antibiotics (n = 55), Aminoacyl-tRNA biosynthesis (n = 16), Pyruvate metabolism (n = 13), Pyrimidine metabolism (n = 12), Glycerophospholipid metabolism (n = 12), Amino sugar and nucleotide sugar metabolism (n = 11) and cysteine and methionine metabolism (n = 11). Also, gene ontology including, cellular process (n = 1464), metabolic process (n = 1383), cell part (n = 790), organelle (n = 609), membrane (n = 588), biological regulation (n = 467), localization (n = 398), response to stimulus (n= 285), cellular component organization or biogenesis (n = 222), and signalling (n = 218). The finding offers miRNA genes involved in the regulation of critical physiological functions specific towards *G. pallidipes* development, which could eventually guide other scientist in the use of miRNA genes as a combination of novel approaches and the available technique to combat tsetse-borne diseases.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Tsetse flies are the cyclical vectors of African trypanosomes, the causative agent of sleeping sickness in humans and nagana in livestock and wild animals (Krafsur & Maudlin, 2018). The disease is widespread across 38 countries in sub-Saharan region that is endemic with tsetse flies (Wamwiri & Changasi, 2016) The African climates supports diverse tsetse species, such as *Glossina brevipalpis, G. morsitans,G. austeni,G. swynnertoni, G. pallidipes,G. longipalpis,G. Tachinoides* and *G. palpalis*, which are responsible for AAT transmission among livestock in Africa. It is the major constraint in livestock production and has led to economic loss of US$ 4.75 billion per year (Krafsur & Maudlin, 2018).

African trypanosomiasis is re-emerging due to lack of proper control strategies, vaccination and effective chemotherapeutic agents (Meyer *et al.,* 2016). The cheapest and effective way to control the disease is to reduce the vector in the environment and their ability to transmit trypanosomes (Meyer *et al.,* 2016). According, to the Pan-African Tsetse and Trypanosomiasis Eradication Campaign (PATTEC, 2013), vector control remains the most economically sustainable option for combating the diseases. Since 1980s, tsetse flies have been controlled by different methods including: sterile insect technique (SIT) and tiny target control baits, sequential aerial spraying (SAS) and insecticide-treated traps and targets (ITT), which are cost effective and sustainable to the environment. However, the control of trypanosomiasis has still been elusive. The genetics of the vector plays a crucial role in developing better vector control strategies, for example SIT, which was built on the basis of understanding the fecundity of the vector (Feldmann *et al.,* 2021).

MicroRNAs comprise a large family of endogenous, evolutionarily conserved, ncRNA that post-transcriptionally regulate mRNAs and influence fundamental cellular processes and gene expression programs in metazoan animals and plants (Achkar *et al.*, 2016). They modulate gene expression by binding to their target mRNAs (Oliveira *et al.*, 2017; Seok *et al.*, 2016). Because of its versatility, miRNAs have evolved as a major class of gene-regulatory molecules critical for diverse biological processes such as cell proliferation, differentiation, apoptosis, stress and immune response (E. Aksoy *et al.*, 2016; Meki *et al.*, 2018). Comparing miRNA expression profiles in infected and uninfected tissues provides an overall picture of cellular miRNAs that change following infection with a microorganism (Carthew et al., 2017; X. Li, 2017; Seok et al., 2016). Hence, identification of miRNAs and their physiological and pathological roles have become an essential research topic.

After the discovery of the first miRNA in *C. elegans*, thousands of miRNAs have been identified (Kozomara et al., 2019). The major approaches to identifying miRNAs include genetic screening, direct cloning, bioinformatics analyses, and deep sequencing (Feldmann et al., 2021; Sheng Li et al., 2019). The earliest discovery of miRNA genes was through direct cloning, which was time consuming and expensive. This led to the introduction of computational approaches and bioinformatics analysis that were based on complete genome sequences. However, these approaches were only available to a limited number of model species, limiting the application of these programs for the numerous non-model species (Kozomara *et al.*, 2019). Next generation sequencing (NGS) has provided an innovative look into the genome with an unprecedented depth of coverage (Slatko et al., 2018). It allows for a comprehensive coverage of miRNAs of any species because miRNAs can be detected in any organism without prior sequence or secondary structure information (Feldmann et al., 2021; Sheng Li et al., 2019; Z. Li et al., 2021). Consequently, this new approach has opened the door to functional genomics analyses of non-model species. It is widely used for profiling miRNAs in populations in various developmental stages (Akkaya et al., 2020; Carthew et al., 2017). This study aims to increase knowledge on genetics especially in the study of miRNA genes in tsetse

flies. Hence, the knowledge acquired could be eventually used to improve the control methods available for reducing trypanosomiasis transmission (E. Aksoy *et al.,* 2016; Meki *et al.,* 2018).

1.2 Statement of the problem

MicroRNAs are non protein coding small RNAs of length 18–24 nucleotides that are produced by plants, animals and viruses (Kozomara *et al.,* 2019) . They regulate numerous biological processes such as growth and development, differentiation, disease progression, apoptosis and immunity (W. Liu *et al.,* 2017; Shirjang *et al.,* 2019). There are currently 99 miRNAs identified in the *Glossina pallidipes* genome and they are highly stable molecules, which makes them potential candidates for the development of new vector control tools (E. Aksoy et al., 2016). More importantly, they are regulators of gene expression and changes in their expression pattern may have great impact on cellular pathways crucial to the development of the *Glossina* spp (Yang et al., 2021). The vast majority of studies on miRNAs in these species have only focused on the innate immunity of the insect. In 2016, the first tsetse miRNAs was characterized in *G. morsitan* in response to trypanosome infection (E. Aksoy et al., 2016). Furthermore, they were able to characterise miR-275, which was highly expressed when the vector was infected by the trypanosome parasites. In addition, a recent study investigating the role miRNAs play in immunity of *G. pallidipes*, when challenged with salivary gland hypertrophy virus (*Hytrosavirus*) (Meki *et al.,* 2018), found that miRNA gene such as miR-184-3p may play a crucial in regulating important biological process. Despite this, numerous studies need to be conducted to understand the roles miRNAs play in the development of *Glossina* spp. However, no studies have focused on understanding the regulation of miRNAs across developmental stages of *Glossina* spp. The identification, characterisation and functional analysis of dysregulated miRNAs across the life stages of *G. pallidipes* would yield information on tsetse biology and new vector control strategies.

1.3 Objectives

1.3.1 General objective

To identify and characterize microRNA genes in the developmental stages of the insect vector *G. pallidipes*.

1.3.2 Specific objectives

1. To identify miRNAs from *G. pallidipes* developmental stages (larva, pupa, teneral, adult, non-teneral adult and gravid adult)
2. To determine stage- specific and sex- specific miRNA expression profile in *G. pallidipes*
3. To investigate potential functions that correlates to stage-specific and sex-specific miRNAs of the *G. pallidipes*.

1.4 Null hypothesis

1. MicroRNAs are not expressed in the developmental stages of *G. pallidipes* species.
2. The target genes of the miRNAs has no important functional roles in the development of *G. pallidipes* species

1.5 Research Questions

1. Do miRNA genes differ in the different life cycle stages of *G. pallidipes* ?
2. Are the expression profiles of miRNAs genes stage-specific in the life cycle of the *G. pallidipes* ?
3. What are the functions of the differentially expressed miRNA genes ?

1.6 Justification of the study

African animal trypanosomiasis (AAT) affects livestock productivity in sub-Saharan Africa (Cayla *et al.,* 2019). Despite, AAT begin of economic importance, little is done on the way of reducing this diseases in the environment (Büscher *et al.,* 2017). Control measures such as vaccination and chemotherapeutic agent have been misused by farmers causing resistance by the parasite towards these agents (Venturelli *et al.,* 2022). Although, vector control such as traps, baits and SIT measures are viable and more sustainable, this technique requires expertise and constant monitoring to effectively reduce the population in the environment (Krinsky, 2019). A better understanding of tsetse biology, including the developmental cycle and the involved genes can greatly facilitate the development of new control strategies. Moreover, basic research can help to identify other regulatory elements and genes involved in tsetse metamorphosis. MicroRNAs regulate different biological process such as development and metamorphosis in other insect species (Sun *et al.,* 2016).

Majority of the known miRNAs have been classified in the recent past years using traditional methods such as cloning method and PCR technique, which is time consuming and labour intensive (Jain *et al.,* 2015). However, improvement in technologies have led to an innovative tool such as NGS technique to view the genome with an unbiased depth of coverage (Jain *et al.,* 2015). The Illumina sequencing approach is one of these high throughput technologies by which miRNAs in any organism can be detected without prior sequence or secondary structure information (Jain *et al.,* 2015). Previous profiling studies using this method have found specific miRNAs (e.g. miR-275, miR-184) playing a critical role in the innate immunity of this species (Aksoy *et al.,* 2016; Meki *et al.,* 2018). However, to date, no published study has quantified the concentration of miRNAs in *G. pallidipes* developmental cycle. Thus, conducting miRNA profiling studies for the life cycle stages of *G. pallidipes* will assist researchers on more informed decision on the use of miRNA genes to development additional tools for trypanosomiasis contol.

CHAPTER TWO

LITERATURE REVIEW

2.1 Epidemiology

Trypanosomiasis is an epizootic diseases caused by a hemoprotozoan belonging to the genus *Trypanosoma* in the family of Trypanosomatidae (Wamwiri & Auma, 2021). it was first discovered in 1895, by a pathologist known as David Bruce (Büscher *et al.,* 2017). The disease is confined in regions with the biological vector (tsetse fly) or the biting flies (Figure 2.1). It has become both an economic and medical burden in the African continent. It is categorized in two forms such as Human African trypanosomiasis (HAT) and Animal African trypanosomiasis (AAT). By 1998, almost 40,000 cases of HAT were reported and the prevalence in countries such as Sudan was about 50% of the population in several villages (Büscher *et al.,* 2017). This encouraged surveillance of the disease and the number of cases dropped from 40,000 to 17,616 in the year 2010 (Büscher *et al.,* 2017). In 2015, there have been 2,804 cases recorded due to proper surveillance and monitoring of the disease (Moran & Parker, 2016).

However, AAT remains as one of the biggest constraints to livestock production and a threat to food security in sub-Saharan Africa (Wamwiri & Auma, 2021). The epidemiology of the disease is characterized by three elements including biological vector, trypanosome and the vertebrate host. The AAT-causing parasites include: *T. vivax, T. simiae, T. brucei brucei*, *T. conglense*, *T. evansi* that mainly affects livestock and wild animals (Cayla *et al.,* 2019). In Africa, *T. vivax* and *T. congolense* is transmitted by *Glossina* species and mechanically by horse flies (tabanids and *Stomoxy* spp.) (Cayla *et al.,* 2019). It is found in countries such as Ethiopia, Kenya, Nigeria, Tanzania, Uganda, Zambia, Zimbabwe, and Latin America (Attardo *et al.,* 2014). They mortality rate is high in cattle, sheep, goat and dogs (Nimpaye *et al.,* 2011). *T. evansi* is mostly found in dry areas such as north-eastern part of Kenya and Ethiopia and this

parasite species is mostly transmitted by hematophagous biting flies resulting in high mortality rates in camels (Bargul *et al.*, 2016). *T. simiae* is transmitted by tsetse flies and mostly affects pigs, sheep, goats and cattle. *T. simiae* is widely found in central Africa (Nimpaye *et al.*, 2011).

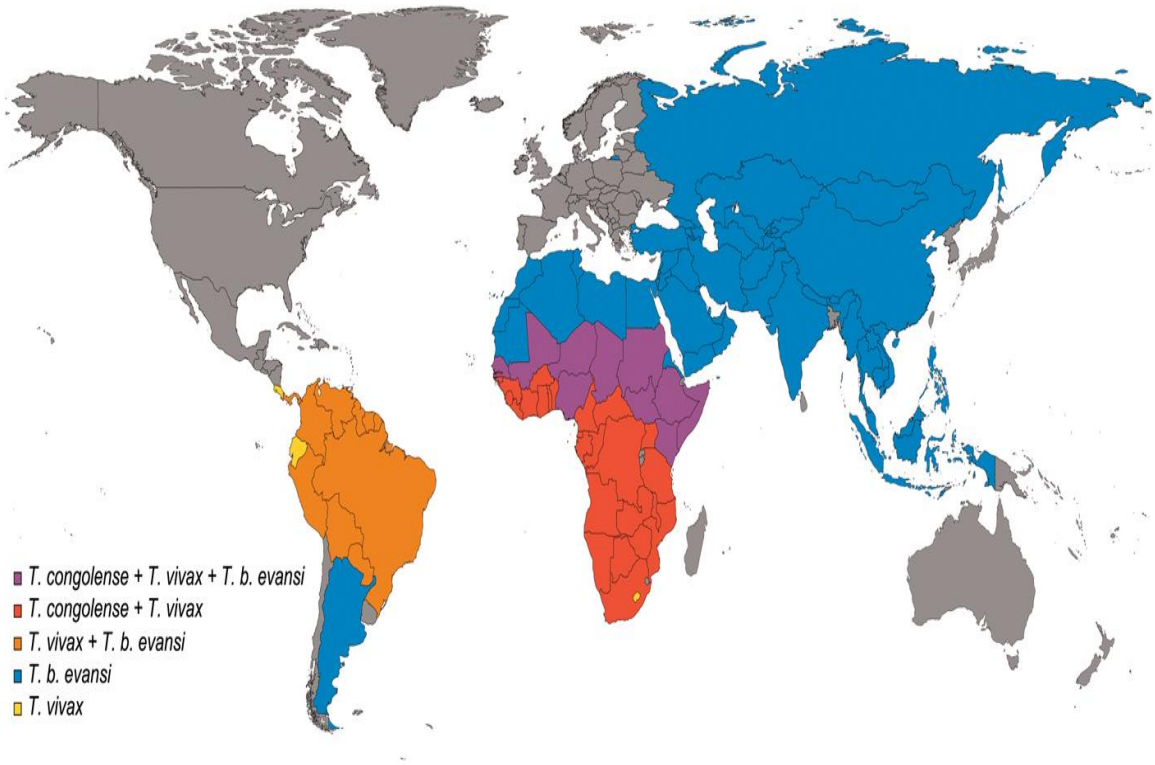


Figure 2.1: Geographical representation of animal trypanosomiasis.

The figure shows the distribution of trypanosomasis transmitted by tsetse flies and other bitting flies across different continent. Tsetse flies are only endemic to sub-Saharan Africa, specifically in 38 countries, and are not found in other continents. In South America/Brazil and in Asia, for instance, the transmission of animal trypanosomosis is caused by the hematophagous biting flies such as tabanids and *Stomoxys* spp, among

others. And the infections are caused mostly by *T. vivax, T. evansi,* and *T. congolense* – not *T. brucei* that requires cyclical transmission by tsetse fly.

For instance; *Trypanosoma evansi* and *T. vivax* infections are commonly reported in horses, mules, camels, and cattle in Asia and South America (Desquesnes and Dia, 2003; Mihok *et al.,* 1995; Jones and Davila, 2001)The key illustrates the different species of trypanosomes (*T. Vivax , T. brucei,T. congolense, T. b. evansi*) (Giordani *et al.,* 2016)

2.2 African trypanosomes

African trypanosomes are flagellated protozoan parasites that cause nagana and sleeping sickness (Cayla *et al.,* 2019). Due to the large difference between the two hosts, the trypanosome undergoes complex changes during its life cycle to facilitate its survival in the insect's gut and mammalian blood stream. After, the tsetse fly takes a blood meal on an infected host, the ingested trypanosomes reach the midgut of the tsetse fly. The parasite differentiates into procyclic forms from stumpy trypomastigotes. It is then differentiated into several sequential forms during its life cycle in the tsetse fly. The early procyclic form is differentiated in to the late procyclic form in the mid gut of the insect, followed by mesocyclic form (Büscher *et al.,* 2017). Once the mesocyclic form is developed into the anterior midgut, it is proliferated into epimastigote form in the salivary gland or proboscis depending on the trypanosome species.

The late stage, which is infectious to the mammalian host differentiates into non-proliferating metacyclic form (Büscher *et al.,* 2017). Once the infected fly feed on the host, the parasite moves from the saliva of the fly into the blood stream as a trypomastigote and the parasite are carried throughout the body reaching the body fluids (lymph and spinal fluid) (Geiger *et al.,* 2015). The parasite replicates by binary fission causing either nagana or sleeping sickness. Throughout its developmental cycle, the trypanosome variant surface glycoprotein is modified to evade the immune response of the vector and mammalian host (Horn, 2014). To date, there is no effective vaccine

against the African trypanosomes owing to the ever changing huge repertoire of expression of the parasite's surface Variant Surface Glycoproteins (VGS) throughout its life cycle (Horn, 2014).

2.3 Tsetse flies

Tsetse flies (*Glossina* spp.) are exclusively blood feeding vectors that transmit trypanosomiasis (Geoffrey M. Attardo *et al.,* 2019). They are distributed in frequently within 38 countries in Africa (Figure 2.2).Tsetse fly species belong to three different groups according to their habitat; (i) *fusca* "forest" flies,(ii) *morsitans* "savannah" flies, and (iii) *palpalis* "riverine" flies (Gooding & Krafsur, 2005). The *palpalis* group is the major vector of *Trypanosoma brucei gambiense,* which is responsible for the chronic form of HAT in 24 countries of western and central Africa (Bottieau & Clerinx, 2019; Cayla *et al.,* 2019) and poorly transmits *T. brucei brucei* and *T congolense* (Silvester *et al.,* 2018). The '*morsitans* group' is the major vector for trypanosomes of the *Trypanosoma brucei brucei* (Tbb), *T. congolese,* and *T. vivax,* which are prevalent in cattle and the main nagana-causing parasites in sub-Saharan Africa (Cayla *et al.,* 2019; Mbewe *et al.,* 2018; Saini *et al.,* 2017). This disease is responsible for 3 million cattle deaths yearly and an estimated loss of US$4.75 billion in revenue (Saini *et al.,* 2017).

Furthermore, the *morsitans* group is the vector of *T. brucei rhodesiense* (Tbr), the causative agent of the acute form of HAT that is endemic in 13 east African countries (de Koning, 2020). Data on tsetse fly distribution in Africa indicates that the species are wide spread in eastern and western countries in Africa due to their geographical sites. In addition, recent studies indicate there about eight species across eastern African region alone (*G. pallidipes*, *G. morsitans morsitans*,*G. austeni* and*G. swynnertoni* (*morsitans* group),*G. fuscipes fuscipes* and *G. tachinoides* (palpalis group), *G. brevipalpis,* and *G. longipennis* (fisca group – subgenus *Austenina*) (Cecchi *et al.,* 2015). In Kenya, data on distribution is scarce and much of the available information relies on locally reported cases, topographical information and

entomological surveys. This is due to the variations in Kenya's weather patterns in (Langley *et al.*, 2017).

2.3.1 Tsetse distribution

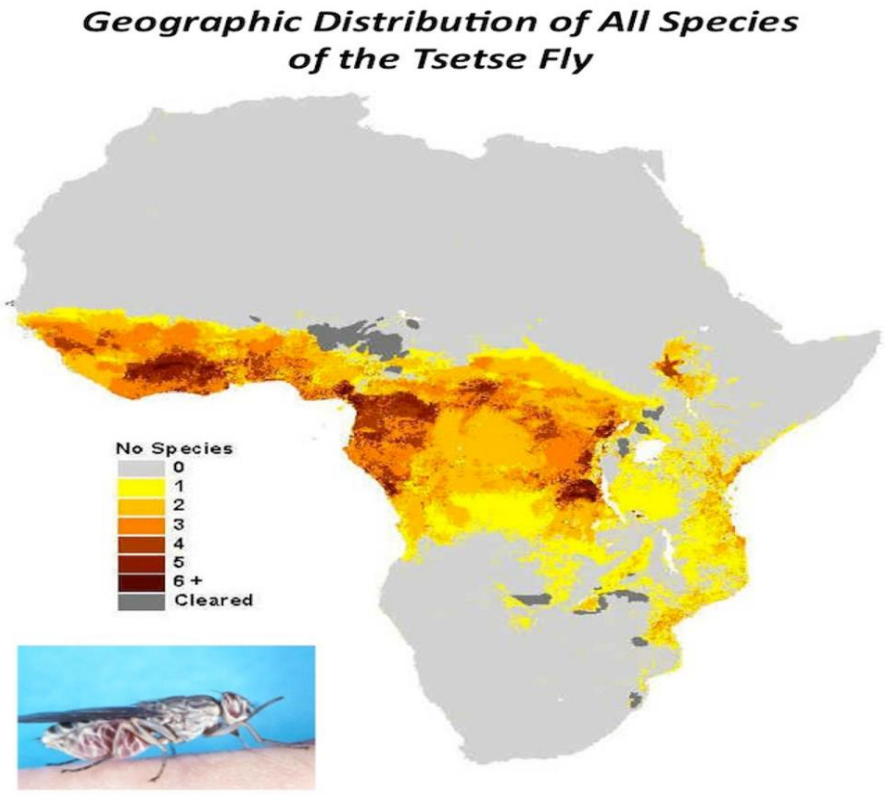


Figure 2.2: Tsetse fly distribution in sub-Saharan Africa demonstrates the distribution of the tsetse fly species in the sub-Saharan region of the African continent. The key demonstrates the different number of tsetse species in the sub-Saharan African countries (Conca James, 2015).

2.3.2 The life cycle of the tsetse fly

All tsetse, *Glossina* go through four stages in their life cycle, namely egg, larva, pupa, and adult (Figure2.3). The first two stages develop in the gravid female by a process

referred to as adenotrophic viviparity and last 7-9 days, depending on the species and the ambient temperature (Tirados Estebanez, 2014).

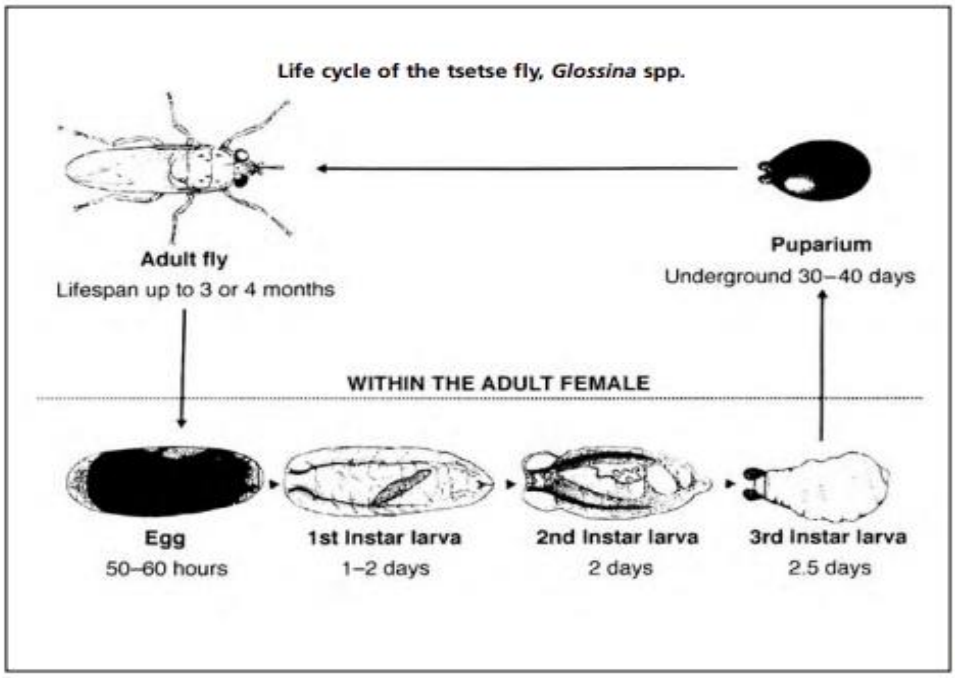


Figure 2.3: The life cycle of a tsetse fly (Estebanez et al., 2014). Hence, the complete metamorphosis cycle of tsetse includes: egg, 1st larva, 2nd larva, 3rd larva, pupa, and adults. Source: Estebanez, 2014.

2.3.2.1 Egg

The tsetse flies species **(**Glossinidae) belonging to the Hippoboscoidea family. In addition, they have a unique reproductive cycle where it carries the young and give birth to live larva referred to as adenotrophic viviparous (Bonnet *et al.,* 2015). Once mating occurs the female tsetse flies stores the sperm in the spermathecal and can reproduce young larvae throughout its life time, even without mating again (de Beer *et al.,* 2017). They fertilizes one ovary every gonadotrophic cycle (Bonnet *et al.,* 2015). Therefore,

producing only one offspring each cycle, after ovulation the right ovary is fertilized and the egg, which is about 1.6 mm long, is released to the uterus and embryogenesis occurs within 3-4 days depending on the nutrition and species of tsetse (Bonnet *et al.*, 2015).

2.3.2.2 Larva

As the tsetse larva grows, it undergoes several developmental stages referred to as instars. There are three larval instars that develop in the gravid female, namely first, second and third instar. The larva has a mouth at the anterior and two posterior spiracles. The first instar larva emerges from the egg by breaking out of the chronion using the sharp egg tooth (Krinsky, 2019). It can grow up to 1.8 mm, where it gets rid of its old skin after about 1 day to moves to the next instar (Caljon *et al.,* 2017; Krafsur *et al.,* 2016). The second instar grows up to 4.5 mm and lasts for approximately two days (Lee R Haines *et al.,* 2020). This stage is characterized by rapid growth and development through the posterior spiracles swelling and between the spiracles in the area of small spines (English *et al.,* 2016). The third instar also is characterized by rapid growth and development. The larva can grow up to a length of 6-7 mm (Lee R Haines *et al.,* 2020).

The weight and volume may be due to the gut containing large amount of unassimilated food (English *et al.,* 2016). Once it is fully grown a pair of large black swelling posterior end can be observed referred to as polypneustic lobes, which carries many small holes through which larva breathes. The air is supplied by the female vulva (English *et al.,* 2016). All the three instars are fed by food from the egg and milky secretion from the milk gland of the female fly (S. Aksoy, 2019). The larva sucks the secretion and passes it straight to the midgut. Here it is slowly digested and assimilated for larval growth (Krinsky, 2019). The entire process could take about 6-9 days in the uterus of the female prior to larviposition (Lee R Haines *et al.,* 2020).

2.3.2.3 Pupa

The pupa is dark brown in colour and is round-shaped. Its posterior end contains the polypneustic lobes, whose shape is distinct for each species of tsetse (Deken, 2014). Once the larva is larviposted and burrows into the ground, within less than thirty minutes, it molts into pupa with a smaller size compared to the 3rd instar larvae (Moran & Parker, 2016). The pupa develops a hard casing referred to as a puparium (Moran & Parker, 2016). The remaining food in the midgut is digested and assimilated to form organs of the adult fly (Hargrove & Vale, 2020). The pupal stage usually lasts about four to five weeks, according to the temperature of the environment (Hargrove & Vale, 2020). Higher temperatures shorten the pupal period; whereas lower temperatures lengthen it for up to 50 days or more in some climates. extreme high or low temperatures will result in death (Are & Hargrove, 2020).

2.3.2.4 Adult

The young adult emerges from the puparium casting and using its ptilinum, to exits from the surrounding (Lee Rafuse Haines, 2013). The body of the adult teneral fly (juvenile flies) is divided into three parts, namely head, thorax, and abdomen (Ta *et al.,* 2021). At this stage, the body is still soft and the wings are not fully developed (Lee Rafuse Haines, 2013). The underside of the abdomen appears whitish in colour and semi-transparent (Krafsur & Maudlin, 2018). The head is modified to acquire sensory information and for feeding. Hence, it contains eyes, a pair of long and many segmented antennae and an extended proboscis (Moran & Parker, 2016). Once it has emerged from the ground, the teneral fly uses its extended proboscis to acquire a meal from a mammalian host. The underside of the abdomen appears creamier yellow and when held up in light, the dark shape of the blood meal can be observed.. The muscles begin to develop making the throax feel a little firmer and harder (Moran & Parker, 2016).

Also, the ptilinum disappears and the wing become fully developed to assist in movement of the fly (Moran & Parker, 2016). Consequently, this stage is referred to as non-teneral adult stage or mature stage (Lee Rafuse Haines, 2013). The flies are obligate hematophagous species and solely feeds on blood (Rio *et al.,* 2016) . In the wild the adult females survive longer than adult male and can survive for several month depending on the environment condition (S. Aksoy *et al.,* 2013).

2.3.3 The insect vector *Glossina pallidipes*

Glossina pallidipes is one of the 33 species of tsetse flies belonging to the morsitans "savannah" group (Krafsur *et al.,* 2016). The *G. pallidipes* are zoophagic and most of its population feed on livestock and wild animals but other sub-populations of the species may feed on human blood(Okeyo *et al.,* 2017). They are the major vector of the pathogenic *Trypanosoma* species, *T. vivax* and *T. congolense,* for animals and a minor vector of *T. brucei rhodesiense* for humans (Silvester *et al.,* 2018). They have become the most crucial vector for transmission of AAT in eastern and western African region contributing to loss of livestock and lively-hood for the people in the community (Mbahin *et al.,* 2013; Mbewe *et al.,* 2018; Saini *et al.,* 2017). The species plethora depends on suitable condition and the availability of their mammalian host. Therefore, their populations are scattered and patchy throughout the sub-Saharan region of Africa. Although, their population are scattered, these species remain the most virulent vector of trypanosomiasis. According to a study conducted, the species has become both an economic and medical burden in countries such as Ethiopia, Kenya, Tanzania, Uganda, Zambia and Zimbabwe (Bateta *et al.,* 2020; Chilongo *et al.,* 2021; Mbewe *et al.,* 2018).

2.4 Vector control measures

The tsetse fly has a unique life cycle as previously described. The absence of eggs and free larval stage in nature and the pupal stage development occurring in the soil, the adult fly is the only phase targeted by most control strategies (English *et al.,* 2016).

Early methods for control include use of insectides such as DDT and hydro carbon, clearing of bushes and wild life curling, which was harmful to the environment (S. Aksoy, 2019). Due to the environmental effects, these controls prompted for a more effective and environment friendly techniques.

The current control strategies developed include the use of, sequential aerial spraying (SAS) and insecticide-treated traps and targets (ITT) and introduction to natural predators such as bombylid flies, *Exhylanthrax* spp., *Syntomosphyrum* and *Muttilla glossinae* (Engel & Moran, 2013), which is eco-friendly. Despite the effectiveness of these controls, cumbering the disease has been difficult due to environment challenges and dependence on local participants. Improvement in molecular technology has developed new technique in vector control (S. Aksoy, 2019). This includes use of sterile insect technique (SIT) and incompatible insect technique (IIT). The SIT technique has been used in countries such as Burkina Faso and Zanzibar eradicating up to 70% of the tsetse population (Pagabeleguem *et al.,* 2016). This technique has focused on the unique reproductive cycle of the tsetse fly. The female tsetse can only mate once and reproduce offspring throughout it is life span. The SIT involves mass production of target species to reduce the population.

The male species are obtained in the pupa stage and exposed to irradiation (gamma rays), which affects the spermatogenesis of the male. Once the adult male species have emerged, they are released to the environment to mate with the female. The population of the tsetse decreases drastically, because sterile sperms cause embryo abortion in the female species (Demirbas-Uzel *et al.,* 2018). The vectoral capacity of the male tsetse is not affected by irradiation, therefore, they are able to transmit trypanosome once they feed on an infected host.

The IIT involves the use of microbial symbionts such as *Wolbachia, Wiggleworthia* to limit the prevalence of the pathogenic microbes in the insect vector. The microbiota infers immunity in the invertebrate hosts via different approaches including the

modulation of immune responses, niche competition or production of inhibitory molecules (Demirbas-Uzel *et al.,* 2018; Doudoumis *et al.,* 2017; Hoffmann *et al.,* 2011). Therefore, microbial symbionts could have a great impact on the vector competence (Geoffrey M. Attardo *et al.,* 2019). In contrast most of the symbionts such as *Wolbachia*, *Wigglesworthia* has become difficult to cultivate in the laboratory for mass production (Doudoumis *et al.,* 2017; G. Zhang *et al.,* 2013).

2.5 Vector-parasite interaction

The trypanosomes parasite undergoes several series of development within the vector before it can be transmitted to the host. The developmental process of parasite in tsetse begins in the vertebrate host by differentiation of the long slender bloodstream forms (BSF) into non-dividing stumpy forms (ST) (Bateta *et al.,* 2017). The BSF are digested in the gut while stumpy parasites are capable of differentiating into a procyclic form (PCF). The major bottleneck for parasites development occurs in the gut, where most of the parasites are eliminated within days post acquisition due to the innate immunity of the fly (Bateta *et al.,* 2017). The PCF can evade the midgut of the vector in susceptible flies and colonize the proventiculus organ developing into a mesocylic trypamastigotes that migrates to the proventiculus, where the fly midgut is aligned with proventriculus membrane (Bateta *et al.,* 2017).

The short epimastigotes are formed and enter into the salivary glands, where they attach to the epithelial cells and differentiate into metacyclic form infectious to the mammal. However studies have shown that only small number are able to cross the proventiculus membrane to infect the salivary gland (Geoffrey M. Attardo *et al.,* 2019; Bateta *et al.,* 2017; Lee Rafuse Haines, 2013). The reduction may be influenced by the sophisticated immune factors existing in tsetse fly (Bateta *et al.,* 2017)(Bateta *et al.,* 2017). A class of small non-coding RNA (ncRNA) molecules known as microRNA (MicroRNAs) has recently been altered during *Glossina* response to trypanosome parasite infection (E. Aksoy *et al.,* 2016), suggesting that MicroRNAs may be likely candidates for serving as

regulates of defence mechanism to the parasite by modulating the expression level or being a part of the sensing process (E. Aksoy *et al.*, 2016).

2.6 Pathogenesis of trypanosomiasis

The infected tsetse injects metacyclic trypanosomes into the skin of the animals causing inflammation (chancres) (Ntantiso *et al.,* 2014). The parasite enter the lymph and lymph nodes, then the bloodstream, where they divide rapidly by binary fission (Prayag *et al.,* 2020). *Trypanosoma brucei* species and *T vivax* invade tissues and cause tissue damage in several organs (Nimpaye *et al.,* 2011). Whilst *T congolense* attaches to endothelial cells and localize in capillaries and small blood vessels. The immune response is vigorous, and immune complexes cause inflammation, which contributes to fever and other signs and lesions of the disease (Morrison *et al.,* 2016). Antibodies against the surface-coat glycoproteins kill the trypanosomes (Morrison *et al.,* 2016). However, trypanosomes have a large family of genes that code for variable surface-coat glycoproteins that are switched in response to the antibody evading immunity. This antigenic variation results in persistence of the organism (Horn, 2014). Antigenic variation has prevented development of a protective vaccine and permits reinfections when animals are exposed to a new antigenic type (E. Aksoy *et al.*, 2016).

2.7 Trypanosomiasis management

2.7.1 Clinical presentation of Trypanosomiasis

The severity of trypanosomiasis varies with species and age of the animal infected and species of trypanosome involved (Tirados Estebanez, 2014). The incubation period is usually 1–4 weeks (Cayla *et al.*, 2019). The primary clinical signs are intermittent fever, anemia, and weight loss. Cattle usually have a chronic course with high mortality,

especially if there is poor nutrition or other stress factors (Odeniran & Ademola, 2018). Ruminants may gradually recover if the number of infected trypanosome parasite is low; however, stress results in relapse (Morrison *et al.,* 2016). Necropsy findings vary and are nonspecific. In acute, fatal cases, extensive petechiation of the serosal membranes, especially in the peritoneal cavity, may occur (Morrison *et al.,* 2016). Also, the lymph nodes and spleen are usually swollen. In chronic cases, swollen lymph nodes, serous atrophy of fat, and anemia are seen (Morrison *et al.,* 2016).

2.7.2 Primary prevention

Prevention and control focuses on the eradication of the parasitic host and the tsetse fly. Hence there is a need for regular surveillance, involving case detection and treatment (G. M. Attardo *et al.,* 2014). Two alternative strategies have been used to reduce this disease. One tactic is primarily medical care that targets the disease directly using monitoring, prophylaxis, treatment, and surveillance to reduce the number of organisms that carry the disease (Rodriguez *et al.,* 2022). The second strategy is generally entomological and seeks to disrupt the cycle of transmission by reducing the number of flies (Krinsky, 2019).

2.7.3 Secondary prevention

2.7.3.1 Clinical examination and Laboratory tests

Early diagnosis is difficult because signs and symptoms in the first stage are non-specific and the insensitive nature of the diagnostics (Desquesnes *et al.,* 2022). Diagnosis requires confirming the presence of the parasite in any body fluid (Desquesnes *et al.,* 2022). The classic approach for diagnosing an infection is by light-microscopic detection of the parasite in a lymph node aspirate (usually, from a posterior cervical node) (Sadek *et al.,* 2021). Another method is through serologic testing for parasite, which is applied for screening purposes only and the definitive diagnosis rests is on microscopic observation of the parasite (Bottieau & Clerinx, 2019). Hence

with advancement in technology, rapid diagnostic tests for *Trypanosome* infection such as the Surra Sero K-SeT for AAT diagnostics and the SD Bioline HAT 1.0 for HAT diagnosis. are developed and introduced in endemic areas, where laboratories are inaccessible (Birhanu *et al.,* 2015; Boelaert *et al.,* 2018).

Staging for Trypanosome (i.e., assessment of neurological infection) is performed by microscopic examination of Celebral Spinal Fluid (CSF) is collected by lumbar puncture on a wet preparation looking for motile trypomastigotes and white blood cells (WBC), with severe symptoms (meningo-encephalic stage symptoms) (Bonnet *et al.,* 2015). Patients with five or fewer WBC per microliter and no trypomastigotes are considered to be in the first stage, and those with more than five WBCs per microliter or trypomastigotes are considered to be in the second stage (Bonnet *et al.,* 2015). CSF testing is done after a parasitologic diagnosis has been made by microscopic examination of blood, lymph node aspirates, chancre fluid, or bone marrow or when indications of infection are present that justify a lumbar puncture (e.g., clinical signs and symptoms of sleeping sickness or strong serologic suspicion). Isolation of the parasite by inoculation of rats or mice is a sensitive diagnostic method, but its use is limited to *some trypanosome species such as (T. b. rhodesiense)* (Bonnet *et al.,* 2015; Desquesnes *et al.,* 2022).

2.7.4 The potential roles of microRNA in trypanosomiasis management

An investigation of stage-specific miRNAs may provide an understanding of tsetse biology and provide tsetse-specific targets for disease control. Significant stage-specific expression was observed for miRNAs in various species (Asgari, 2013). In *Anopheles*, aan-miR-2943 and afu-miR-980 were only expressed in the egg stage in *An. anthropophagus* (Liu et al., 2017) and *An. funestus*, (Allam *et al.,* 2016) respectively. ast-miR-2943 and ast-miR-2945 were highly expressed in *An. stephensi* embryos, and ast-miR-1890 had a peak expression in *An. stephensi* pupae. Jain and colleagues (Jain *et al.,* 2015) reported that 36 miRNAs were differentially expressed

among various developmental stages of *An. stephensi*, including larval male and female, pupal male and female, and adult male and female. Among them, ast-miR-1891, ast-miR-190-3p, ast-miR-285, ast-miR-988-3p, and ast-miR-989 were absent in the larval stage, but ast-miR-8-3p was the most abundant in the male and female larval stages. ast-bantam-3p was the most abundant in the male and female pupal stages of development, and ast-miR-281-5p and ast-miR-bantam-3p were the most abundant in adult males and females, respectively, ast-miR-14 had a relatively strong signal from the late embryonic to adult stages (Jain *et al.,* 2015).

Expression analysis of miRNAs revealed distinct patterns from early embryo to adult stages in Aedes. In *Ae. albopictus*, aal-mir-M1 was only expressed in embryos, and aal-mir-9a was mainly expressed in embryo and larval stages. aal-let-7 was only expressed in pupal and adult stages, and aal-miR-1175 was widely expressed in all of the life stages, except for embryos (Megy *et al.,* 2012). Conserved miRNAs are likely to be involved in important functions in tsetse lineages. In *Glossina morsitans,* the expression of *miR275* was experimentally reduced in tsetse's cardia and midgut through the provisioning of synthetic anti-*miR275* antagomirs (antagomir-275) or VSG purified from BSF trypanosomes, formation of the fly's PM was impaired. This process disrupted blood meal digestion and enhanced the ability of trypanosomes to establish an infection in the fly's midgut (E. Aksoy *et al.,* 2016). In 2018, a study showed a significant change in expression profile of 15 *G. pallidipes* miRNAs in response to *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) invasion (Meki *et al.,* 2018). Furthermore, when they generated a mimic of one miRNA gene (miR-184-3p), which was highly abundant in symptomatically infected flies compared to the asymptomatic individuals, they observed better survival of the virus suggesting involvement of these tiny ncRNAs as a new layer in the regulation of tsetse fly defence mechanism against hypertrophy virus (Meki *et al.,* 2018). Therefore, miRNAs have key roles in the regulation of distinct processes in tsetse. They provide a key and powerful tool in gene regulation and thus a potential novel class of therapeutic targets.

2.7.5 Treatment of Trypanosomiasis

Livestock infection with *Trypanosome* spp., namely *T. congolense, T. vivax, and T. evansi* , is seldom screened in several developing countries, allowing these animals to continue acting as sources of sporadic infections in animals (Birhanu *et al.,* 2015). The narrow chemotherapeutic spectrum available for AAT disease includes diminazene aceturate, isomethamidium chloride, and ethidium bromide (homidium salts) (Venturelli *et al.,* 2022). Trypanocidal drug applicability in the veterinary field ranges from prophylaxis to treatment of infected animals (Venturelli *et al.,* 2022). Prophylactic administrations allow reducing the risk of infection in areas of intense tsetse activity and high risk of transmission, or when infected animals cannot be reached (Bustamante *et al.,* 2022). The treatment options, diminazene aceturate, isomethamidium chloride, and ethidium bromide (homidium salts) accounts for 33%, 40%, and 26% of the veterinary prescriptions for AAT treatment (Venturelli *et al.,* 2022). Diminazene aceturate for trypanosomiasis prevention is applied when the other drugs are lacking for therapeutic administration (Bustamante *et al.,* 2022). However, Trypanocides are being applied on animals based on unspecific health indicators and biased choices for drug selection leading to the misuse and abuse of these drugs (Bustamante *et al.,* 2022; Venturelli *et al.,* 2022).

2.8 MicroRNAs

Ribonucleic acid (RNA) have been known to have diverse function such as genetic regulation and storage (Y. Zhang *et al.,* 2018). The major classes of RNA: messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) are responsible for information flow from DNA to proteins (Y. Zhang *et al.,* 2018). However, there are many small ncRNAs expressed in eukaryotic cells (Y. Zhang *et al.,* 2018). They have been involved in crucial roles such as catalytic, structural and regulatory roles in the cell (Y. Zhang *et al.,* 2018). Many small RNA may be involved in dosage compensation, imprinting modulating RNA polymerase and stress responses. Due to continuous stream

of novel RNA being reported, MicroRNA has revolutionized the scientific community with its ability to regulate cellular events (Allam *et al.,* 2016).

2.8.1 Discovery of microRNAs

The short stranded MicroRNA lin-4 was first discovered in 1993 as a result of an expedition to clone the lin-4 gene (Vinod Chandra, 2021). The gene was linked to having a defect in the developmental timing of the organism (*Caenorhaditis elegans*), once it was mutated. In addition, it was established that the gene could not code for a protein rather a 21-RNA nucleotide(Vinod Chandra, 2021). However it was soon discovered that the gene lin-4 was a target of the lin-14 at the 3'-untranslated region with seven complimentary nucleotide sequence giving the idea that the small RNA was functional (Xu *et al.,* 2016).

In 2000, almost seven years later a second MicroRNA, evolutionary conversed MicroRNA known as let-7 was discovered (Sun *et al.,* 2016). The gene was not only elucidated in *C. elegans*, but is also high eukaryotic species such as mammals and insects (Y. Zhang *et al.,* 2018). However, this led to scientist studying the function of MicroRNA in different organism (Asgari, 2013).

2.8.2 MicroRNA Biogenesis

MicroRNA genes are actually encoded by miR genes and are characterized in many organism (Vishnoi & Rani, 2017). Interestingly, these genes can be expressed as an individual or a cluster of 2-7 genes referred to as a MicroRNA family (Kabekkodu *et al.,* 2020). In addition, experimental studies suggest that MicroRNA families may be controlled by common regulatory process (Vishnoi & Rani, 2017). These genes can be produced from exons and introns of non-coding RNA, the introns of protein coding RNA or the 3'UTR of protein coding genes (Lai, 2015) and the repetitive regions as in the case in mammals (Kabekkodu *et al.,* 2020).

MicroRNA biogenesis occurs when the miR genes are processed by RNA polymerase II (Avesson *et al.*, 2012). These class II gene transcripts can form two types of miRNA, canonical and non-canonical types. The canonical MicroRNAs transcripts are folded into hair-loop structures known as the primary miRNA (pri-miRNA), and are processed in the nucleus by Drosha enzyme liberating the precursor miRNA (pre-MicroRNA) (Avesson *et al.*, 2012). While the non-canonical miRNA transcripts are processed by spliceosomes and intronic RNA lariant debranching enzymes to form pre-miRNAs (Vishnoi & Rani, 2017), which are exported to the cytoplasm. The pre-miRNAs are further processed by dicer enzyme to form the mature miRNA-miRNA duplexes. The miRNA duplexes are selected by the argonaute protein which is loaded on to a RISC component to form a guide strand that is incorporated in the mRNA for down regulation (Figure 2.4)

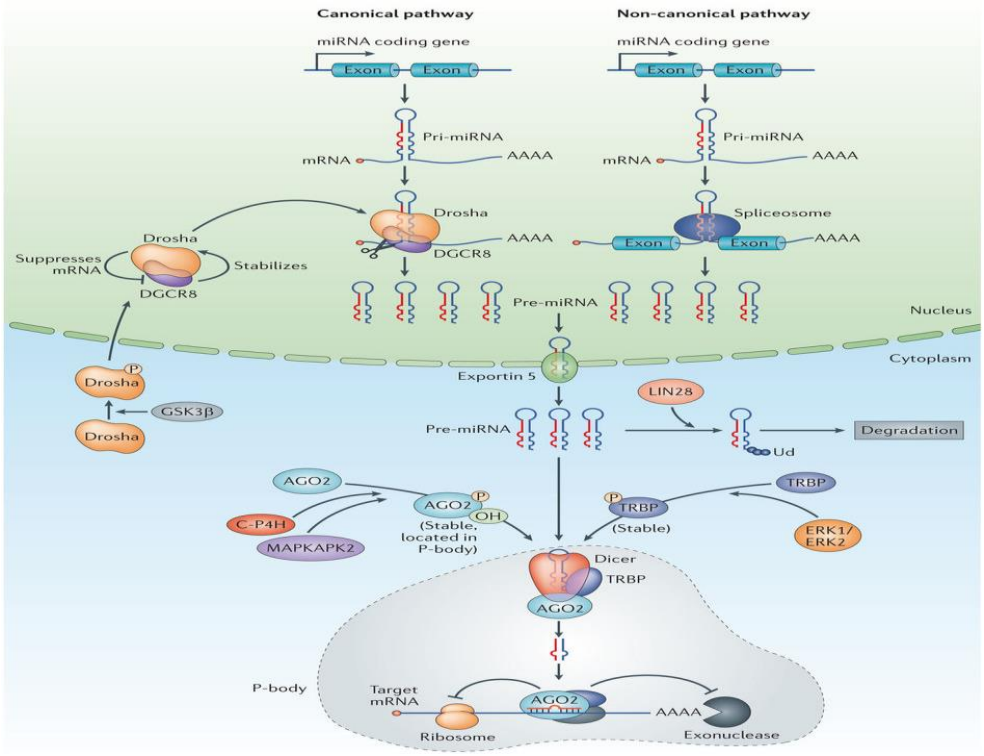


Figure 2.4: Canonical and non-canonical microRNA biogenesis pathways for biosynthesis of miRNA.

In the canonical pathway miRNA are transcribed by RNA polymerase II enzyme to produce primary miRNA hair pins. Which are then processed by the Drosha-DGR8 enzyme complex to form pre – miRNA. They are transported by exportin 5 into the cytoplasm, where they are further processed by Dicer – TRBP and loaded into argonaute 2-containing RNA induced silencing complex (RISC) factor to suppress the down streaming target gene expression. Non canonical pathway uses the spliceosome dependent mechanism. These pathways are tightly regulated. DGR8 stabilizes Drosha protein (Judice *et al.,* 2016). AGO2 is stabilised by CP4H and phosphorylated by MAPK-activated protein kinase2 and TRBP is stabilised by extracellular signal regulated kinase EPK1 or EPK2 phosphorylation (Judice *et al.,* 2016).

2.8.3 Function of microRNAs

MicroRNA genes are small non-coding RNA genes of about 18-25 base pairs, with the main function being to regulating protein production (Vishnoi & Rani, 2017). These genes have been known to bind at the 3’ UTR of the mRNA genes (Riffo-Campos *et al.,* 2016). However recently studies have shown that the miRNA can bind at any specific region on the mRNA including the 5’ UTR (Gu *et al.,* 2014). The complementary binding of the miRNA depends on the targets and itself. If the complementary is high the miRISC will degrade the mRNA. However, if the complementary is low or not sufficient enough for degradation then translation inhibition occurs. Furthermore, an mRNA can contain multiple sites for the same or different miRNAs. Consequently, several different miRNAs can act together to repress the same gene (Vishnoi & Rani, 2017).

2.8.4: Characteristics of MicroRNA genes

Small RNA such as miRNA (18 -24), Piwi-interacting RNAs (26-31nt) and small interfering RNAs (siRNA) (21-34 nt) can be easily distinguished by their length (Choudhuri, 2010). However, the precursors (pre-miRNA) have a length of 60-400 NT which is easily undistinguishable between other RNA (mRNA, tRNA, and rRNA). Although, the precursors of RNA is not easily differentiated by their structures, the minimum folding energy of the precursor serves as a means of fully characterizing the RNA. Hence, adjusting the minimum folding energy of the precursor: miRNA (0.97) and their lengths could be a precise way of distinguishing between miRNA and other RNA (Riffo-Campos *et al.,* 2016).

Majority of mature miRNA genes can be cleaved at either the 3' site or the 5' site of the pre miRNA (Riffo-Campos *et al.,* 2016). Interestingly, many mature miRNAs are believed to be evolutionarily conserved from species to species in both animal and plant kingdoms (Riffo-Campos *et al.,* 2016). The majority of the known miRNAs have high complementary site at their targeted mRNAs, and this complementarily is conserved among organism (Riffo-Campos *et al.,* 2016). In plants the complementary between miRNA and mRNA is perfect or nearly perfect leading to the mRNA degradation, but complementarily in animals is imperfect and mostly leads to translational inhibition of the mRNA (Vishnoi & Rani, 2017).

2.8.5 Identification of microRNA genes

Since miRNA discovery in the late 20th century, there has been continuous increase in the identification of miRNA genes in different organism. This has seen approaches such as biochemical methods based on purification of RNAs (Kennell *et al.,* 2012), Direct cloning (Avesson *et al.,* 2012) and computation approaches centering on the conservation of DNA region between two clearly related species being developed (Ninova *et al.,* 2016; Tirumalai *et al.,* 2020; Vilmos *et al.,* 2013). With advancement in

technology, better methods of identifying and validating have been introduced, computational based prediction has revolutionized identification of miRNA genes (Vishnoi & Rani, 2017).

Due to the continuous demand on understanding the importance of miRNA genes in different organism, computational methods have become more sensitive and precise. However, these approaches bases their analysis on major feature characteristic of miRNA genes, for instances: high evolutionary conservation within species, the secondary structure of the pre-miRNA (Carthew *et al.,* 2017; Cowled *et al.,* 2014) the minimal folding energy index and also the structure of the mature miRNA (Yen *et al.,* 2019). Classification of computation approaches is categorized into six categories:

The homology based approach involves identification of miRNA using paralogue and orthologues (Alptekin *et al.,* 2017). This approach relies on already known miRNA genes. The approach is effective enough to identify evolutionary conserved miRNA. The homology searches are classified into two genome-based or expressed sequence tags (EST) search based (Prabu & Mandal, 2010). EST search has become more renowned due to its economic feasibility for gene discovery of different species that lack a genome (Patel *et al.,* 2019). Therefore, identification of miRNA by EST may be more powerful and efficient than genome based.

The neighbour stem-loop search is based on miRNA structure and secondary structure. Many MicroRNA genes are classified into cluster like operons (Yen *et al.,* 2019). This cluster may be useful approach to identify new genes by searching the neighbours of known miRNA genes and their hairpin stem-loops structure. Moreover, this approach has been used to identify human , animal and a few plant species that lack enough clusters of miRNA (Miskiewicz *et al.,* 2017; Rissone *et al.,* 2022).

The gene finding approach identifies conserved regions within a genome region of different organism (Miskiewicz *et al.,* 2017). It involves using a window that can hold

about 110 nucleotides. The window is supplemented with a secondary prediction programs such as RNAfold (Tack *et al.*, 2018) and MFOLD (Gopinath *et al.*, 2017). Interestingly two computational programs was built using this approach include miRseeker and miRscan (Mani, 2021). The miRseeker and miRscan have identified dozens of new miRNA genes in different organisms, with further validation by experimental approaches (Mani, 2021). However, due to the window size limit in these two programs, it is difficult to employ them in the prediction of new miRNAs which have different lengths (Miskiewicz *et al.*, 2017).

Phylogenetic shadowing based approaches are built on the phylogenetic foot printing (cross species sequence comparison). Where it uses transcriptional elements within a non-coding region by comparing it to the orthologous sequences in different species (Rajendiran *et al.*, 2017). The sensitivity of this model decreases with the phylogenetic distances. Hence to overcome this limitations one uses a large number of evolutionary species (Rajendiran *et al.*, 2017). The method is not only based on examining close related species but also allows accurate determination at single nucleotide resolution levels (Rajendiran *et al.*, 2017).

Algorithms based on comparative genomics apply sequence comparison based on the genomes (Saçar Demirci *et al.*, 2017). Although these methods are widely used in computational identification of miRNA genes it is limited to organism that contains genome sequences (Saçar Demirci *et al.*, 2017). Hence, new technology such as machine learning based algorithm predictors can be applied.Machine learning is based on the ranking degree of similarity found on the features of miRNA (Gomes *et al.*, 2013). Hence, identifying both positive and negative dataset is crucial. Machine-learning methods used to predict novel miRNAs include support vector machine (SVM), Hidden Markov Model (HMM), and naïve Bayes classifier (Gomes *et al.*, 2013). Several tools that applies these methods include HHMMiR (Parveen *et al.*, 2020), MiRFinder (Yousef *et al.*, 2022), ProMiR (Akhtar *et al.*, 2016), and MiRRim (Akhtar *et al.*, 2016). Hence

these methods could be applied to identify microRNA genes in different organisms (Parveen *et al.*, 2020).

2.8.6 MicroRNA target

Understanding the interaction of miRNA to its target is quite crucial. The interaction of miRNA and the mRNA genes is a regulatory system conserved in animals, plants and even viruses (Riffo-Campos *et al.,* 2016). However, target complementarily differs from species to species. Identification of miRNA-target in plants is easier than in animals due to the nature of the miRNA gene interaction. Hence, this led to development of computational approaches to discover more miRNA targeted genes (Akhtar *et al.,* 2019; Roberts & Borchert, 2017).

2.8.6.1 Common features for MicroRNA target prediction tools.

Most of the miRNA target prediction features relays on miRNA:mRNA pairing, site location , conservation, site accessibility, multiple target site and the miRNA:mRNA expression profiles (Akhtar *et al.,* 2019). miRNA: mRNA pairing involves the binding of miRNA (5' end) and mRNA (3' UTR) sites. The seed region is located at positions 2-7 from 5' end of the miRNA. In addition this region, guides the miRiSC complex to target mRNA (Akhtar *et al.,* 2019).

The seed region is governed by the Watson-Crick pairing, where the region can be divide according to seed types (Carmel *et al.,* 2012). The seed types include 8mer, 7mer-m8, 7mer-A1, 6mer (Figure2.5).The 8mer has an exact match at position 2-8 nucleotide and has an adenosine (Kim *et al.,* 2017). 7mer-m8 has an exact match at position 2-8 nucleotide and is not followed by an adenosine (Kim *et al.,* 2017). 7mer-A1 has an exact match at position 2-7 and followed by an adenosine (Kim *et al.,* 2017).

However, the 6mer position has an exact match at position 2-7 nucleotide and does not contain an adenosine at the end (Kim *et al.,* 2017). Interestingly, miRNA:mRNA

pairing is also dependent on the type of nucleotide available if there is higher GC content at this region high energy consumption is required, which is not efficient for organism development (Carmel *et al.,* 2012).

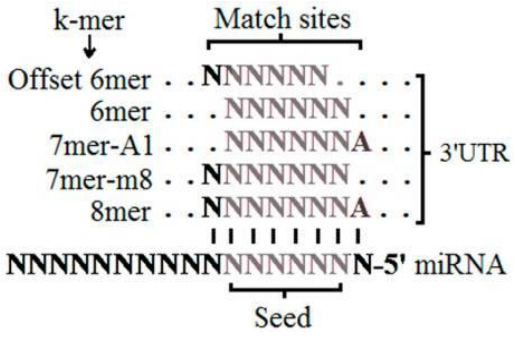


Figure 2.5: Seed match types and numbering system illustrated for MicroRNA.

Kmers positions in the miRNA are numbered 5′-3′. (Seed match 6 mer) Watson-Crick inverse complement of miRNA bases 2-7; (A1) presence of adenosine opposite miRNA base 1; (M8) Watson-Crick match to MicroRNA base source (Riffo-Campos *et al.,* 2016)

Also, the length of the mRNA depends on the binding at the 3' UTR region of the target genes (Akhtar *et al.,* 2019; Roberts & Borchert, 2017). Studies have shown that shorter length of the 3' UTR of the mRNA is believed to reduce the chance of the MicroRNA to bind to the gene.However, longer 3′UTRs sites, of the mRNA genes located at 15-20 nucleotides away from the stop codon increases the chances of miRNA binding to the mRNA (Roberts & Borchert, 2017). Studies have shown that miRNA can bind at the 5 UTR of the mRNA and the coding site of the genes (Roberts & Borchert, 2017).

Hence, correctly identifying potential target sites might reduce siRNA off target detection (Seok *et al.,* 2016). Effective target contains an A:U rich region that require minimum energy to bind (Gopinath *et al.,* 2017). Thus, site accessibility can assist in identifying miRNA target sites that are thermodynamically efficient. Therefore, filtering

putative targets based on thermodynamic stability provides sensitivity and efficient ways to develop miRNA target prediction tool (Akhtar *et al.,* 2019; Roberts & Borchert, 2017)..

2.8.6.2 MicroRNA target prediction tools

There has been an increasing demand of identifying targets for miRNA, which has led to both web based and non-web based approaches (Peterson *et al.,* 2014). Although most tools seek conserved 3' UTR sites with favorable thermodynamic hybridization energies and use of seed matches (Peterson *et al.,* 2014). New approaches such as machine learning techniques attempts to obey the rules of target site recognition from a small set of confirmed miRNA (Riolo *et al.,* 2020). Therefore, the success of the algorithms are determinate on the number of miRNA targets available (Riolo *et al.,* 2020). Hence, comparing different software and understanding their approaches to miRNA target prediction gives scientist the opportunity to choose the best tool for their analysis. The tools are categorized in different groups (Riolo *et al.,* 2020).

2.8.6.2.1 Seed-based approaches

Interestingly, this is the earliest approach for target prediction of miRNAs through applying computation based prediction of complimentary match at the first 8 nucleotide sequences of the miRNA to the 3 UTR of the mRNA (Fridrich *et al.,* 2019). Therefore, this led to discovery of multiple target sites for miRNA, which were validated through experimental analysis (Fridrich *et al.,* 2019) . The MiRanda method is an example of this algorithm that applies the seed based approach to identify miRNA target (Enright *et al.,* 2003). It applies three phases. Phase one includes the hybridization of miRNA and mRNA target gene, by allowing the G:U pairs and indels of considerable size. Phase two calculates the computing thermostabilty of the miRNA:target duplex. The final phase identifies evolutionary conservation of miRNA target across two different species (Oliveira *et al.,* 2017).

Target scan was the first method to explicitly use the concept of seed matches (Quillet *et al.,* 2020). The algorithm scans through a set of orthologous 3'UTR sequences from specific organisms and identifying the perfect complementarity between bases at position 2-8 of miRNA and the target site. The secondary structure of the heteroduplex is predicted by a folding algorithm. Then, a folding free energy value is calculated, and a z-score assigned based on the transcript for each organism. The cut-off values for ranking and the z score are given depending on the candidate sets. The addition of more organism led to the improvement of the ranking system and cut off system (Akhtar *et al.,* 2019).

DIANA-microT was used to predict targets in both human and *Mus musculus* (Paraskevopoulou *et al.,* 2013). The method relays on two hypothesis about MicroRNA:mRNA regulatory association:

1. Association should be structurally restrained due to enzymology of miRisc complex.

2. They should be conserved high-affinity interactions.

The first hypothesis led to the speculation that structural restraints might be reduced to set general rules. The second observation led to the computing of the thermodynamic stability of imperfect mRNA:mRNA pairings (Paraskevopoulou *et al.,* 2013). Hence, the observations resulted in the need to identify near perfect complementarity on the first few nucleotides at the beginning.

RNAhybrid applied the first single genome for identification of targets sites (Shaker *et al.,* 2020). The method is a dynamic programming algorithm that calculates the most energetically favorable hybridization for miRNA and the target. Hence the method allows the user to form perfect helix through determining the statistical significance of the predicted target and extreme value statistics for minimum folding energy of the target length. The poison distribution is used to determine the multiple binding sites of

the miRNA and target (Shaker *et al.,* 2020) . Hence, the statistical treatment is extended with comparative analysis of conserved binding site in orthologous target of related species.

Pictar is a tool that combines both near perfect seed search at the seven stretches of the miRNA and target genes and minimum free energy from heteroduplex complex formed (W *et al.,* 2014). The target site that pass both the filters are referred to as anchor and are ranked using the HMM and maximum likelihood score. Hence the score accounts for the synergistic effect of multiple binding sites for single miRNA and co regulated miRNA and reduces the false positive rank (W *et al.,* 2014).

2.8.6.2.2 Machine learning approaches

Machine learning algorithm predictors applies both positive and negative datasets to identify the binding between target genes and miRNA (Gomes *et al.,* 2013). TargetBoost is a machine learning algorithm that uses both genetic programming with boosting. The algorithm uses the hidden rules of miRNA:target site hybridization (Thakur *et al.,* 2022). The genetic programming component consists of evolving series of patterns sequences, which spawn together to describe properties of miRNA: target hybridization site namely the existence of a nucleus of consecutive paired bases or a bulge of unpaired nucleotides (Gomes *et al.,* 2013). The pattern sequence is referred to as a classifier, combined with the boosting technique; the classifier gives weight to be used as a training dataset. Another, machine learning approach is the use of support vector machine (SVM). miTarget applies SVM algorithm to consider the structural position of the features, and hybridization of the 5'and 3' hybridization site (Yerukala Sathipati & Ho, 2018). However, for SVMs to work well, they normally require a large negative training set, which is not currently available for miRNA targets (Yerukala Sathipati & Ho, 2018). Moreover, due to advancement in technology, Software such as miRaw, has been developed that applying the deep learning algorithm (Pla *et al.,* 2018).This technique does not apply the typical feature of miRNA:mRNA target hybridization site rather it

investigates the entire miRNA and 3'UTR mRNA nucleotides to learn an uninhibited set of feature descriptors related to the targeting process (Pla *et al.,* 2018). However, data availability is crucial for building reliable machine learning classifiers and reducing the false positive (Pla *et al.,* 2018).

2.8.7 The importance of Bioinformatics approaches

The motivation for choosing the algorithm (miRanda and RNAhybrid) as the software for miRNA target prediction is that they are the most frequently used target prediction algorithms and that they are open source which allow scientists to execute them locally and adapt them to different data sets and extract new self-defined features (Feng et al., 2018; Kumar, 2020). Moreover, combining algorithms with variations in prediction results due to different weightages of miRNA–mRNA interaction properties has also led to the reduction of the false positive results. For instance MiRanda is one of the earliest developed large-scale target prediction algorithms for vertebrates. The standard version of miRanda selects target genes based on three properties: sequence complementarity using a position-weighed local alignment algorithm, free energies of RNA-RNA duplexes using the Vienna RNA fold package, and conservation of targets in related genomes (Feng et al., 2018). These features are weighed in a decreasing order. In this application, only the first two filtering layers, i.e. sequence and energy scores are applied to restrict the predictions (Feng et al., 2018). RNAhybrid finds the energetically most favourable hybridization sites between miRNAs and their target mRNAs using integrated powerful statistical models (Feng et al., 2018). It takes candidate target sequences and a set of miRNAs and looks for energetically favourable binding sites (Feng et al., 2018). The RNAhybrid tool to find the minimum free energy hybridization and calculates the effective numbers across species is not performed (Feng et al., 2018). Hence, combining more than one algorithm reduces false positive and increases accuracy and precision.

CHAPTER THREE

MATERIAL AND METHODS

3.1 Study site

Experimental procedures of this study were carried out at the Molecular Biology and Bioinformatics Unit (MBBU), International Centre of Insect Ecology and Physiology (*icipe*), Nairobi, Kenya.

3.2 Ethical approval

It was not applicable in this study

3.3Insect rearing

Glossina pallidipes specimens used in the study were reared in the International Centre of Insect Ecology and Physiology (*icipe,* Kenya) insectary and Kenya Agricultural and Livestock Organization (KARLO) insectary in Nairobi, Kenya under favourable conditions of 75-90% humidity and 23°C ±1°C. The non-teneral adults (both males and females) and gravid females were fed once on defibrinated bovine blood on an *in vitro* membrane feeding system and then starved two days before RNA extraction. The third instar larvae were collected immediately after larviposition and frozen at -80°C to prevent the larvae from transitioning to pupae. Other larvae were left for two days, which were then collected as two-day-old pupae. The new emerged flies (teneral flies) both male and female were captured and immobilised at 4°C for 10 minutes, the whole body was used for total RNA extraction. For the mature adult flies (non-teneral flies), both sexes were captured at 8-days old and immobilised at 4°C for 10 minutes, for total RNA extraction.

3.3.1 Female tsetse fly dissection

The 8 day old gravid females carrying fully developed third instar larvae were observed visually and selected for dissection. They were immobilized by placing them in the refrigerator at 4°C for 10 minutes. Each uterus was dissected under a microscope, and the larvae were carefully removed to maintain the integrity of the uterus of the female specimens (Attardo *et al.,* 2014). The carcasses of the gravid females were used for total RNA extraction.

3.3.2 Study Design

The tsetse flies at different stages of their development, namely third instar larvae, 2 day old pupae, new emerged teneral adults (females and males), 8 days old non teneral adult (females and males) and 8 day old gravid females with a minimum sample size of three numbers in each group were collected and pooled as one sample. The tsetse flies specimens were collected a minimum of three times during different insect rearing cycle for each stage of development. Collecting specimens from different cycles minimized possible altered rearing condition and hence served as the biological replicates. Although, flies found in the field are known to be infected with trypanosomes. The study did not utilize these symptomatic flies because it was reviewing the biological aspects of microRNAs, focusing on their roles as regulators of gene expression during development of *G. pallidipes* species and not their implication in immune responses.

3.4 RNA isolation and small RNA library preparation

Total RNA was extracted from 21 samples using Trizol® Reagent as per the manual's protocol. The quanity of the total RNA was measured using Eppendorf Biospectrometer® fluoresces Germany by the ratio of OD260 and OD280, the value of all samples ranged from 1.8 to 2.2 indicating good quanity of the RNA. RNA degradation and contamination were monitored on 1% agarose gels Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). RNA quality assessment

was determined by using Agilent 2100 Bioanalyzer RNA Nano 6000 kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Total RNA with an RIN numbers of 7.5-8.9 were used for sequencing of sRNA libraries. The sRNA fragments of <40 bp were separated by polyacrylamide gel. The sRNA fragments of <40 bp were separated on polyacrylamide gels. Subsequently, the sRNA fragments were ligated with 3′ RNA adaptors and used to prepare libraries that were sequenced using Illumina sequencing platform Hiseq 2500 (Macrogen, South Korea). Raw sequencing data were submitted to NCBI Sequence Read Archives (SRA) under Bioproject accession number PRJNA590626.

3.5 Sequence data processing and analysis

3.5.1 Bioinformatics

The sequence read data was analysed using a pipeline based on miRNA seq analysis(Figure 3.1). The pipeline was designed with distinct phases: (1) Quality control;(2) Filtering other type of small RNA (tRNA, rRNA, mRNA);(3) Identification of miRNA genes and expression profiling;(4) Functional analysis.

3.5.1.1 Quality control

The sequence read data (fastq files) obtained from 24 *G. pallidipes* samples were exposed to standard Illumina quality control procedures. The sequence adaptors and low quality reads were filtered by two software namely: cutadapt (v4.1) (Martin, 2011) and FASTX toolkit (v0.0.13) (http://hannonlab.cshl.edu/fastx_toolkit/), with the following parameters for adapter trimming: **"**-e 0.80, –m 15, --match-read-wildcards". Option –e 0.80 retained the sequence with quality greater than 80% of the total phred value from each nucleotide. The length of the sequence to be kept was defined by option'-m 15' which corresponded to the nucleotides of a read sequence. While '-match-read-mismatch' allows the letters that are not the common nucleotide (AGCT) to be considered in the adaptor sequence during trimming.

3.5.1.2 Mapping the reads to the reference genome

All obtained reads were mapped to the *G. pallidipes* genome which was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genome/). Sequence reads were mapped to the genome using miRDeep2 (Friedländer *et al.*, 2012). Briefly, the mapper module was used first to test the format of the input files (the FASTQ files from each library and the reference genome file). The sequencing file was then converted into a FASTA format. To ensure there was no redundancy, identical sequences were collapsed to have a unique read. The number displayed in the new FASTA identifiers determined the quantity of the corresponding sequence in the data set. Reads were mapped to the reference genome (*G. pallidipes* genome) with bowtie, utilizing the following parameters; bowtie '-f ','-n 1', '-e 80','-l17','-a','-m5','-best -strata'. The alignments of the reads to the genome retained only a mismatch or none at the seed region as illustrated by the option '-n 1'. Option '-l 17' ensured that the length of the sequence was not below 17 nucleotide long. The option '-e 80' determines the maximum sum of the quality values at each mismatch position. For the 'non-seed region', the parameter was kept at a default of 40, allowing two mismatches in the regions of the read to the genome. Only reads that do not map more than five times to the genome were accepted as per the parameter option '-m5'. Option'–best –strata' ordered the mappings from best to worse alignments according to the strata definition of bowtie. If mappings with zero mismatches occurred, then mappings with one or two mismatches were not reported. All the processed reads and the mappings to the genome were outputted in miRDeep2 format.

3.5.1.3 Small non-coding RNA annotation

The *G. pallidipes* reference genome was annotated using small non-coding RNA from RFAM database v14.0 (E. P. Nawrocki *et al.*, 2015) by the cmscan algorithm from the infernal software packages v1.1.2 (Eric P Nawrocki & Eddy, 2013). First, the cmbuid program used sequences from RFAM database to build covariance models from our reference genome (Eric P Nawrocki & Eddy, 2013). This was followed by compressing

the huge covariance file using cmpress algorithm. The reference genome was annotated by cmscan using the following options: --cut_ga --nohmmonly option "–cut_ga" define what gets included in Rfam full alignments based on searches with Rfam seed models. The option –nohmmonly turns off the accelerated HMM banded alignment strategy. Once the genome was annotated , the clean reads were mapped to the reference genome (*G. pallidipes* genome) with bowtie, using the following options: bowtie -f -n 1 -e 80 -l 17-a -m 5 –best –strata. Finally, we used BEDToolsv2.27.1 (Quinlan & Hall, 2010) to determine the overlapping small non-coding RNAs in sequence libraries.

3.5.1.4 Identification of known and novel MicroRNAs

For identication of known and novel miRNAs present in the eight datasets, the miRDeep2.pl algorithm was used. The input files for miRDeep2 script were as following: (i) a FASTA file with deep sequencing reads from each stage; (ii) a FASTA file of the reference genome which is a file of mapped reads to the genome in miRDeep2 format, and (iii) a FASTA file of known miRNAs in all species. After testing the format of all input files, a fast quantification of known miRNAs was done. Potential miRNA precursors were excised from the genome using the genomic coordinates of the mapped miRNAs as guidelines. The genomic coordinates of the mapped miRNAs were first parsed such that only perfect mappings (no mismatches) of at least 18 nucleotides were retained. Furthermore, the genomic coordinates of the mapped miRNAs from reads that map perfectly in more than 5 loci in the genome were discarded.

The two genome strands of each genome contig were scanned separately, from the 5′ to 3′ end. Excision was initiated when a stack of reads (height one or more) was encountered. However, if there was a higher read stack within 70 nucleotides downstream of the current read stack, then this was chosen instead. This downstream search was iterated until no higher read stack was found within 70 nucleotides. In this way, the highest local read stack was identified. The sequence covered by the highest local read stack was excised twice, once including 70 nucleotides upstream and 20

nucleotides downstream flanking sequence, and once including 20 nucleotides upstream and 70 nucleotides downstream flanking sequence.

Subsequently, the genome scanning continues from the position one nucleotide downstream of the last excised sequence. If the total number of potential precursor sequences excised was less than 50000 (two precursors per genomic locus), then this set was output to the downstream analysis. If there were more sequences, then the entire excision step was repeated, with the height of the read stack necessary for initiating excision increased by one. The bowtie-build tool was used with default options to build a bowtie index of the excised potential precursors. The set of sequencing reads were mapped to the index, using bowtie with the following options: bowtie -f -v 1 -a –best –strata –norc. Option '-f' designates a FASTA file as input, option '-v 1' reports the genomic coordinates of the mapped miRNAs with up to one mismatch to the precursors, option '-a' lead to the report of all valid alignments, options '–best –strata' ordered the mappings from best to worse alignments according to the strata definition of bowtie. If reads map perfectly to the precursors, then mappings of the same read with one mismatch were not reported. Option'–norc' specifies that no reads must be mapped to the reverse complement of the precursor sequences in the bowtie index.

The set of known mature miRNAs for the reference species was also mapped to the index, with the following options: bowtie -f -v 0 -a –best –strata –norc. Here the module did not allow any mismatches between the mature miRNA sequence and the potential precursor. The two mapping files were concatenated and all lines were sorted according to the potential precursor ID's. The next step was to predict RNA secondary structures of the potential precursors. This was done with randfold with default options. Optionally, the randfold P-values for a subset of the potential precursors were calculated. This was done by selecting the potential precursors that (i) fold into an unbifurcated hairpin, (ii) can be partitioned into candidate mature, loop and star part based on the genomic coordinates of the mapped miRNAs to it, and (iii) have minimum of 60% of the nucleotides in the candidate mature part base paired.

The randfold p-values were calculated for the subset of potential precursors with these options: randfold -s 99. In the next step, the potential precursors were individually scored or discarded by the miRDeep2 core algorithm. The core algorithm ran with these options: -s -v -50 -y. Option '-s' designates the reference mature miRNAs file in FASTA format as input, option '-v -50' keeps all precursors that have a miRDeep2 score above -50 and option '-y' supplies an additional file with randfold values. Furthermore, 100 rounds of permuted controls were performed as previously described, with the same options as the genuine run. The third step surveyed the score distributions of the genuine run and the control runs. The performance statistics were calculated for all score cut-offs from -10 to 10. The number of known miRNAs present in the data was estimated as the number of known mature miRNAs that map perfectly to one or more excised potential precursors. The number of known miRNAs that were recovered was estimated as the number of known mature miRNAs that map perfectly to one or more hairpins that exceed the given score cut-off. The sensitivity of the run was estimated at se = (known MicroRNAs recovered)/(known miRNAs in data). The number of false positives for a given score cut-off was estimated by the permuted controls. The fraction of true MicroRNAs reported was estimated by t = (novel miRNAs – estimated false-positive novel miRNAs)/novel miRNAs. The signal-to-noise ratio was estimated as n = total miRNAs/estimated total false-positive novel miRNAs (total miRNAs = novel miRNAs + known miRNAs) (Friedländer *et al.,* 2012).

Finally, the miRDeep2 module integrated all these results in a .html file as well as a corresponding tab-separated file. The tab-separated file contained detailed information on every known and novel MicroRNA in the data. In the top of the .html file was a survey of miRDeep2 performance for varying score cut-offs. For each score cut-off the sensitivity and number of true positive novel miRNAs was estimated.

3.5.1.5 MicroRNA differential expression analysis

Read counts generated from the quantifier module in miRDeep2 was used for statistical analysis by Perseus software (v1.6.2.3**)** (Tyanova *et al.,* 2016). The read counts were first transformed into Log2 for quantification analysis; at least three valid counts were used in all the groups. The read count with missing values (zero) were imputed with random numbers from normal distribution using the mean and standard deviation with a noise level (width = 0.3, down shift = 1.8). The Pearson correlation coefficients were calculated within the three biological replicates in each life stage. However, we had to exclude one replicate from the analysis due to low Pearson correlation ($R < 0.70$). Subsequently, we identified differentially expressed genes using edgeR (Empirical Analysis of Digital Gene Expression Data in R; v3.16.5) package (Robinson *et al.,* 2010). The normalised mean values of the miRNA genes were used to determine their abundance in each consecutive life stages. The normalised values were in form counts per millions of the total number of reads (CPM), miRNA with less than ten CPM were filtered from the analysis. The miRNA with a false discovery rate (FDR) adjusted p-value of ≤ 0.05 and the fold change of log2 scale of > 1 or < -1 were determined as either up-regulated or down-regulated in the consecutive life stages

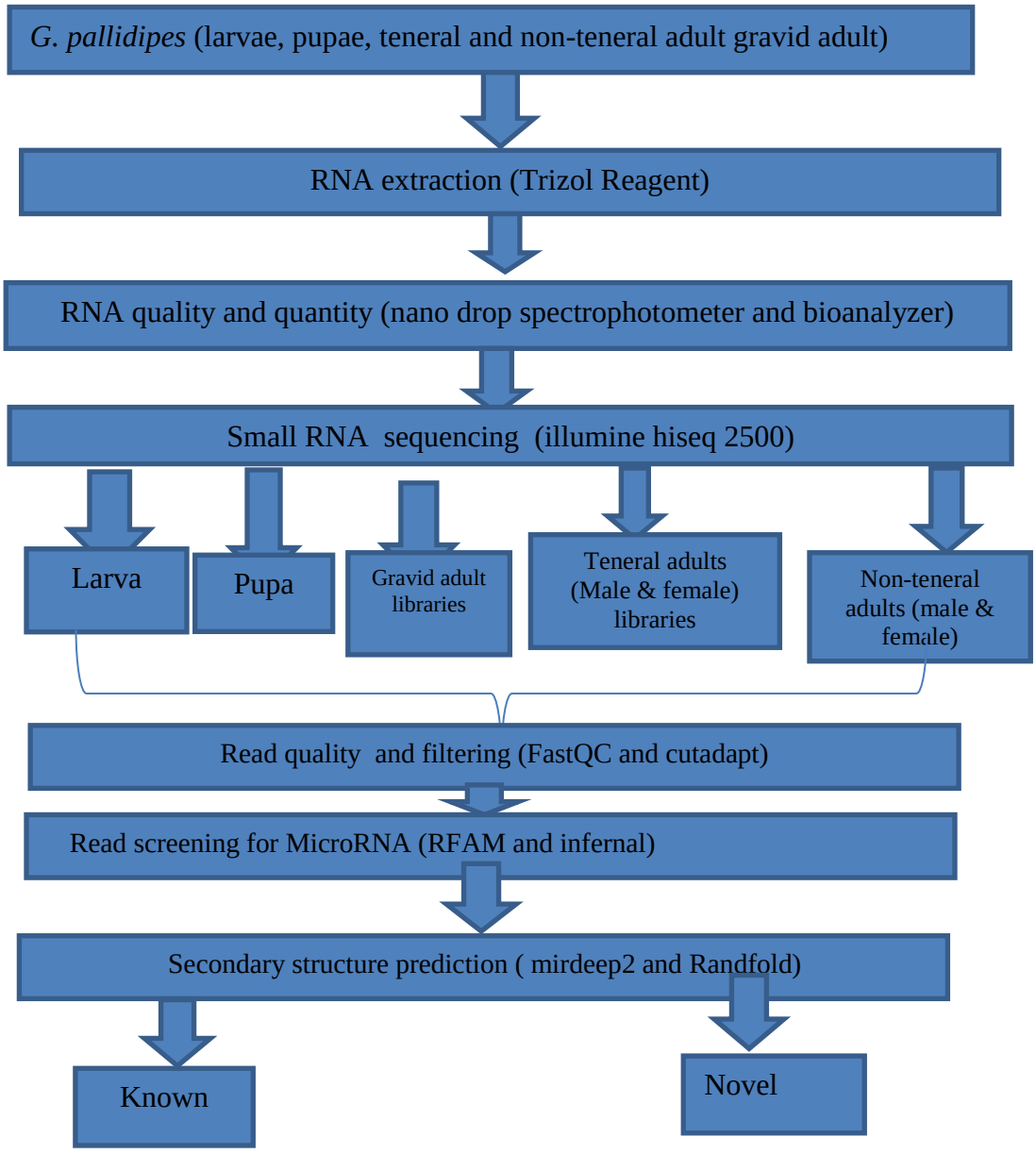


Figure 3.1: Schematic overview of analysis pipeline for identification and charcterisation of *G. pallidipes* microRNAs.

The experimental work (tsetse rearing, RNA extraction, quality, and quantity) was conducted in insectary and Molecular Biology and Bioinformatics Unit (MBBU), International Centre of Insect Physiology and Ecology (*icipe*), Kenya. Small RNA sequencing was done by Macrogen Inc, South Korea and the computational analysis

(read quality, filtering mapping, and identification of know and novel miRNAs) was performed at MBBU, *icipe*.

3.5.1.6 MicroRNA target prediction and functional analysis

The potential candidate target genes for the differentially expressed miRNAs were identified by RNAhybrid (v2.1.2) (Krüger & Rehmsmeier, 2006) and MiRanda (v 3.3a) (Riffo-Campos *et al.,* 2016). The *G. pallidipes* miRNA genes were utilised to determine potential targets from the orthologous genes of *Glossina morsitans morsitans* 3' UTR sequences obtained from Vectorbase (https://www.vectorbase.org/). RNAhybrid software (v2.1.2) was set with the parameters: miRNA: mRNA binding energy <-14 kcal/mol, the p-value was set at 0.05 and > 6 nucleotide match at the seed region. MiRanda software (v 3.3a) as used with the parameters: miRNA: mRNA binding energy value < -14 kcal mol and a score of ≥ 100. The parameters were slightly modified to fit the data set available. The potential candidates, identified by both software, were selected for further analysis. The Gene Ontology (GO) enrichment studies and pathway annotation of the selected candidate genes were performed using Blast2GO version 5.2.5 (Conesa *et al.,* 2005).

CHAPTER FOUR

RESULTS

4.1 Pre-processing of short reads

4.1.1 Sequence quality of the five libraries

Applying next generation sequencing technology, a total of 43.9, 43.6, 41.4 and 38.5 million raw reads were obtained from eggs, larvae, pupae, and adults, respectively (Table 4.1). More than 96% and 91% of the reads had Phred quality values (PQV) of 20 and 30 respectively. The PQV determines the quality of the reads through demonstrating the base call accuracy value of above 20 (Ohta *et al.,* 2017). The length distributions of the reads were 101 nucleotides long. After removing the adapters and poor quality reads the length reduced to 15-35 nucleotide long (Figure 4.1), a total of 5.65 million, 8.13 million, 6.12 million, 5.19 million, 8.88 million, 7.18 million, 9.29 million, high quality reads were retained in larva, pupa, teneral female adult, teneral male adult, non-teneral female adult, non-teneral male adult, gravid adult, respectively (Table 4.1).

4.1.2 Mapping reads from the eight libraries

Characterization of miRNA genes in NGS data start with the first step of aligning the good quality reads to the reference genome (*G. palliidpes*). Furthermore, mapping the reads to the reference genome demonstrates an unbiased option, allowing for identification of both known and still undiscovered miRNA. The total number of mapped reads constitutes only 38.54%, 30.96%, 38.04%, 29.46%, 47.33%, 34.50%, 57.90% of the total high-quality reads from larva, pupa, teneral female adult, teneral male adult, non-teneral female adult, non-teneral male adult, and gravid adult, respectively.

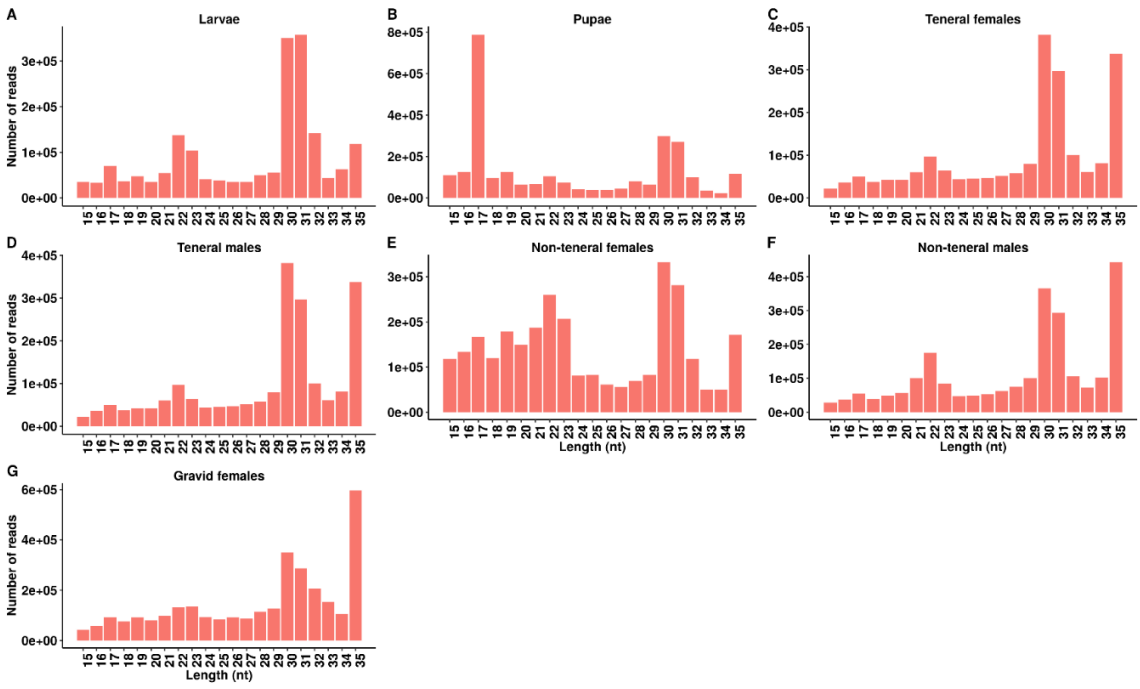


Figure 4.1: The length distribution of *G. pallidipes* small RNA sequences pooled in (A) larvae libraries, (B) pupae libraries, (C) Teneral females libraries, (D) Teneral males libraries, (E) Non-teneral females libraries, (F) Non-teneral males libraries and (G) Gravid females libraries. The Y-axis represent the absolute number of reads (x 100,000) in each library and X-axis represents the length of the reads in nucleotides.

Table 4.1: The table summarises the information about our libraries

Stage	Sexes	Replicates	Total raw counts	Reads post-adaptor trimming	High-quality reads (18-35)	Mapped to the genome
Larvae		3	20,361,758	19,691,042(96.71%)	5,652,804 (28.71%)	2,178,368(38.54%)
Pupae		3	23,201,822	20,262,789(87.33%)	8,128,822(40.60%)	2,516,749(30.96%)
Teneral adults	Female	3	20,989,380	20,019,810(95.38%)	6,119,010(30.56%)	2,327,379(38.04%)
Non - teneral adults	Female	3	21,971,711	19,918,765(90.66%)	8,879,712(44.58%)	4,203,245(47.33%)
	Male	3	19,771,800	19,058,192(96.91%)	7,184,851(37.70%)	2,478,727(34.50%)
Gravid	Female	3	22,267,407	19,710,796(88.52%)	9,298,295(47.17%)	5,271,811(56.70%)
Total		21	149,478,030	139,538,937 (93.35%)	50,460,899(36.16%)	20,507,417 (40.60%)

The quality of the library and the percentage of alignment on to the genome and provides statistical summary.

4.1.3 Annotation of small ncRNAs in eight libraries

To annotate small ncRNA in the eight libraries, all clean reads were aligned to the annotated reference genome. Hence, to quantify the reads, statistical analysis was conducted using R software programming. Due to the lack of rRNA depletion, as shown in Figure4.2, the most abundant class in all the libraries were the rRNA genes. However, the most abundant miRNAs genes were observed in the non-teneral female library, while the tRNAs were abundant in the pupa library.

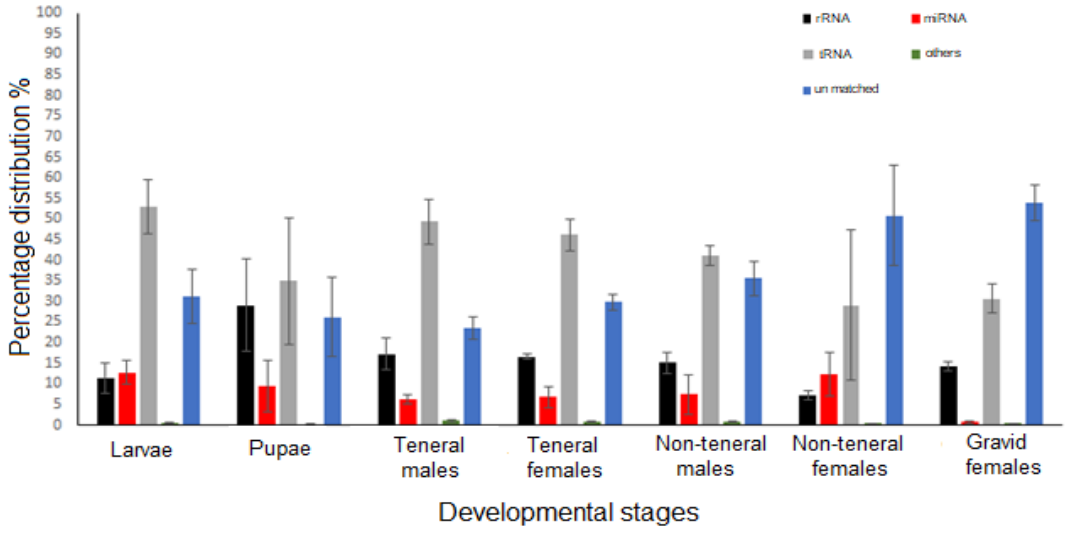


Figure 4.2: The small non-coding RNA distribution from Rfam database in the different developmental stages.

Others and unmatched represents other types of small non-coding RNA establish in Rfam and unaligned reads from the libraries respectively.

4.2 Identification of known microRNAs in the eight libraries

The 56 known *G. pallidipes* miRNAs in the Meki *et al*, 2018 study were detected in the sequenced short reads (Appendix III). For all characterized miRNAs, the full precursor structure (mature, loop and star sequence) or parts of it were detected.

4.3 Detection of MicroRNA and its isoforms

We computationally characterised a total of 157 miRNA genes. Out of 157, we identified 148, 137, 144, 139, 137, 156 and 141 miRNA genes in larvae, pupae, teneral female adults, teneral male adults, non-teneral adult males, non-teneral adult females, and gravid adult females, respectively. These included 99 miRNA genes already computationally predicted in a recent study by Meki and colleagues (Meki *et al.*, 2018) and annotated in the *G. pallidipes* genome hosted by vectorbase (Megy *et al.*, 2012). In

addition, a total of 47 mature miRNA gene candidates were identified as orthologues from other insects. However, eight of the 47 miRNA genes (miR-375-3p, miR-7-5p, miR-957-3p, miR-970-3p, miR-980-3p, miR-993-3p, miR- 306-3p, miR-307-3p) were conserved in more than one insect species (*Drosophila melanogaster*, *Anopheles gambiae*, *Aedes aegypti*, *Apis melifera*, and *Bombyx mori*) from miRBase. Many pre-miRNAs have been reported to yield two kinds of mature miRNAs, although the two products, miR-#-5p and miR-#-3p, may vary in expression levels. Employing information from miRBase, we were able to identify computational evidence based on the 5p- and 3p- associated read abundance for *G .pallidipes* miRNA except for two miRNA (miR-2521, miR-2760). The 143 miRNA genes revealed that both miR-#-5p and miR-#-3p were conserved (Appendix2), with most sites showing minimal variation in nucleotide composition. Herein, 20 dominant miRNAs were identified as 5p-miRNA from 5p-arm.Whereas others (40 miRNAs) were identified as 3p-MicroRNA were observed to have abundant expression profile (number of miRNA genes, which was represented as 3p-arm (Appendix II)

4.4 MicroRNA expression profiles

The expression patterns of miRNAs provide clues of their functions (Yao *et al.*, 2007). To obtain insight into possible stage-dependent and sex-dependent roles of miRNAs in *G. pallidipes*, the expression patterns of miRNAs in different developmental stages including

larva, pupa, and teneral adult, non-teneral adult and gravid adult were examined based on the number of reads obtained. The heat maps summarize the expression of the known and novel microRNAs in the five developmental stage libraries. The majority of miRNAs were sequenced between 1-200 times (Figure4.3). Altogether, 99 known miRNAs were present at more than five counts for at least one developmental stage. However, we did not utilize RT-PCR for validation of the 157 miRNA genes.

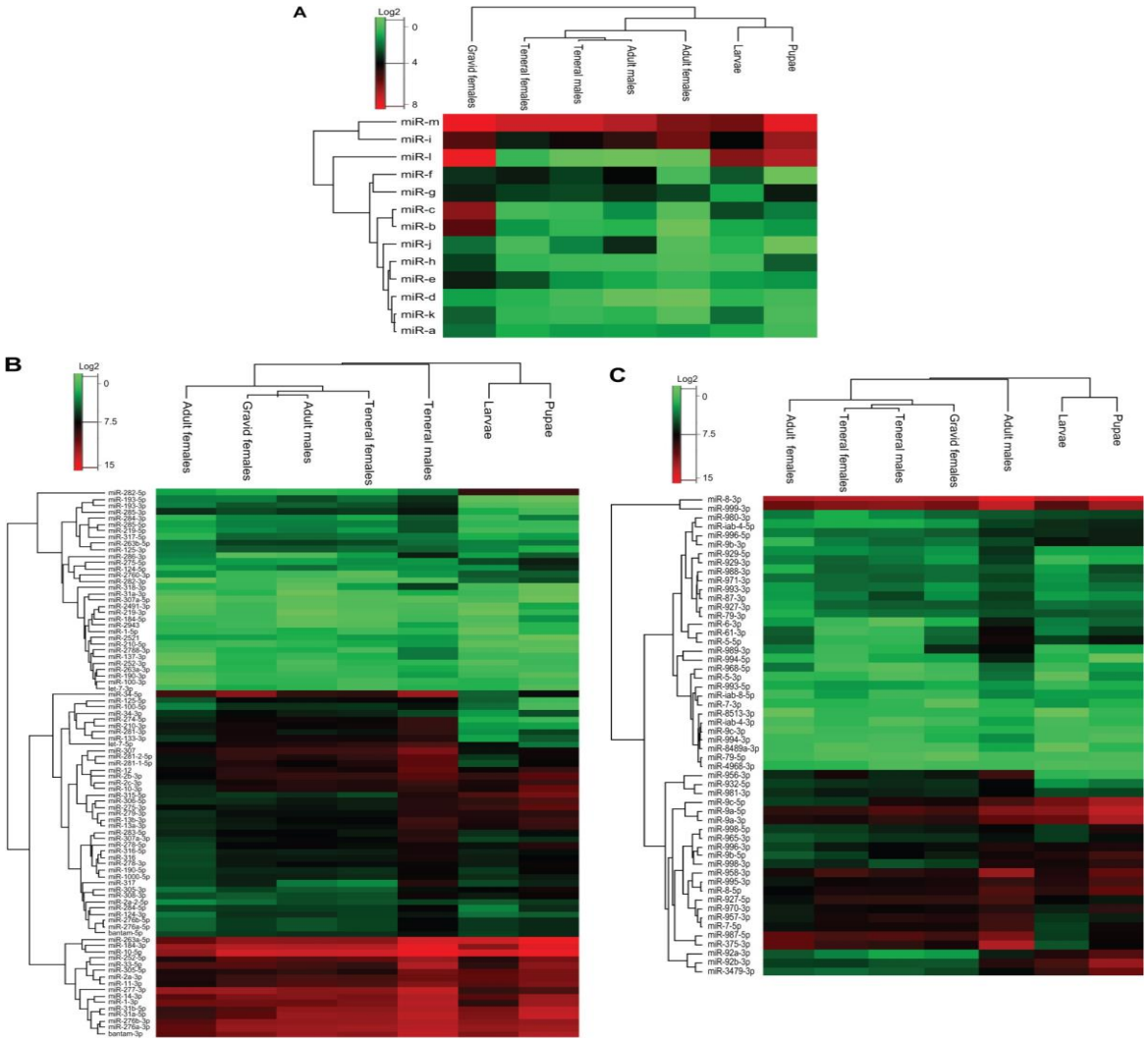


Figure 4.3: Heat maps clustering of miRNAs expressed in the five developmental stage libraries of *G. pallidipes*.

The clustering was performed on all (panel a), (panel b) and (panel c) MicroRNAs based on the log_2 transformation of the read counts of miRNA across the developmental stages. The miRNA clustering is shown on top. The colour scale (shown on the left and the middle) illustrates the log_2 transformation value (sequencing frequency) and the five development stage samples and their sexes. Each row represents a stage and each column represents a miRNA gene.

4.4.1. Sex-specific miRNA expression profiles

The newly emerged flies (teneral flies) had only three miR-5-5p, miR-998-3p and miR-3-3p up regulated in the teneral female adult compared to the teneral male adult and only one miR-87-3p and down regulated in teneral female compared to teneral male as demonstrated in Figure 4.4 A. Both gender during the mature adult stage (non-teneral stage) had significant miRNA being expressed miR-286-5p, miR-5-5p, miR-5-3p, miR-6-3p, miR-309-3p, miR-3-3p, miR-318-5p, miR-31b-5p, miR-33-5p, miR-3479-3p, miR-375-3p, miR-375-5p, miR-8-3p, miR-92a-3p,miR-92b-3p, miR-968-5p, miR-989-3p, miR-994-5p, miR-9c-5p, miR-h, miR-I, miR-iab-4-5p, and miR-j were up regulated in the non-teneral females (Figure 4.4 B). Seven miRNAs, namely miR-133-3p, miR-125-5p, miR-125-3p, miR-34-5p, miR-34-3p, miR-317-5p, miR-263b-5p, and miR-g were significantly down-regulated in the non-teneral females compared to the non-teneral males (Figure 4.4 B).

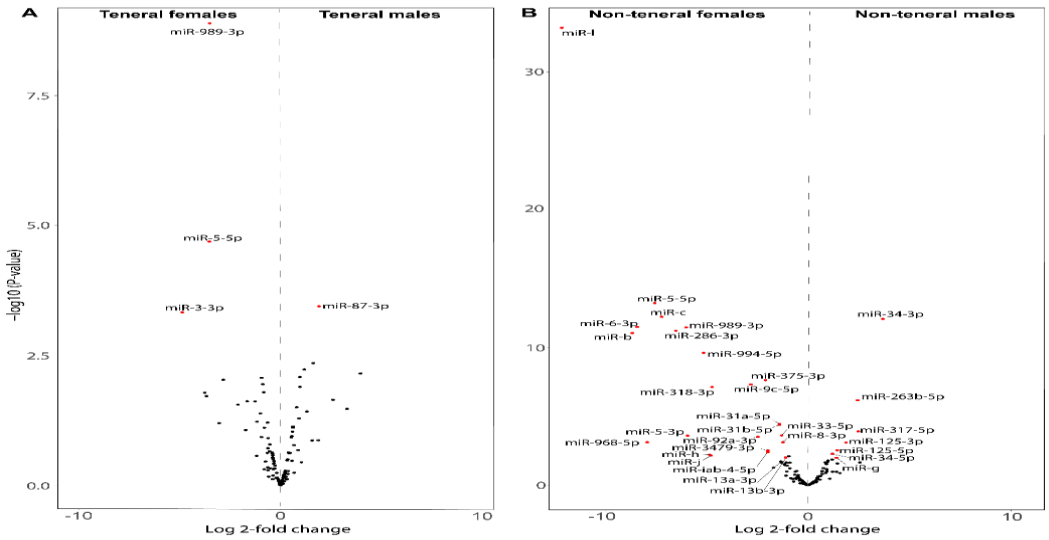


Figure 4.4: Volcano plots of statistically differentially regulated *G. Pallidipes* sex-specific miRNAs.

In the volcano plots, the results demonstrate the measure of differences (Log2 of the fold change (FC), x-axis) and significance (shows the −Log10 p-values (p), y-axis) expression profile of miRNA genes in different sexes of *G. pallidipes*. Red dots represent significant miRNA features at p-value < 0.05 and −1 > FC > +1.

4.4.2 Stage-specific MicroRNA expression profile

The pupation process revealed only four significantly expressed miRNA, three miRNAs, let-7-5p, miR-100-5p and miR-125-5p up-regulated in the pupal stage (Figure 4.5 A). One novel miRNA (miR-i) was down regulated in the pupal stage. The metamorphosis of pupa to teneral female adults showed the most significant shift in miRNA expression. We identified 18 miRNAs that were highly expressed in the teneral female stage compared to the pupal stage and 32 miRNA genes that were down regulated in the teneral female stage compared to the pupal stage (Figure 4.5B). In addition, the transition of pupae to teneral male adults was associated with the up-regulation of 24 miRNA and the down regulation of 28 miRNA (Figure 4.5 C). The transition from newly emerged (teneral) flies into a mature (non- teneral) flies was associated with41 miRNA genes significantly expressed, 29 miRNA genes were up regulated in non-teneral stage and 12 miRNA genes were down regulated in the non-teneral female stage (Fig 4.5 D). In contrast, 11 miRNA genes were

differentially expressed in the transition of teneral male adults and to non-teneral male adults; five were up regulated and in non-teneral male adults and six MiRNA were down regulated in non-teneral male adults compared to the teneral male flies (Figure 4.5 F). We identified significantly expressed miRNA genes during the transition of the female specimens from non-gravid to gravid state. 25 genes were significantly up regulated and 18 MiRNA genes that were down regulated in the gravid females compared to non-teneral females.

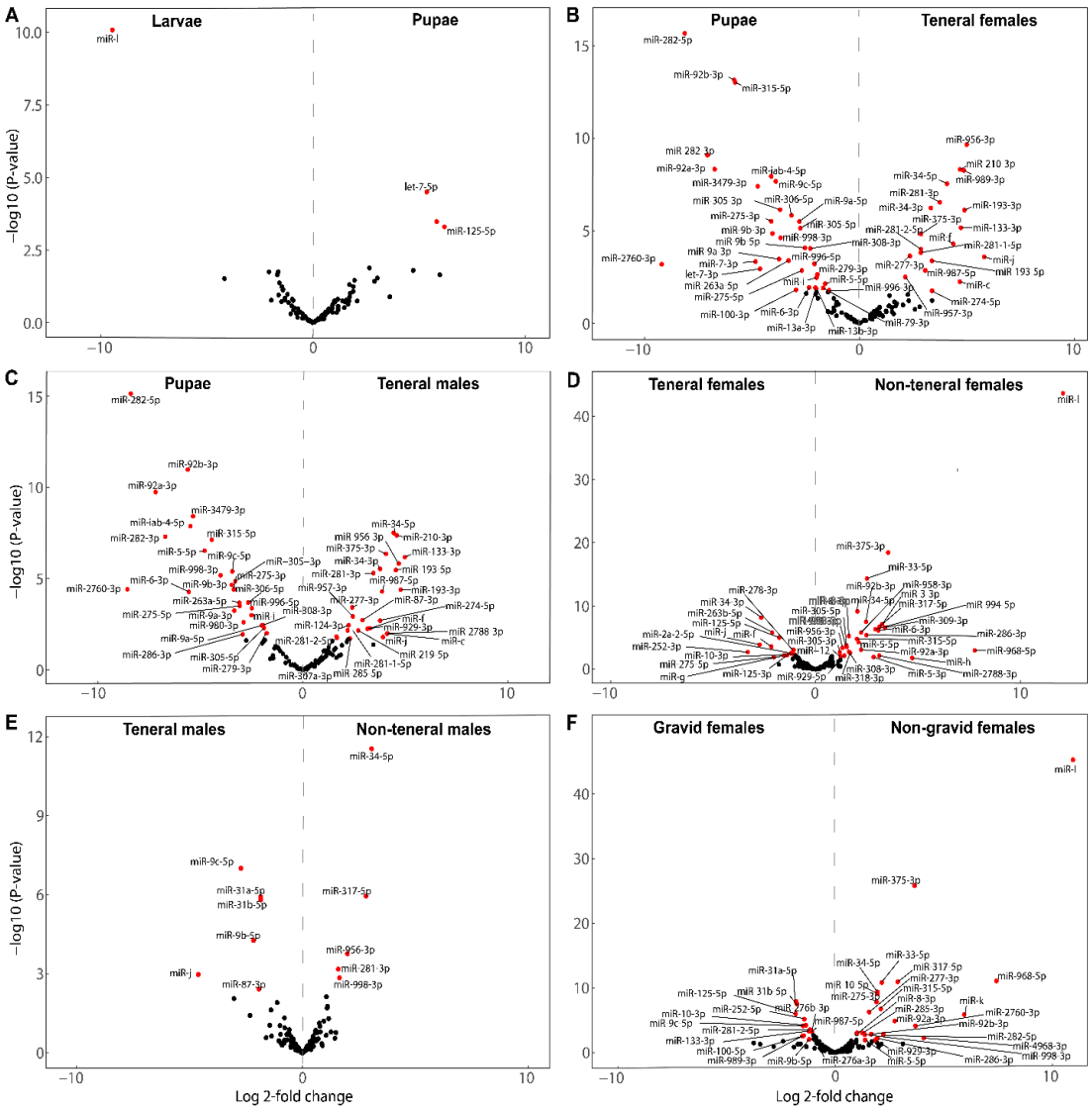


Figure 4.5: Volcano plots of statistically differentially regulated G. pallidipes stage-specific miRNAs.

In the volcano plots, the results demonstrate the measure of differences (Log2 of the fold change (FC), x-axis) and significance (shows the −Log10 p-values (p), y-axis) expression profiles for miRNA genes from the different consecutive life stage. Red dots represent significant miRNA feature pFDR< 0.05 and −1> FC > +1.

4.5 Target prediction and enrichment pathways

The putative target genes identified from 93 differentially expressed miRNA genes included 5550 genes, further analysis using Blast2GO yield KEGG pathways such as Purine metabolism (n = 183), Thiamine metabolism (n = 163), Biosynthesis of antibiotics (n = 55), Aminoacyl-tRNA biosynthesis (n = 16), Pyruvate metabolism (n = 13), Pyrimidine metabolism (n = 12), Glycerophospholipid metabolism (n = 12), Amino sugar and nucleotide sugar metabolism (n = 11) and cysteine and methionine metabolism (n = 11) (Figure 4.6). In addition, the top 10 gene ontology terms , 'cellular process' (n = 1464), 'metabolic process' (n = 1383), 'cell part' (n = 790), 'organelle' (n = 609), 'membrane' (n = 588), 'biological regulation' (n = 467), 'localization' (n = 398), 'response to stimulus' (n= 285), 'cellular component organization or biogenesis' (n = 222), and 'signalling' (n = 218) in (Fig 4.6). We developed eight miRNA:mRNA networks (Appendix IV & V) to demonstrate their interaction at different stages of development and sexes. The number of targets for each network varied depending on the number of miRNA genes. These networks also indicated that several miRNA were predicted to target more than one mRNAs gene transcript such as miR-34-5p, miR-125-5p, miR-100-5p, and miR-275-3p targeted 6, 13, 19 and 11 mRNA transcripts, respectively (Appendix V). Conversely, the mRNA transcripts GMOY007199, GMOY001123, and GMOY011632 had in silico binding sites for more than one miRNA gene. For instance, the mRNA GMOY007199, a nucleoporin 160 kD is targeted by three miRNA genes, miR-100-5p, let-7-5p, and miR-I (Appendix IV).

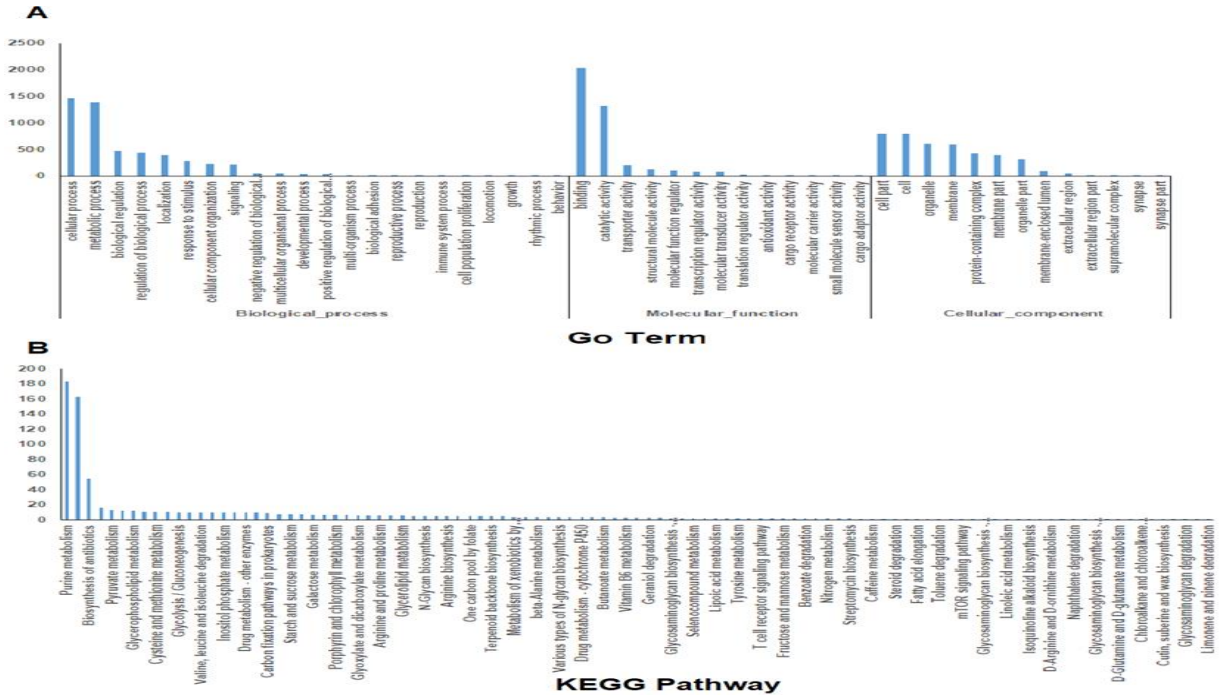


Figure 4.6: Go and KEGG pathway analysis of the 5550 target genes. (A) The GO terms for all the target genes: The y axis represents gene counts and the x axis represents GO term annotation. (B) The KEGG pathways annotation; the y axis represents the gene counts in number value and the x axis represents the KEGG pathway definitions identified by Blast2GO.

CHAPTER FIVE

DISCUSSION

G. pallidipes is an important vector of trypanosome parasite and is distributed throughout the sub-Saharan region (Cayla *et al.,* 2019; Mbewe *et al.,* 2018; Saini *et al.,* 2017). Various strategies were planned to control trypanomiasis spread, few of which were directed towards controlling the expansion of vector population (Demirbas-Uzel *et al.,* 2018). Such strategies require detailed understanding of tsetse biology pertaining to its developmental stages. Tsetse have a complete metamorphic cycle goes through four stages of development (Geoffrey M. Attardo *et al.,* 2019). All four stages show characteristic morphological and physiological differences, resulting from differential gene expression at each stage of development (Geoffrey M. Attardo *et al.,* 2019). To understand molecular mechanisms mediating such differences, we studied the stage specific differences in miRNA expression which are known regulators of gene expression.

Currently, no tsetse miRNAs predicted has been deposited in the current miRBase release (Kozomara & Griffiths-Jones, 2014). However, studies such as Meki et al 2018 and Aksoy et al 2016 have been able to characterize miRNA in the *Glossinidae* species (Aksoy *et al.,* 2016; Meki *et al.,* 2018). Among these known 99 miRNA genes were computationally predicted and 48 novel miRNAs were identified in this study using a high-throughput sequencing approach.This approach is more powerful than other conventional technologies previously used in other insects (Jain *et al.,* 2015), as it is able to identify new miRNAs, which are beyond the capabilities of older traditional methods. The genes identified as miRNAs had to, contain mature, loop and star strands and fit the parameters set in miRDeep2 software to be confirmed as a miRNA gene (Friedländer et al., 2012).

In the study, approximately 50 million high quality reads from each developmental stage were obtained by deep sequencing. The size distribution of the sequence reads peaked between 18-35 nucleotides. In addition, using similar deep sequencing techniques, other studies on insects showed similar peaks in sRNA genes (Asgari, 2013; Zhao *et al.*, 2015). However, the high-quality reads, mapped to the reference genome at low percentages. Such low mapping percentages may be associated with to the short nature of the reads and the high complexity of un-masked reference genome as observed in other studies (Allam et al., 2016; Calla & Geib, 2015; Cordero et al., 2012; Feng, Wu, et al., 2018). Further studies are needed to understand to what extent the 30-35 nucleotides class representing siRNA populations (the most abundant and diverse class of small ncRNAs sequenced in small RNA libraries that masked the abundance of miRNA populations in our libraries) (Kozomara *et al.,* 2019).

Subsequently, miRNA genes were predominantly expressed in at least one specific developmental cycle stages such as *miR-263a-3p,* Which was highly expressed in the larva libraries (> 6e+04 counts), and *miR-8* was highly expressed in the non-teneral female adults library (>1.6e+05 reads). MiR-277-3p was high expressed miRNA gene in Gravid adults (> 9.0e+03 reads). Nevertheless, *miR-iab-8-5p* had the lower expression profile in the four libraries (<19 reads). The expression profiles of miRNAs varied from highly specific to ubiquitous across all the five developmental stages. Seven miRNAs were detected ubiquitously in all libraries with comparable expression (e.g. *miR-10-5p*). In addition, we were able to characterize some of the novel miRNAs such as *miR-2a-3p, miR-14-3p,* miR-307-3p, *miR-375-3p,* which had the most expressed miRNAs respectively. This similarities was observed in insect species such as *Anopheles gambiae* and *Grapholita molesta* (Biryukova *et al.,* 2014; Wang *et al.,* 2017). In vast majority of cases, mature miRNAs were more abundant than the loop and star sequences.

Additionally, miR-219-5p and miR-309-5p were not found in larvae and pupae libraries. Hence, it can be speculated that miRNA may be involved in regulation of function and

dysfunction of growth and development of the insect species (Sun *et al.*, 2016). Arm selection preference refers to selection of the 5p or 3p arms of miRNA genes to improve the modulation of miRNA biological functions and its regulatory networks (Medley *et al.*, 2021). The analysis of the read numbers revealed that arm selection preference in *G. pallidipes* had biasness towards 3p arm compared to the 5p arm. The arm selection preference has been shown to vary across different dipterans such as *Drosophila*, having a slight bias toward 5′ arm usage (Medley *et al.*, 2021) , *An. Gambia* having a slight biasness toward 3' arm usage (Biryukova *et al.*, 2014) and other types of insects such as *Tribalism castaneum* having an equal number of mature miRNAs preferring either the 5′ and 3′ arm, suggesting that miRNAs may be expressed under a condition-specific manner and arm selection can significantly enrich the regulatory capacity of miRNA.

The characterisation of novel miRNA is a challenging problem for the understanding of post-transcriptional gene regulation. However, different studies have shown that miRNA precursors (hairpin structure) can be used to predict novel miRNAs (Allam *et al.*, 2016; Bisgin *et al.*, 2018) and using miRDeep2, we predicted novel mRNAs by exploring the secondary structure, the Dicer cleavage site, and the minimum free energy of the unannotated small RNA tags that could be mapped to the genome.

According to our evaluation, 58 sequences were retrieved and were regarded as novel MicrRNA candidates. These novel candidates displayed a concentrated length distribution between 18 nucleotides and 25 nucleotides, with a peak at 22 nucleotides. We identified the precursor sequences for some of the novel miRNAs. The detection of miRNA star, loop and mature miRNA is a strong clue to predict the formation of the precursor hairpin. In addition, this increases the accuracy of the predicted candidates (Douglass *et al.*, 2016). Two studies proposed that these miRNAs might differ from their sense partners by acting on different mRNA targets (Douglass *et al.*, 2016). We searched the miRBase databases for homologs to determine whether these novel miRNAs are conserved among other animal species. This search indicated that most of the candidates

(46 (75%)) are conserved in other insect species but not the *Glossina* genus, suggesting that these are insect-specific miRNAs.

Among these new *Glossina* miRNAs, we detected the three *Drosophilid* specific miRNAs (miR-284, miR-274, miR-958 and miR-987) (Shengjie Li *et al.,* 2017). This is the first description of these three miRNAs in the *Glossinidae* species. Secondary structure prediction analysis of some reads resulted in the identification of new stem-loop precursors. The remaining new miRNAs lacks a seed homology to any known miRNAs in insect or other animal species. Counting redundant miRNA reads revealed that expression varies significantly among different stages and sexes. We identified sex-specific miRNA in adult stages (teneral and non-teneral) only. A female-biased (highly expressed in female compared to male) miR-989 as reported in other insects such as *D. melanogaster* and *A. Gambiae* (Feng, Zhou, et al., 2018; Jain et al., 2015; Kugler et al., 2013) had similar expression pattern in the teneral and non-teneral stage. The function of this miRNA has already been characterised in *drosophila* as one of the regulators in the migration of cell borders in the ovary (Kugler *et al.,* 2013). Moreover, more differentially expressed genes were observed in the reproductive adult stage (non-teneral stage) with 18 female-specific miRNA and 7 male-specific miRNA genes. Among these miRNAs miR-994-5p, miR-318-3p, miR-92b-3p, miR-92a-3p, miR-9c-5p also were expressed in a gender-biased pattern in *D. melanogaster* (Marco, 2014).

miRNA clusters are evolutionarily conserved in metazoan species such as nematodes, flies and mammalian species and tend to be localised in tandem clusters of less than 10kb in the genome (Axtell *et al.,* 2011). Recently, differences between male and female miRNA clusters expression profiles have been observed in various insect species (Marco, 2014; Peng *et al.,* 2016), suggesting their role in sex development. In our study, we identified female-biased clusters (miR-309-3p/miR-3-3p/miR-286-5p/miR-5-5p/miR-5-3p/miR-6-3p, miR-318-5p/miR-994-5p, miR-92a-3p/miR-92b-3p/miR-3479-3p, and miR-375-3p/miR-375-5p) and male-biased clusters (miR-125-5p/miR-125-3p, miR-34-5p/miR-34-3p), in the genome assembly of *G. pallidipes.* The pan-drosophilid

miR 309~6 cluster of members (miR-309, miR-3, miR-286, miR-5, miR-6) were identified to align at an interval of 4.31 kb in scaffold 26 of the *G. pallidipes* genome assembly and is associated with insect development through promoting mRNA turnover during the maternal to zygotic transition (Peng *et al.,* 2016).

This study is the first report on how miRNA regulates complete metamorphosis in both sexes in tsetse fly species. miRNA cluster let-7-complex locus (let-7-5p, miR-100-5p and miR-125-5p) is highly conserved among animal species especially insect species. In *D. melanogaster,* the let-7 complex is a spatial-temporal gene that is highly expressed in the pupal stage and adult stages (Chawla & Sokol, 2012; Oliveira *et al.,* 2017) and regulate the temporal identity of *Drosophila* mushroom body neurons via regulating the chinmo genes (Riffo-Campos *et al.,* 2016). In *G. pallidipes,* we identified these orthologue genes clustered in interval of 1.284 kb in scaffold 60 with significant expression during the transition of larvae to pupae, indicating its role in metamorphosis of the insect.

In *G. pallidipes*, sexual dimorphism cannot be easily observed in the early developmental stages (larvae to pupae) as demonstrated in other holometabolous insects such as *D. melanogaster* (Perry *et al.,* 2014), *Musca domestica* (de Jonge *et al.,* 2020) and *Stomoxys calcitrans* (Skovgård & Nachman, 2017). Hence, the identification of stage-specific miRNA expression profile was between unsexed whole pupae and teneral adult stages (female and male). The metamorphosis process involves organ and tissue re-organisation, which requires a lot of molecular interaction as demonstrated with the significant number of miRNAs being expressed. Among these miRNA, miR-iab-4-5p, miR-193-3p/miR-193-5p, miR-210-3p and miR-281-3p/miR-281-1-5p/miR-281-2-5p showed significant expression in adult stages (teneral) compared to the pupae stage. MiR-193 and miR-210 were also similarly expressed in *Anopheles sinensis* (Jain *et al.,* 2015). Previous studies have demonstrated that miR-iab-4-5p and miR-193 has a crucial role in post pupation stage development in other Diptera species such as *Heliconius melpomene* (Surridge *et al.,* 2011) and *Drosophila* (Ninova *et al.,* 2016). The ortholog

miR-210 appears in *D. melanogaster* to affects behavioural circadian rhythmicity and the morphology of the pigment dispersing factor (PDF) positive ventral lateral neuron (Lnv) in the adult stage (Cusumano *et al.,* 2018).Interestingly, miR-282-5p, which was highly expressed in the pupal stage compared to the adult teneral stage, had similar expression patterns in both *D. melanogaster* and *Bactrocera dorsalis* (Calla & Geib, 2015). This miRNA has been shown to regulate viability and production of eggs through the targeting of the nervous-specific adenylate cyclase in pupae during metamorphosis (Vilmos *et al.,* 2013).

When teneral flies take up a blood meal and use the nutrients for body maturation several changes occur especially in terms of gene expression (Lee Rafuse Haines, 2013) , which influences its morphology by transitioning into a mature adult (non-teneral adults). The morphological differences play a crucial role in vector competence and reproduction (Lee Rafuse Haines, 2013), transformation of teneral adult to the non-teneral adult is tightly regulated by a multitude of miRNA expression profiles. In our study, miR-317-5p, miR-34-5p, miR-956-3p and miR-958-3p were upregulated in non-teneral stages (male and female) compared to teneral stage (male and female). In *Drosophila*, miR-956 regulates viral replication by suppression modulates Ectoderm-expressed 4 (Monsanto-Hearne *et al.,* 2017).

Interestingly, miR-34 and other miRNAs were significantly expressed during *Plasmodium* parasite invasion in *A. gambiae* (Biryukova *et al.,* 2014). MiR-958 inhibits Toll signalling and Drosomycin expression by direct targeting of *Toll* and *Dif* proteins in *D. Melanogaster* (Shengjie Li *et al.,* 2017). The miR-275 has been characterised in *G. morsitans* to regulate the synthesis of the proventricular membrane (PM) by modulating the Wnt signalling pathway and the Iroquois/IRX-family of transcription factors, affecting the mechanism by which trypanosomes escape the tsetse PM (E. Aksoy *et al.,* 2016). However, an ortholog gene in *Aedes aegypti* females has been shown to regulate the blood meal-activated physiological events that is involved in the complete maturation of the egg (Bryant *et al.,* 2010). In our study, miR-275 was upregulated in the

non-teneral female stage compared to teneral female stage indicating that the gene might not only be involved in immune response but also in the development of the female reproduction.

The adult female undergoes viviparous reproduction, generating one larva each gonotrophic cycle throughout its lifetime to ensure the survival of the species (Geoffrey M. Attardo *et al.,* 2012). During, which the utero of the female has to expand 100 times to fit the growing larvae (Benoit *et al.,* 2015) and lipid reservoirs from their fat bodies are metabolised for the larval nourishment. In our study, we identified miR-277, which was abundantly expressed in gravid female compared to non-gravid (teneral) female. In addition, miR-277 was among the highly expressed miRNA genes in the gravid female stage. In *Ae. aegypti* mosquitoes, miR-277 targets insulin-like peptides 7 and 8 to control reproduction and lipid metabolism (Ling *et al.,* 2017). The predicted novel miRNAs exhibited much lower expression levels, consistent with the evidence that non-conserved miRNAs are often expressed at a lower level than conserved miRNAs (Allam *et al.,* 2016; Inukai *et al.,* 2012). The low abundance of novel miRNAs might suggest a specific role for these miRNAs under various growth conditions or during developmental stages.

In our study, we mainly focused on identifying the target genes sites using two algorithms (miRanda and RNAhybrid) with the reference genome of *G. morsitans morsitans* as it is well-annotated compared to *G. Pallidipes* genome. Other studies have exploited similar avenues to reduce limitations of genome annotation in identifying putative target genes (Feng, Zhou, et al., 2018; Rebijith et al., 2016; Wang et al., 2017). Also, Algorithm combination is the most plausible way to achieve the best accuracy for conserved miRNA target sites identification for species specific miRNA (Fridrich et al., 2019; Kumar, 2020). Hence this approach was used in the study to reduced the false positive results (Fridrich et al., 2019; Kumar, 2020). The differentially expressed miRNAs were used to identify 5550 mRNA transcripts that have potential target site for our miRNA genes, with numerous biological functions and pathways. During the

metamorphosis of larvae to non-teneral adults, we found that biological processes such as cellular process, metabolic process, organelle, membrane, cellular component organization or biogenesis and enriched pathways such as purine metabolism, thiamine metabolism and aminoacyl-tRNA biosynthesis were among the predicted functional roles of the miRNA target gene list. These biological processes and key pathways might be involved in the formation and regeneration of tissues in the various life stages of *G. pallidipes*.

Networks of the miRNA and mRNA were analysed and revealed that most of miRNAs targeted more than one mRNA transcript and vice versa. The networks demonstrated that miRNA-target study can guide the inference and discovery of the regulatory roles of clustered or co-localized miRNA genes. It is possible to use these maps to improve experimental designs that seek to assess and infer relationship and possible roles of clustered miRNA genes (Figure 4.6). For example, GMOY009801,an orthologue gene of GPAI002137, which is targeted by more than one miRNA (i.e Gpa-miR-190, Gpa-miR-9c, Gpa-miR-219) , is an annotated gene (TSC2 complex) in *G. pallidipes* and many insects plays a vital role in negative regulation of mammalian target of rapamycin (mTOR) signaling pathway (Cui *et al.,* 2019; Dubinsky *et al.,* 2014; James *et al.,* 2018; LoRusso, 2016). The correlation between miRNA genes and its target genes may reflect a regulatory mechanism used by the insect to alter different biological process for its development (Lai, 2015).

To conclude, miRNAs represent a very promising tool in the fight against trypanosomiasis and have the potential not only to allow the detection of the parasites in tsetse (E. Aksoy et al., 2016), but also to determine controlling key genes required for tsetse life cycle, which will significantly boost the success of developing tsetse control strategies.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

In this study, miRNAs in *G. pallidipes* were identified and their expression described in the four developing stages (larva, pupa, adults, gravid adults). Furthermore, this study offers a substantially expansive list of miRNAs in *Glossina* species and provides evidence of an additional 58 new miRNA in this species. These results significantly expand the set of miRNAs in the *G. pallidipes* species to 157 miRNAs. Among theses miRNAs there are 46 miRNAs that match previously to known miRNAs in *D. Melanogaster, bombxy mori, An. gambiae and An. stephensi*. Twelve new miRNAs in *Glossina,* did not match to any known miRNAs in any organism. The expression profile analysis of *G. pallidipes* miRNAs, including the new miRNAs, revealed distinct patterns of expression from larvae to adult stages. Many miRNAs were identied as stage-specific or sex-specific in their expression profiles. In this manner, the study contributed to improving miRNA targets predication in *G. pallidipes.* Combining two miRNA target prediction programs resulted in identifying 5550 target genes. Moreover, functions of the predicted target genes differed but were mostly involved in metabolic process and biological regulation process. Silencing such molecules in a specific developmental stage could decrease the vector population and therefore interrupt trypanosomiasis transmission.

6.2 Recommendations

Based on the findings from this study, the following recommendations are put forward:

1. Further validation should be exploited on the miRNAs characterised using methods such as miRNA luciferase knock-down.

2. Further studies should be done to investigate the target genes and biological functions of the identified miRNA repertoire

REFERENCES

Achkar, N. P., Cambiagno, D. A., & Manavella, P. A. (2016). miRNA Biogenesis: A Dynamic Pathway. *Trends in Plant Science*, *21*(12), 1034–1044. https://doi.org/10.1016/J.TPLANTS.2016.09.003

Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F., & Procopio, A. D. (2016). Bioinformatic tools for microRNA dissection. *Nucleic Acids Research*, *44*(1), 24–44. https://doi.org/10.1093/NAR/GKV1221

Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F., & Procopio, A. D. (2019). A Practical Guide to miRNA Target Prediction. *Methods in Molecular Biology*, *1970*, 1–13. https://doi.org/10.1007/978-1-4939-9207-2_1/COVER

Akkaya, M., Bansal, A., Sheehan, P. W., Pena, M., Cimperman, C. K., Qi, C. F., Yazew, T., Otto, T. D., Billker, O., Miller, L. H., & Pierce, S. K. (2020). Testing the impact of a single nucleotide polymorphism in a Plasmodium berghei ApiAP2 transcription factor on experimental cerebral malaria in mice. *Scientific Reports 2020 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-70617-7

Aksoy, E., Vigneron, A., Bing, X., Zhao, X., O'Neill, M., Wu, Y., Bangs, J. D., Weiss, B. L., & Aksoy, S. (2016). Mammalian African trypanosome VSG coat enhances tsetse's vector competence. *Proceedings of the National Academy of Sciences*, *113*(25), 6961–6966. https://doi.org/10.1073/pnas.1600304113

Aksoy, S. (2019). Tsetse peritrophic matrix influences for trypanosome transmission. *Journal of Insect Physiology*, *118*, 103919. https://doi.org/10.1016/J.JINSPHYS.2019.103919

Aksoy, S., Caccone, A., Galvani, A. P., & Okedi, L. M. (2013). Glossina fuscipes populations provide insights for human African trypanosomiasis transmission

in Uganda. *Trends in Parasitology*, *29*(8). https://doi.org/10.1016/j.pt.2013.06.005

Allam, M., Spillings, B. L., Abdalla, H., Mapiye, D., Koekemoer, L. L., & Christoffels, A. (2016). Identification and characterization of microRNAs expressed in the African malaria vector Anopheles funestus life stages using high throughput sequencing. *Malaria Journal*, *15*(1), 542. https://doi.org/10.1186/s12936-016-1591-0

Alptekin, B., Akpinar, B. A., & Budak, H. (2017). A comprehensive prescription for plant miRNA identification. *Frontiers in Plant Science*, *7*, 2058. https://doi.org/10.3389/FPLS.2016.02058/BIBTEX

Are, E. B., & Hargrove, J. W. (2020). Extinction probabilities as a function of temperature for populations of tsetse (Glossina spp.). *PLOS Neglected Tropical Diseases*, *14*(5), e0007769. https://doi.org/10.1371/JOURNAL.PNTD.0007769

Asgari, S. (2013). MicroRNA functions in insects. *Insect Biochemistry and Molecular Biology*, *43*(4), 388–397. https://doi.org/10.1016/j.ibmb.2012.10.005

Attardo, G. M., Abila, P. P., Auma, J. E., Baumann, A. A., Benoit, J. B., Brelsfoard, C. L., Ribeiro, J. M. C., … & Zhou, J.-J. (2014). Genome Sequence of the Tsetse Fly (Glossina morsitans): Vector of African Trypanosomiasis. *Science*, *344*(6182), 380–386. https://doi.org/10.1126/science.1249656

Attardo, Geoffrey M., Abd-Alla, A. M. M., Acosta-Serrano, A., Allen, J. E., Bateta, R., Benoit, J. B., Bourtzis, K., … & Aksoy, S. (2019). Comparative genomic analysis of six Glossina genomes, vectors of African trypanosomes. *Genome Biology*, *20*(1), 1–31. https://doi.org/10.1186/s13059-019-1768-2

Attardo, Geoffrey M., Benoit, J. B., Michalkova, V., Yang, G., Roller, L., Bohova, J., Takáč, P., & Aksoy, S. (2012). Analysis of lipolysis underlying lactation in the tsetse fly, Glossina morsitans. *Insect Biochemistry and Molecular Biology*, *42*(5), 360–370. https://doi.org/10.1016/j.ibmb.2012.01.007

Avesson, L., Reimegård, J., Wagner, E. G. H., & Söderbom, F. (2012). MicroRNAs in Amoebozoa: Deep sequencing of the small RNA population in the social amoeba Dictyostelium discoideum reveals developmentally regulated microRNAs. *RNA*, *18*(10), 1771–1782. https://doi.org/10.1261/rna.033175.112

Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, *12*(4), 221. https://doi.org/10.1186/gb-2011-12-4-221

Bargul, J. L., Jung, J., McOdimba, F. A., Omogo, C. O., Adung'a, V. O., Krüger, T., Masiga, D. K., & Engstler, M. (2016). Species-Specific Adaptations of Trypanosome Morphology and Motility to the Mammalian Host. *PLoS Pathogens*, *12*(2), e1005448. https://doi.org/10.1371/journal.ppat.1005448

Bateta, R., Saarman, N. P., Okeyo, W. A., Dion, K., Johnson, T., Mireji, P. O., Okoth, S., Malele, I., Murilla, G., Aksoy, S., & Caccone, A. (2020). Phylogeography and population structure of the tsetse fly Glossina pallidipes in Kenya and the Serengeti ecosystem. *PLOS Neglected Tropical Diseases*, *14*(2), e0007855. https://doi.org/10.1371/JOURNAL.PNTD.0007855

Bateta, R., Wang, J., Wu, Y., Weiss, B. L., Warren, W. C., Murilla, G. A., Aksoy, S., & Mireji, P. O. (2017). Tsetse fly (Glossina pallidipes) midgut responses to Trypanosoma brucei challenge. *Parasites & Vectors*, *10*(1), 614. https://doi.org/10.1186/s13071-017-2569-7

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. https://doi.org/10.1146/annurev-ento-010814-020834

Birhanu, H., Rogé, S., Simon, T., Baelmans, R., Gebrehiwot, T., Goddeeris, B. M., & Büscher, P. (2015). Surra Sero K-SeT, a new immunochromatographic test for serodiagnosis of Trypanosoma evansi infection in domestic animals. *Veterinary Parasitology*, *211*(3–4), 153–157. https://doi.org/10.1016/J.VETPAR. 2015.05.008

Biryukova, I., Ye, T., & Levashina, E. (2014). Transcriptome-wide analysis of microRNA expression in the malaria mosquito Anopheles. In *BMC Genomics*. BioMed Central. https://doi.org/10.1186/1471-2164-15-557

Bisgin, H., Gong, B., Wang, Y., & Tong, W. (2018). Evaluation of Bioinformatics Approaches for Next-Generation Sequencing Analysis of microRNAs with a Toxicogenomics Study Design. *Frontiers in Genetics*, *9*, 22. https://doi.org/10.3389/fgene.2018.00022

Boelaert, M., Mukendi, D., Bottieau, E., Kalo Lilo, J. R., Verdonck, K., Minikulu, L., Barbé, B., Gillet, P., Yansouni, C. P., Chappuis, F., & Lutumba, P. (2018). A Phase III Diagnostic Accuracy Study of a Rapid Diagnostic Test for Diagnosis of Second-Stage Human African Trypanosomiasis in the Democratic Republic of the Congo. *EBioMedicine*, *27*, 11. https://doi.org/10.1016/J.EBIOM .2017.10.032

Bonnet, J., Boudot, C., & Courtioux, B. (2015). Overview of the Diagnostic Methods Used in the Field for Human African Trypanosomiasis: What Could Change in the Next Years? *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/583262

Bottieau, E., & Clerinx, J. (2019). Human African Trypanosomiasis: Progress and Stagnation. *Infectious Disease Clinics of North America*, *33*(1), 61–77. https://doi.org/10.1016/j.idc.2018.10.003

Bryant, B., Macdonald, W., & Raikhel, A. S. (2010). microRNA miR-275 is indispensable for blood digestion and egg development in the mosquito Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(52), 22391–22398. https://doi.org/10.1073/pnas.1016230107

Büscher, P., Cecchi, G., Jamonneau, V., & Priotto, G. (2017). Human African trypanosomiasis. *The Lancet*, *390*(10110), 2397–2409. https://doi.org/10.1016/ S0140-6736(17)31510-6

Bustamante, J. M., Padilla, A. M., White, B., Auckland, L. D., Busselman, R. E., Collins, S., Malcolm, E. L., Wilson, B. F., Saunders, A. B., Hamer, S. A., & Tarleton, R. L. (2022). Prophylactic low-dose, bi-weekly benznidazole treatment fails to prevent Trypanosoma cruzi infection in dogs under intense transmission pressure. *PLOS Neglected Tropical Diseases*, *16*(10), e0010688. https://doi.org/10.1371/JOURNAL.PNTD.0010688

Caljon, G., Stijlemans, B., De Trez, C., & Van Den Abbeele, J. (2017). Early Immunological Responses Upon Tsetse Fly–Mediated Trypanosome Inoculation. In *Arthropod Vector: Controller of Disease Transmission, Volume 2* (pp. 115–132). Elsevier. https://doi.org/10.1016/B978-0-12-805360-7.00007-1

Calla, B., & Geib, S. M. (2015). MicroRNAs in the oriental fruit fly, Bactrocera dorsalis: extending Drosophilid miRNA conservation to the Tephritidae. *BMC Genomics*, *16*(1), 740. https://doi.org/10.1186/s12864-015-1835-3

Carmel, I., Shomron, N., & Heifetz, Y. (2012). Does base-pairing strength play a role in

microRNA repression? *RNA*, *18*(11), 1947–1956. https://doi.org/10.1261/rna.032185.111

Carthew, R. W., Agbu, P., & Giri, R. (2017). MicroRNA function in Drosophila melanogaster. *Seminars in Cell & Developmental Biology*, *65*, 29–37. https://doi.org/10.1016/J.SEMCDB.2016.03.015

Cayla, M., Rojas, F., Silvester, E., Venter, F., & Matthews, K. R. (2019). African trypanosomes. *Parasites & Vectors 2019 12:1*, *12*(1), 1–8. https://doi.org/10.1186/S13071-019-3355-5

Cecchi, G., Paone, M., Argilés Herrero, R., Vreysen, M., & Mattioli, R. (2015). Developing a continental atlas of the distribution and trypanosomal infection of tsetse flies (Glossina species). *Parasites & Vectors*, *8*(1), 284. https://doi.org/10.1186/s13071-015-0898-y

Chawla, G., & Sokol, N. S. (2012). Hormonal activation of let-7-C microRNAs via EcR is required for adult Drosophila melanogaster morphology and function. *Development (Cambridge, England)*, *139*(10), 1788–1797. https://doi.org/10.1242/dev.077743

Chilongo, K., Manyangadze, T., & Mukaratirwa, S. (2021). Effects of Human Settlements and Spatial Distribution of Wing Vein Length, Wing Fray Categories and Hunger Stages in Glossina morsitans morsitans (Diptera: Glossinidae) and Glossina pallidipes (Diptera: Glossinidae) in Areas Devoid of Cattle in North-Eastern Zambia. *Journal of Medical Entomology*, *58*(2), 891–899. https://doi.org/10.1093/JME/TJAA228

Choudhuri, S. (2010). Small noncoding RNAs: Biogenesis, function, and emerging significance in toxicology. In *Journal of Biochemical and Molecular Toxicology* (Vol. 24, Issue 3, pp. 195–216). John Wiley and Sons Inc.

https://doi.org/10.1002/jbt.20325

Conca James. (2015). *Irradiating And Eradicating The Tsetse Fly Scourge*. Forbes. https://www.forbes.com/sites/jamesconca/2015/04/02/irradiating-and-eradicating-the-tsetse-fly-scourge/?sh=36e065cf5c04

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Cordero, F., Beccuti, M., Arigoni, M., Donatelli, S., & Calogero, R. A. (2012). Optimizing a Massive Parallel Sequencing Workflow for Quantitative miRNA Expression Analysis. *PLoS ONE*, *7*(2), e31630. https://doi.org/10.1371/journal.pone.0031630

Cowled, C., Stewart, C. R., Likic, V. A., Friedländer, M. R., Tachedjian, M., Jenkins, K. A., Tizard, M. L., … & Wang, L. (2014). Characterisation of novel microRNAs in the Black flying fox (Pteropus alecto) by deep sequencing. *BMC Genomics*, *15*(1), 682. https://doi.org/10.1186/1471-2164-15-682

Cui, G., Shu, B., Veeran, S., Yuan, H., Yi, X., & Zhong, G. (2019). Natural β-carboline alkaloids regulate the PI3K/Akt/mTOR pathway and induce autophagy in insect Sf9 cells. *Pesticide Biochemistry and Physiology*, *154*, 67–77. https://doi.org/10.1016/J.PESTBP.2018.12.005

Cusumano, P., Biscontin, A., Sandrelli, F., Mazzotta, G. M., Tregnago, C., De Pittà, C., & Costa, R. (2018). Modulation of miR-210 alters phasing of circadian locomotor activity and impairs projections of PDF clock neurons in Drosophila melanogaster. *PLOS Genetics*, *14*(7), e1007500. https://doi.org/10.1371/journal.pgen.1007500

de Beer, C. J., Moyaba, P., Boikanyo, S. N. B., Majatladi, D., Yamada, H., Venter, G. J., & Vreysen, M. J. B. (2017). Evaluation of radiation sensitivity and mating performance of Glossina brevipalpis males. *PLOS Neglected Tropical Diseases*, *11*(3), e0005473. https://doi.org/10.1371/journal.pntd.0005473

Desquesnes, M. & Dia, M. L. Mechanical transmission of *Trypanosoma congolense* in cattle by the African tabanid *Atylotus agrestis*. Exp. Parasitol. 105, 226–231 (2003).

.De Deken, R. (n.d.). *Tsetse flies*. Retrieved July 26, 2017, from http://www.afrivip.org/sites/default/files/10_tsetse_references.pdf

de Jonge, N., Michaelsen, T. Y., Ejbye-Ernst, R., Jensen, A., Nielsen, M. E., Bahrndorff, S., & Nielsen, J. L. (2020). Housefly (Musca domestica L.) associated microbiota across different life stages. *Scientific Reports 2020 10:1*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-64704-y

de Koning, H. P. (2020). The Drugs of Sleeping Sickness: Their Mechanisms of Action and Resistance, and a Brief History. *Tropical Medicine and Infectious Disease 2020, Vol. 5, Page 14*, *5*(1), 14. https://doi.org/10.3390/TROPICALMED5010014

Demirbas-Uzel, G., De Vooght, L., Parker, A. G., Vreysen, M. J. B., MacH, R. L., Van Den Abbeele, J., & Abd-Alla, A. M. M. (2018). Combining paratransgenesis with SIT: Impact of ionizing radiation on the DNA copy number of Sodalis glossinidius in tsetse flies. *BMC Microbiology*, *18*(1), 39–53. https://doi.org/10.1186/S12866-018-1283-8/FIGURES/8

Desquesnes, M., Sazmand, A., Gonzatti, M., Boulangé, A., Bossard, G., Thévenon, S., Gimonneau, G., … & Hébert, L. (2022). Diagnosis of animal trypanosomoses: proper use of current tools and future prospects. *Parasites & Vectors 2022*

15:1, *15*(1), 1–22. https://doi.org/10.1186/S13071-022-05352-1

Doudoumis, V., Blow, F., Saridaki, A., Augustinos, A., Dyer, N. A., Goodhead, I., Solano, P., … & Tsiamis, G. (2017). Challenging the Wigglesworthia, Sodalis, Wolbachia symbiosis dogma in tsetse flies: Spiroplasma is present in both laboratory and natural populations. *Scientific Reports 2017 7:1*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-04740-3

Douglass, S., Hsu, S. W., Cokus, S., Goldberg, R. B., Harada, J. J., & Pellegrini, M. (2016). A naïve Bayesian classifier for identifying plant microRNAs. *The Plant Journal*, *86*(6), 481–492. https://doi.org/10.1111/TPJ.13180

Dubinsky, A. N., Dastidar, S. G., Hsu, C. L., Zahra, R., Djakovic, S. N., Duarte, S., Esau, C. C., … & La Spada, A. R. (2014). Let-7 coordinately suppresses components of the amino acid sensing pathway to repress mTORC1 and induce autophagy. *Cell Metabolism*, *20*(4), 626–638. https://doi.org/10.1016/j.cmet.2014.09.001

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, *37*(5), 699–735.

English, S., Cowen, H., Garnett, E., & Hargrove, J. W. (2016). Maternal effects on offspring size in a natural population of the viviparous tsetse fly. *Ecological Entomology*, *41*(5), 618–626. https://doi.org/10.1111/EEN.12333

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., & Marks, D. S. (2003). MicroRNA targets in Drosophila. *Genome Biology*, *5*(1), R1. https://doi.org/10.1186/gb-2003-5-1-r1

Feldmann, U., Dyck, V. A., Mattioli, R. C., Jannin, J., & Vreysen, M. J. B. (2021). Impact of Tsetse Fly Eradication Programmes Using the Sterile Insect

Technique. *Sterile Insect Technique*, 1051–1080. https://doi.org/10.1201/9781003035572-32

Feng, X., Wu, J., Zhou, S., Wang, J., & Hu, W. (2018). Characterization and potential role of microRNA in the Chinese dominant malaria mosquito Anopheles sinensis (Diptera: Culicidae) throughout four different life stages. *Cell and Bioscience*, *8*(1), 29. https://doi.org/10.1186/s13578-018-0227-1

Feng, X., Zhou, X., Zhou, S., Wang, J., & Hu, W. (2018). Analysis of microRNA profile of Anopheles sinensis by deep sequencing and bioinformatic approaches. *Parasites & Vectors*, *11*(1), 172. https://doi.org/10.1186/s13071-018-2734-7

Fridrich, A., Hazan, Y., & Moran, Y. (2019). Too Many False Targets for MicroRNAs: Challenges and Pitfalls in Prediction of miRNA Targets and Their Gene Ontology in Model and Non-model Organisms. *BioEssays*, *41*(4), 1800169. https://doi.org/10.1002/BIES.201800169

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Research, 40(1), 37–52. https://doi.org/10.1093/ nar/gkr688.

Geiger, A., Ponton, F., & Simo, G. (2015). Adult blood-feeding tsetse flies, trypanosomes, microbiota and the fluctuating environment in sub-Saharan Africa. *The ISME Journal*, *9*(10), 1496–1507. https://doi.org/10.1038/ismej.2014.236

GIORDANI, F., MORRISON, L. J., ROWAN, T. G., DE KONING, H. P., & BARRETT, M. P. (2016). The animal trypanosomiases and their chemotherapy: a review. *Parasitology*, *143*(14), 1862–1889. https://doi.org/10. 1017/S0031182016001268

Gomes, C. P. C., Cho, J. H., Hood, L., Franco, O. L., Pereira, R. W., & Wang, K. (2013). A review of computational tools in microRNA discovery. In *Frontiers in Genetics* (Vol. 4, Issue MAY). https://doi.org/10.3389/fgene.2013.00081

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. https://doi.org/10.1146/annurev.ento.50.071803.130443

Gopinath, S. C. B., Lakshmipriya, T., Md Arshad, M. K., Voon, C. H., Adam, T., Hashim, U., Singh, H., & Chinni, S. V. (2017). Shortening full-length aptamer by crawling base deletion – Assisted by Mfold web server application. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, *23*(1), 37–42. https://doi.org/10.1016/J.JAUBAS.2016. 07.001/ SUPPL_FILE/TABS19_A_11985723_SM0001.PPTX

Gu, W., Xu, Y., Xie, X., Wang, T., Ko, J. H., & Zhou, T. (2014). The role of RNA structure at 5′ untranslated region in microRNA-mediated gene regulation. *RNA*, *20*(9), 1369–1375. https://doi.org/10.1261/rna.044792.114

Haines, Lee R, Vale, G. A., Barreaux, A. M. G., Ellstrand, N. C., Hargrove, J. W., English, S., Haines, L. R., … & Ellstrand, N. C. (2020). Big Baby, Little Mother: Tsetse Flies Are Exceptions to the Juvenile Small Size Principle. *BioEssays*, *42*(11), 2000049. https://doi.org/10.1002/BIES.202000049

Haines, Lee Rafuse. (2013). Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection. *Frontiers in Cellular and Infection Microbiology*, *3*, 84. https://doi.org/10.3389/fcimb.2013.00084

Hargrove, J. W., & Vale, G. A. (2020). Models for the rates of pupal development, fat consumption and mortality in tsetse (Glossina spp). *Bulletin of Entomological Research*, *110*(1), 44–56. https://doi.org/10.1017/S0007485319000233

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte Für Chemie Chemical Monthly*, *125*(2), 167–188. https://doi.org/10.1007/BF00818163

Hoffmann, A. A., Montgomery, B. L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P. H., Muzzi, F., Greenfield, M., … & O'Neill, S. L. (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature*, *476*(7361), 454–457. https://doi.org/10.1038/ nature 10356

Horn, D. (2014). Antigenic variation in African trypanosomes. *Molecular and Biochemical Parasitology*, *195*(2), 123–129. https://doi.org/10.1016/j. molbiopara .2014.05.001

Inukai, S., de Lencastre, A., Turner, M., & Slack, F. (2012). Novel MicroRNAs Differentially Expressed during Aging in the Mouse Brain. *PLoS ONE*, *7*(7), e40028. https://doi.org/10.1371/journal.pone.0040028

Jain, S., Rana, V., & Tridibes, A. (2015). Dynamic expression of miRNAs across immature and adult stages of the malaria mosquito Anopheles stephensi. *Parasites &*. https://parasitesandvectors.biomedcentral.com/ articles/ 10.1186/ s13071-015-0772-y

James, E. B., Feng, H., & Wilson, A. C. C. (2018). mTOR Complex 1 Implicated in Aphid/Buchnera Host/Symbiont Integration. *G3 Genes|Genomes|Genetics*, *8*(9), 3083–3091. https://doi.org/10.1534/G3.118.200398

Jones, T. W. & Dávila, A. M. R. (2001).*Trypanosoma vivax* - Out of Africa. *Trends Parasitol. 17*, 99–101

Judice, C. C., Bourgard, C., Kayano, A. C. A. V, Albrecht, L., & Costa, F. T. M. (2016). MicroRNAs in the Host-Apicomplexan Parasites Interactions: A Review of Immunopathological Aspects. *Frontiers in Cellular and Infection Microbiology*, *6*, 5. https://doi.org/10.3389/fcimb.2016.00005

Kabekkodu, S. P., Shukla, V., Varghese, V. K., Adiga, D., Vethil Jishnu, P., Chakrabarty, S., & Satyamoorthy, K. (2020). Cluster miRNAs and cancer: Diagnostic, prognostic and therapeutic opportunities. *Wiley Interdisciplinary Reviews: RNA*, *11*(2), e1563. https://doi.org/10.1002/WRNA.1563

Kennell, J. A., Cadigan, K. M., Shakhmantsir, I., & Waldron, E. J. (2012). The microRNA miR-8 is a positive regulator of pigmentation and eclosion in Drosophila. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *241*(1), 161–168. https://doi.org/10.1002/dvdy.23705

Kim, D., Chang, H. R., & Baek, D. (2017). Rules for functional microRNA targeting. In *BMB Reports* (Vol. 50, Issue 11, pp. 554–559). The Biochemical Society of the Republic of Korea. https://doi.org/10.5483/BMBRep.2017.50.11.179

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141

Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(Database issue), D68-73. https://doi.org/10.1093/nar/gkt1181

Krafsur, E. S., Marquez, J. G., & Ouma, J. O. (2016). Phylogeography and genealogy of the tsetse fly Glossina pallidipes (Diptera: Glossinidae). *International Journal of Tropical Insect Science*, *36*(1), 32–47. https://doi.org/10.1017/ S174275841

5000223

Krafsur, E. S., & Maudlin, I. (2018). Tsetse fly evolution, genetics and the trypanosomiases - A review. *Infection, Genetics and Evolution*, *64*, 185–206. https://doi.org/10.1016/J.MEEGID.2018.05.033

Krinsky, W. L. (2019). Tsetse Flies (Glossinidae). *Medical and Veterinary Entomology*, 369–382. https://doi.org/10.1016/B978-0-12-814043-7.00018-2

Krüger, J., & Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research*, *34*(Web Server issue), W451-4. https://doi.org/10.1093/nar/gkl243

Kugler, J.-M., Verma, P., Chen, Y.-W., Weng, R., & Cohen, S. M. (2013). miR-989 is required for border cell migration in the Drosophila ovary. *PloS One*, *8*(7), e67075. https://doi.org/10.1371/journal.pone.0067075

Kumar, V. (2020). Mini Review Journal of Translational Science Reducing false positive prediction of miRNA target genes with RNA sequencing: Hype or hope. *J Transl Sci*, 2020. https://doi.org/10.15761/JTS.1000409

Lai, E. C. (2015). Two decades of miRNA biology: lessons and challenges. *RNA (New York, N.Y.)*, *21*(4), 675–677. https://doi.org/10.1261/rna.051193.115

Langley, S. A., Messina, J. P., & Moore, N. (2017). Using meta-quality to assess the utility of volunteered geographic information for science. *International Journal of Health Geographics*, *16*(1). https://doi.org/10.1186/S12942-017-0113-9

Li, Sheng, Yu, X., & Feng, Q. (2019). Fat Body Biology in the Last Decade. *Annual Review of Entomology*, *64*, 315–333. https://doi.org/10.1146/ANNUREV-ENTO-011118-112007

Li, Shengjie, Li, Y., Shen, L., Jin, P., Chen, L., & Ma, F. (2017). miR-958 inhibits Toll signaling and Drosomycin expression via direct targeting of *Toll* and *Dif* in *Drosophila melanogaster*. *American Journal of Physiology-Cell Physiology*, *312*(2), C103–C110. https://doi.org/10.1152/ajpcell.00251.2016

Li, X. (2017). Next-Generation Sequencing Sheds New Light on Small RNAs in Plant Reproductive Development. *Current Issues in Molecular Biology 2018, Vol. 27, Pages 143-170*, *27*(1), 143–170. https://doi.org/10.21775/CIMB.027.143

Li, Z., Chen, D., Wang, Q., Tian, H., Tan, M., Peng, D., Tan, Y., Zhu, J., Liang, W., & Zhang, L. (2021). mRNA and microRNA stability validation of blood samples under different environmental conditions. *Forensic Science International: Genetics*, *55*, 102567. https://doi.org/10.1016/J.FSIGEN.2021.102567

Ling, L., Kokoza, V. A., Zhang, C., Aksoy, E., & Raikhel, A. S. (2017). MicroRNA-277 targets insulin-like peptides 7 and 8 to control lipid metabolism and reproduction in Aedes aegypti mosquitoes. *Proceedings of the National Academy of Sciences*, *114*(38), E8017–E8024. https://doi.org/10.1073/PNAS.1710970114

Liu, W., Hao, Z., Huang, L., Chen, L., Wei, Q., Cai, L., & Liang, S. (2017). Comparative expression profile of microRNAs in Anopheles anthropophagus midgut after blood-feeding and Plasmodium infection. *Parasites & Vectors*, *10*(1), 86. https://doi.org/10.1186/s13071-017-2027-6

LoRusso, P. M. (2016). Inhibition of the PI3K/AKT/mTOR Pathway in Solid Tumors. *Journal of Clinical Oncology*, *34*(31), 3803. https://doi.org/10.1200/JCO.2014.59.0018

Lucas, K., & Raikhel, A. S. (2013). Insect microRNAs: biogenesis, expression profiling and biological functions. *Insect Biochemistry and Molecular Biology*, *43*(1),

24–38. https://doi.org/10.1016/j.ibmb.2012.10.009

Mani, I. (2021). Role of Bioinformatics in MicroRNA Analysis. *Advances in Bioinformatics*, 365–373. https://doi.org/10.1007/978-981-33-6191-1_19

Marco, A. (2014). Sex-biased expression of microRNAs in Drosophila melanogaster. *Open Biology*, *4*(4), 140024. https://doi.org/10.1098/rsob.140024

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mbahin, N., Affognon, H., Andoke, J., Tiberius, M., Mbuvi, D., Otieno, J., Muasa, P., & Saini, R. K. (2013). Parasitological Prevalence of Bovine Trypanosomosis in Kubo Division of Kwale County of Coastal: Baseline Survey. *American Journal of Animal and Veterinary Sciences*, *8*(1), 28–36. https://doi.org/10.3844/ajavssp.2013.28.36

Mbewe, N. J., Saini, R. K., Torto, B., Irungu, J., Yusuf, A. A., & Pirk, C. W. W. (2018). Sticky small target: an effective sampling tool for tsetse fly Glossina fuscipes fuscipes Newstead, 1910. *Parasites & Vectors*, *11*(1), 268. https://doi.org/10.1186/s13071-018-2840-6

Medley, J. C., Panzade, G., & Zinovyeva, A. Y. (2021). microRNA strand selection: Unwinding the rules. *Wiley Interdisciplinary Reviews: RNA*, *12*(3), e1627. https://doi.org/10.1002/WRNA.1627

Megy, K., Emrich, S. J., Lawson, D., Campbell, D., Dialynas, E., Hughes, D. S. T., Koscielny, G., … & VectorBase Consortium. (2012). VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Research*, *40*(D1), D729–D734. https://doi.org/10.1093

/nar/gkr1089

Meki, I. K., İnce, İ. A., Kariithi, H. M., Boucias, D. G., Ozcan, O., Parker, A. G., Vlak, J. M., van Oers, M. M., & Abd-Alla, A. M. M. (2018). Expression Profile of Glossina pallidipes MicroRNAs During Symptomatic and Asymptomatic Infection With Glossina pallidipes Salivary Gland Hypertrophy Virus (Hytrosavirus). *Frontiers in Microbiology*, *9*, 2037. https://doi.org/10.3389/fmicb.2018.02037

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*(12), e0005247. https://doi.org/10.1371/journal.pntd.0005247

Mihok, S., Maramba, O., Munyoki, E. & Kagoiya, J. (1995), Mechanical transmission of *Trypanosoma* spp. by African Stomoxyinae (Diptera: Muscidae). *Trop. Med. Parasitol. 46*, 103–105

Miskiewicz, J., Tomczyk, K., Mickiewicz, A., Sarzynska, J., & Szachniuk, M. (2017). Bioinformatics Study of Structural Patterns in Plant MicroRNA Precursors. *BioMed Research International*, *2017*. https://doi.org/10.1155/2017/6783010

Monsanto-Hearne, V., Tham, A. L. Y., Wong, Z. S., Asgari, S., & Johnson, K. N. (2017). Drosophila miR-956 suppression modulates Ectoderm-expressed 4 and inhibits viral replication. *Virology*, *502*, 20–27. https://doi.org/10.1016/J.VIROL.2016.12.009

Moran, Z. R., & Parker, A. G. (2016). Near Infrared Imaging As a Method of Studying Tsetse Fly (Diptera: Glossinidae) Pupal Development. *Journal of Insect Science (Online)*, *16*(1). https://doi.org/10.1093/jisesa/iew047

Morrison, L. J., Vezza, L., Rowan, T., & Hope, J. C. (2016). Animal African Trypanosomiasis: Time to Increase Focus on Clinically Relevant Parasite and Host Species. *Trends in Parasitology*, *32*(8), 599–607. https://doi.org/10.1016/J.PT.2016.04.012

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., Floden, E. W., … & Finn, R. D. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, *43*(D1), D130–D137. https://doi.org/10.1093/nar/gku1063

Nawrocki, Eric P, & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics (Oxford, England)*, *29*(22), 2933–2935. https://doi.org /10. 1093/bioinformatics/btt509

Nimpaye, H., Njiokou, F., Njine, T., Njitchouang, G. R., Cuny, G., Herder, S., Asonganyi, T., & Simo, G. (2011). Trypanosoma vivax, T. congolense "forest type" and T. simiae: prevalence in domestic animals of sleeping sickness foci of Cameroon. *Parasite (Paris, France)*, *18*(2), 171–179. https://doi.org/10.1051/parasite/2011182171

Ninova, M., Ronshaugen, M., & Griffiths-Jones, S. (2016). MicroRNA evolution, expression, and function during short germband development in Tribolium castaneum. *Genome Research*, *26*(1), 85–96. https://doi.org/10.1101 /gr.193367.115

Ntantiso, L., De Beer, C., Marcotty, T., & Latif, A. A. (2014). Bovine trypanosomosis prevalence at the edge of Hluhluwe-iMfolozi Park, KwaZulu-Natal, South Africa. *Onderstepoort J Vet Res*, *81*(1), 8 pages. https://doi.org/10.4102 /ojvr.v81i1.762

Odeniran, P. O., & Ademola, I. O. (2018). A meta-analysis of the prevalence of African

animal trypanosomiasis in Nigeria from 1960 to 2017. *Parasites & Vectors*, *11*(1), 280. https://doi.org/10.1186/s13071-018-2801-0

Okeyo, W. A., Saarman, N. P., Mengual, M., Dion, K., Bateta, R., Mireji, P. O., Okoth, S., … & Caccone, A. (2017). Temporal genetic differentiation in Glossina pallidipes tsetse fly populations in Kenya. *Parasites and Vectors*, *10*(1), 1–13. https://doi.org/10.1186/S13071-017-2415-Y/TABLES/6

Oliveira, A. C., Bovolenta, L. A., Nachtigall, P. G., Herkenhoff, M. E., Lemke, N., & Pinhal, D. (2017). Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Frontiers in Genetics*, *8*(MAY), 59. https://doi.org/10.3389/FGENE.2017.00059/BIBTEX

Pagabeleguem, S., Gimonneau, G., Seck, M. T., Vreysen, M. J. B., Sall, B., Rayaissé, J.-B., Sidibé, I., Bouyer, J., & Ravel, S. (2016). A Molecular Method to Discriminate between Mass-Reared Sterile and Wild Tsetse Flies during Eradication Programmes That Have a Sterile Insect Technique Component. *PLOS Neglected Tropical Diseases*, *10*(2), e0004491. https://doi.org/10.1371/ournal.pntd.0004491

Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., Filippidis, … & Hatzigeorgiou, A. G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Research*, *41*(Web Server issue). https://doi.org/10.1093/nar/gkt393

Parveen, A., Mustafa, S. H., Yadav, P., & Kumar, A. (2020). Applications of Machine Learning in miRNA Discovery and Target Prediction. *Current Genomics*, *20*(8), 537–544. https://doi.org/10.2174/1389202921666200106111813

Patel, M., Mangukia, N., Jha, N., Gadhavi, H., Shah, K., Patel, S., Mankad, A., Pandya,

H., & Rawal, R. (2019). Computational identification of miRNA and their cross kingdom targets from expressed sequence tags of Ocimum basilicum. *Molecular Biology Reports 2019 46:3*, *46*(3), 2979–2995. https://doi.org/10.1007/S11033-019-04759-X

Peng, W., Tariq, K., Xie, J., & Zhang, H. (2016). Identification and Characterization of Sex-Biased MicroRNAs in Bactrocera dorsalis (Hendel). *PLOS ONE*, *11*(7), e0159591. https://doi.org/10.1371/journal.pone.0159591

Perry, J. C., Harrison, P. W., & Mank, J. E. (2014). The ontogeny and evolution of sex-biased gene expression in Drosophila melanogaster. *Molecular Biology and Evolution*, *31*(5), 1206–1219. https://doi.org/10.1093/molbev/msu072

Peterson, S. M., Thompson, J. A., Ufkin, M. L., Sathyanarayana, P., Liaw, L., & Congdon, C. B. (2014). Common features of microRNA target prediction tools. In *Frontiers in Genetics* (Vol. 5, Issue FEB). https://doi.org/10.3389/fgene.2014.00023

Pla, A., Zhong, X., & Rayner, S. (2018). miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. *PLOS Computational Biology*, *14*(7), e1006185. https://doi.org/10.1371/ journal. pcbi.1006185

Prabu, G. R., & Mandal, A. K. A. (2010). Computational identification of miRNAs and their target genes from expressed sequence tags of tea (Camellia sinensis). *Genomics, Proteomics and Bioinformatics*, *8*(2), 113–121. https://doi.org/10.1016/ S1672-0229(10)60012-5

Prayag, K., Surve, D. H., Paul, A. T., Kumar, S., & Jindal, A. B. (2020). Nanotechnological interventions for treatment of trypanosomiasis in humans and animals. *Drug Delivery and Translational Research 2020 10:4*, *10*(4),

945–961. https://doi.org/10.1007/S13346-020-00764-X

Quillet, A., Saad, C., Ferry, G., Anouar, Y., Vergne, N., Lecroq, T., & Dubessy, C. (2020). Improving Bioinformatics Prediction of microRNA Targets by Ranks Aggregation. *Frontiers in Genetics*, *10*, 1330. https://doi.org/10.3389/FGENE.2019.01330/BIBTEX

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rajendiran, A., Chatterjee, A., & Pan, A. (2017). Computational Approaches and Related Tools to Identify MicroRNAs in a Species: A Bird's Eye View. *Interdisciplinary Sciences: Computational Life Sciences 2017 10:3*, *10*(3), 616–635. https://doi.org/10.1007/S12539-017-0223-X

Rebijith, K. B., Asokan, R., Ranjitha Hande, H., & Krishna Kumar, N. K. (2016). The first report of miRNAs from a thysanopteran insect, Thrips palmi Karny using high-throughput sequencing. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0163635

Riffo-Campos, Á. L., Riquelme, I., & Brebi-Mieville, P. (2016). Tools for Sequence-Based miRNA Target Prediction: What to Choose? *International Journal of Molecular Sciences*, *17*(12). https://doi.org/10.3390/ijms17121987

Rio, R. V. M., Attardo, G. M., & Weiss, B. L. (2016). Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy. *Trends in Parasitology*, *32*(9), 739–749. https://doi.org/10.1016/J.PT.2016.05.002

Riolo, G., Cantara, S., Marzocchi, C., & Ricci, C. (2020). miRNA Targets: From Prediction Tools to Experimental Validation. *Methods and Protocols 2021,*

Vol. 4, Page 1, *4*(1), 1. https://doi.org/10.3390/MPS4010001

Rissone, P., Bizarro, C. V., & Ritort, F. (2022). Stem-loop formation drives RNA folding in mechanical unzipping experiments. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(3), e2025575119. https://doi.org/10.1073/PNAS.2025575119/SUPPL_FILE/PNAS.2025575119.SAPP.PDF

Roberts, J. T., & Borchert, G. M. (2017). Computational prediction of microRNA target genes, target prediction databases, and web resources. *Methods in Molecular Biology*, *1617*, 109–122. https://doi.org/10.1007/978-1-4939-7046-9_8/COVER

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rodriguez, M. E., Tekiel, V., & Campo, V. A. (2022). In vitro evaluation of Resveratrol as a potential pre-exposure prophylactic drug against Trypanosoma cruzi infection. *International Journal for Parasitology: Drugs and Drug Resistance*, *20*, 54–64. https://doi.org/10.1016/J.IJPDDR.2022.08.003

Saçar Demirci, M. D., Baumbach, J., & Allmer, J. (2017). On the performance of pre-microRNA detection algorithms. *Nature Communications 2017 8:1*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-00403-z

Sadek, A., el-Khabaz, K. A. S., el-Genedy, S. M., & el-Gioushy, M. M. (2021). Comparative Diagnostic Performance of Microscopic Examination, Polyclonal Antigen-ELISA, and Polymerase Chain Reaction for the Detection of Trypanosoma evansi in Camels (Camelus dromedarius). *Advances in*

Animal and Veterinary Sciences, *9*(7), 1004–1011. https://doi.org/10.17582/JOURNAL.AAVS/2021/9.7.1004.1011

Saini, R. K., Orindi, B. O., Mbahin, N., Andoke, J. A., Muasa, P. N., Mbuvi, D. M., Muya, C. M., Pickett, J. A., & Borgemeister, C. W. (2017). Protecting cows in small holder farms in East Africa from tsetse flies by mimicking the odor profile of a non-host bovid. *PLOS Neglected Tropical Diseases*, *11*(10), e0005977. https://doi.org/10.1371/journal.pntd.0005977

Seok, H., Ham, J., Jang, E. S., & Chi, S. W. (2016). MicroRNA Target Recognition: Insights from Transcriptome-Wide Non-Canonical Interactions. *Molecules and Cells*, *39*(5), 375. https://doi.org/10.14348/MOLCELLS.2016.0013

Shaker, F., Nikravesh, A., Arezumand, R., & Aghaee-Bakhtiari, S. H. (2020). Web-based tools for miRNA studies analysis. *Computers in Biology and Medicine*, *127*, 104060. https://doi.org/10.1016/J.COMPBIOMED.2020.104060

Shirjang, S., Mansoori, B., Asghari, S., Duijf, P. H. G., Mohammadi, A., Gjerstorff, M., & Baradaran, B. (2019). MicroRNAs in cancer cell death pathways: Apoptosis and necroptosis. *Free Radical Biology & Medicine*, *139*, 1–15. https://doi.org/10.1016/J.FREERADBIOMED.2019.05.017

Silvester, E., Ivens, A., & Matthews, K. R. (2018). A gene expression comparison of Trypanosoma brucei and Trypanosoma congolense in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. *PLOS Neglected Tropical Diseases*, *12*(10), e0006863. https://doi.org/10.1371/JOURNAL.PNTD.0006863

Simarro, P. P., Diarra, A., Postigo, J. A. R., Franco, J. R., & Jannin, J. G. (2011). The Human African Trypanosomiasis Control and Surveillance Programme of the World Health Organization 2000–2009: The Way Forward. *PLoS Neglected*

Tropical Diseases, *5*(2). https://doi.org/10.1371/JOURNAL.PNTD.0001007

Skovgård, H., & Nachman, G. (2017). Modeling the Temperature- and Age-Dependent Survival, Development, and Oviposition Rates of Stable Flies (Stomoxys calcitrans) (Diptera: Muscidae). *Environmental Entomology*, *46*(5), 1130–1142. https://doi.org/10.1093/EE/NVX118

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59. https://doi.org/10.1002/CPMB.59

Sun, X., Liu, J., Xu, C., Tang, S. C., & Ren, H. (2016). The insights of Let-7 miRNAs in oncogenesis and stem cell potency. *Journal of Cellular and Molecular Medicine*, *20*(9), 1779–1788. https://doi.org/10.1111/JCMM.12861

Surridge, A. K., Lopez-Gomollon, S., Moxon, S., Maroja, L. S., Rathjen, T., Nadeau, N. J., Dalmay, T., & Jiggins, C. D. (2011). Characterisation and expression of microRNAs in developing wings of the neotropical butterfly Heliconius melpomene. *BMC Genomics*, *12*, 62. https://doi.org/10.1186/1471-2164-12-62

Ta, B. T. D., Kaba, D., Berte, D., Djohan, V., Acapovi-Yao, G. L., Rayaisse, J. B., Salou, E., Solano, P., & Dujardin, J. P. (2021). Tsetse Flies: Comparative Morphometric Information from Traits Collected on Wings and Pupae. *African Entomology*, *29*(2), 522–533. https://doi.org/10.4001/003.029.0522

Thakur, A., Kamboj, S., & Kumar, M. (2022). Computational resources for analysis of miRNA targetome. *MicroRNA in Human Malignancies*, 125–139. https://doi.org/10.1016/B978-0-12-822287-4.00020-7

Thirugnanasambantham, K., Hairul-Islam, V. I., Saravanan, S., Subasri, S., & Subastri, A. (2013). Computational Approach for Identification of Anopheles gambiae

miRNA Involved in Modulation of Host Immune Response. *Applied Biochemistry and Biotechnology*, *170*(2), 281–291. https://doi.org/10.1007/s12010-013-0183-5

Tirados Estebanez, J. I. (2014). *The role of host cues in the transmission of sleeping sickness*.

Tirumalai, V., Prasad, M., & Shivaprasad, P. V. (2020). Rna blot analysis for the detection and quantification of plant micrornas. *Journal of Visualized Experiments*, *2020*(161), 1–8. https://doi.org/10.3791/61394

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, *13*(9), 731–740. https://doi.org/10.1038/nmeth.3901

Venturelli, A., Tagliazucchi, L., Lima, C., Venuti, F., Malpezzi, G., Magoulas, G. E., Santarem, N., … & Costi, M. P. (2022). Current Treatments to Control African Trypanosomiasis and One Health Perspective. *Microorganisms 2022, Vol. 10, Page 1298*, *10*(7), 1298. https://doi.org/10.3390/MICROORGANISMS10071298

Vilmos, P., Bujna, Á., Szuperák, M., Havelda, Z., Várallyay, É., Szabad, J., Kucerova, L., … & Erdélyi, M. (2013). Viability, Longevity, and Egg Production of *Drosophila melanogaster* Are Regulated by the miR-282 microRNA. *Genetics*, *195*(2), 469–480. https://doi.org/10.1534/genetics.113.153585

Vinod Chandra, S. S. (2021). Bioinformatics of Prediction of Secondary Structures of Non-coding RNAs. *Bioinformatics and Human Genomics Research*, 222–241. https://doi.org/10.1201/9781003005926-11

Vishnoi, A., & Rani, S. (2017). MiRNA biogenesis and regulation of diseases: An overview. *Methods in Molecular Biology*, *1509*, 1–10. https://doi.org/10.1007/978-1-4939-6524-3_1/COVER

W, X., A, S. L., Z, W., & Y, L. (2014). Identifying microRNA targets in different gene regions. *BMC Bioinformatics*, *15 Suppl 7*(Suppl 7). https://doi.org/10.1186/1471-2105-15-S7-S4

Wamwiri, F. N., & Auma, J. E. (2021). *Overview of the Vectors and Their Role in Transmission of African Animal Trypanosomiasis*. 53–72. https://doi.org/10.4018/978-1-7998-6433-2.CH003

Wamwiri, F. N., & Changasi, R. E. (2016). Tsetse Flies (Glossina) as Vectors of Human African Trypanosomiasis: A Review. *BioMed Research International*, *2016*. https://doi.org/10.1155/2016/6201350

Wang, X., Li, Y., Zhang, J., Zhang, Q., Liu, X., & Li, Z. (2017). De novo characterization of microRNAs in oriental fruit moth Grapholita molesta and selection of reference genes for normalization of microRNA expression. *PloS One*, *12*(2), e0171120. https://doi.org/10.1371/journal.pone.0171120

Wu, J., Liu, Q., Wang, X., Zheng, J., Wang, T., You, M., Sheng Sun, Z., & Shi, Q. (2013). mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. *RNA Biology*, *10*(7), 1087–1092. https://doi.org/10.4161/rna.25193

Wu, Y.-C., Chen, C.-H., Mercer, A., & Sokol, N. S. (2012). let-7-Complex MicroRNAs Regulate the Temporal Identity of Drosophila Mushroom Body Neurons via chinmo. *Developmental Cell*, *23*(1), 202–209. https://doi.org/10.1016/j.devcel.2012.05.013

Yang, L., Weiss, B. L., Williams, A. E., Aksoy, E., de Silva Orfano, A., Son, J. H., Wu, Y., … & Aksoy, S. (2021). Paratransgenic manipulation of a tsetse microRNA alters the physiological homeostasis of the fly's midgut environment. *PLOS Pathogens*, *17*(6), e1009475. https://doi.org/10.1371/JOURNAL. PPAT. 1009475

Yen, P. S., Chen, C. H., Sreenu, V., Kohl, A., & Failloux, A. B. (2019). Assessing the Potential Interactions between Cellular miRNA and Arboviral Genomic RNA in the Yellow Fever Mosquito, Aedes aegypti. *Viruses 2019, Vol. 11, Page 540*, *11*(6), 540. https://doi.org/10.3390/V11060540

Yerukala Sathipati, S., & Ho, S. Y. (2018). Identifying a miRNA signature for predicting the stage of breast cancer. *Scientific Reports 2018 8:1*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-34604-3

Yousef, M., Parveen, A., & Kumar, A. (2022). Computational Methods for Predicting Mature microRNAs. *Methods in Molecular Biology*, *2257*, 175–185. https://doi.org/10.1007/978-1-0716-1170-8_9/COVER

Zhang, B., Pan, X., Wang, Q., Cobb, G. P., & Anderson, T. A. (2006). Computational identification of microRNAs and their targets. In *Computational Biology and Chemistry* (Vol. 30, Issue 6, pp. 395–407). https://doi.org/10.1016/j. compbiolchem.2006.08.006

Zhang, G., Hussain, M., O'Neill, S. L., & Asgari, S. (2013). 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*, *110*(25), 10276–10281. https://doi.org/10.1073/pnas. 1303603110

Zhang, Y., Yun, Z., Gong, L., Qu, H., Duan, X., Jiang, Y., & Zhu, H. (2018).

Comparison of miRNA Evolution and Function in Plants and Animals. *MicroRNA*, *7*(1), 4–10. https://doi.org/10.2174/2211536607666180126163031

Zhao, F., Wang, Z., Lang, H., Liu, X., Zhang, D., Wang, X., Zhang, T., Wang, R., Shi, P., & Pang, X. (2015). Dynamic Expression of Novel MiRNA Candidates and MiRNA-34 Family Members in Early- to Mid-Gestational Fetal Keratinocytes Contributes to Scarless Wound Healing by Targeting the TGF-β Pathway. *PloS One*, *10*(5), e0126087. https://doi.org/10.1371/journal.pone.0126087

APPENDICES

Appendix I: Manuscript publication

The developmentally dynamic microRNA transcriptome of *Glossina pallidipes* tsetse flies, vectors of animal trypanosomiasis

Careen Naitore, Jandouwe Villinger ✉, Caleb K Kibet, Shewit Kalayou, Joel L Bargul, Alan Christoffels, Daniel K Masiga ✉

Bioinformatics Advances, Volume 2, Issue 1, 2022, vbab047,
https://doi.org/10.1093/bioadv/vbab047
Published: 28 December 2021 **Article history**

PDF Split View Cite Share

Abstract

Summary

MicroRNAs (miRNAs) are single stranded gene regulators of 18–25 bp in length. They play a crucial role in regulating several biological processes in insects. However, the functions of miRNA in *Glossina pallidipes*, one of the biological vectors of African animal trypanosomosis in sub-Saharan Africa, remain poorly characterized. We used a combination of both molecular biology and bioinformatics techniques to identify miRNA genes at different developmental stages (larvae, pupae, teneral and reproductive unmated adults, gravid females) and sexes of *G. pallidipes*. We identified 157 mature miRNA genes, including 12 novel miRNAs unique to *G. pallidipes*. Moreover, we identified 93 miRNA genes that were differentially expressed by sex and/or in specific developmental stages. By combining both miRanda and RNAhybrid algorithms, we identified 5550 of their target genes. Further analyses with the Gene Ontology term and KEGG pathways for these predicted target genes suggested that the miRNAs may be involved in key developmental biological processes. Our results provide the first repository of *G. pallidipes* miRNAs across developmental stages, some of which appear to play crucial roles in tsetse fly development. Hence, our findings provide a better understanding of tsetse biology and a baseline for exploring miRNA genes in tsetse flies.

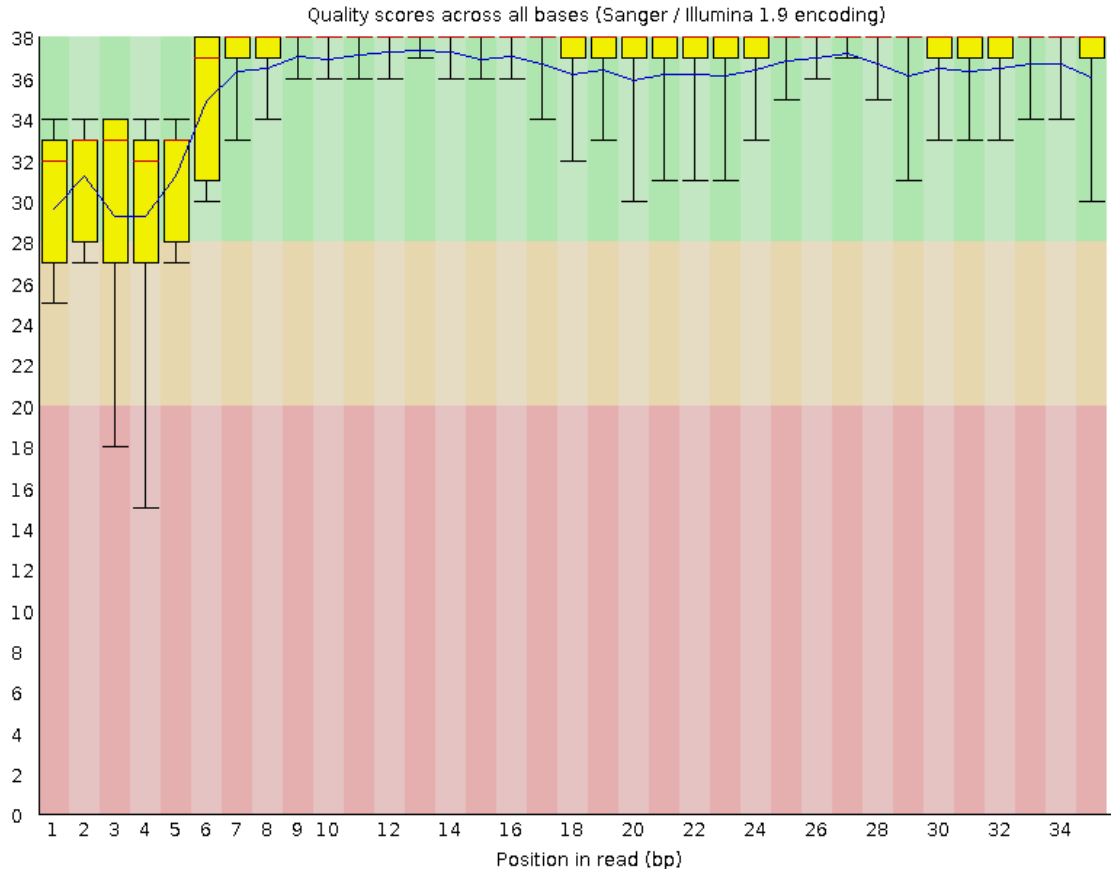
Availability and implementation

Raw sequence data are available from NCBI Sequence Read Archives (SRA) under Bioproject accession number PRJNA590626.

Supplementary information

Supplementary data are available at *Bioinformatics Advances* online.

Appendix II: An overview of the range of quality values across all bases at each position in the Fastq file for library.



For each position a Box Whisker type plot is drawn. The elements of the plot are as follows; the central red line is the median value, the yellow box represents the inter-quartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points, the blue line represents the mean quality. The y-axis on the graph shows the quality scores. The higher the score the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The quality of calls in the begin had as the run progresses, so it is common to see base calls falling into the orange area towards the end of a read

Appendix III: The known MicroRNAs identified in the four developmental stage libraries of G. pallidipes

MicroRNA	Mature sequence	matur e arm	Precursor coordinates: strand position	Precursor sequence
Gpa-miR-283	cggaauuucaguugguaucga	3p	KK499783.1_Scaffold19:1470952-1471009:+	aaauaucagcugguaauucugggauauuuauaacucggaauuucaguugguauaaau
Gpa-miR-8489a	aaaauucauuuugaauauuguu	3p	KK499819.1_Scaffold55:429510-429450:-	Cgaucuucaaaaugaauuuucguagauuuuaacuaucggaaaauucauuuugaauauuguu
Gpa-miR-79	gcuuuggcgcuuuagcuguauga	5p	KK499824.1_Scaffold60:382325-382387:-	Gcuuuggcgcuuuagcuguaugauagauuugauaccuucauaaagcuagauuaccaaagcau
GPA-miR-4968	cagcaacagcagcagcagcaga	3p	KK499844.1_Scaffold80:1472300-1473255+	Uuuuccucagagccgcgacugcuggcguguuauuguugcuguugcagcuagcugccacgguuuguugcuucacugauuacgcgcaaugcagcaacagcagcagcagcagaagcaacagcaacuaacugcagaggagcaauagcag

				cugagaguuc
Gpa-miR-9c	uaaagcuuuuauaccaaagcuc	3p	KK499824.1_Scaffold60:384436-384498:-	Ucuuugguauucuagcuguagauucauaauaaacuaucaucuaaagcuuuuauaccaaagcu
Gpa-miR-307a	acucacucaaccuggugugau	5p	KK499832.1_Scaffold68:383020-383082:+	Acucacucaaccuggugugauguuuuagaucugguauccaucacaaccuccuugaggugagc
GPA-miR-d	uguguacauuguguaguuuuu	-	KK499919.1_Scaffold157:370815-370898:+	aaagcuaccuaauuguacguagcguuuuugaauuucuaaauauaaaauguuuugaauuuugcgcuguguacauuguguaguuuuu
Gpa-miR-iab-4	cgguauaccuucaguauacguaac	3p	KK499939.1_Scaffold177:289199-289258:-	Acguauacugaauguauccugagugcuuccuauccgguauaccuucaguauacguaaca
Gpa-miR-219	aggguugcgacugggcaucgcg	3p	KK499772.1_Scaffold8:1087955-1088027:-	Uaauucgauuuuuagcuaugauugucaaacgcaauucuuguugauauucaauauucaaggguugcgacugggcaucgcggcucgaaauaagaauacaac

Gpa-miR-2491	Caacaacagcagcagcaa	3p	KK499771.1_Scaffold7:2209803-2209856:+	cgcgaaccacuuugugccgcuguugccuugcaguugcuguuuuccauguugcuacugcggccaauuguugcgcaucagcaacaucaacaacaucaacaacagcagcagcaacugcaguagcagcacucgugauucacuuucga
Gpa-miR-31a	cgcuaugcugcaucuagucagu	3p	KK499775.1_Scaffold11:1615005-1615097:+	Uccguugguaaauuggcaagaugucggcauagcugacguugaaaagcgauuuugaagagcgcuaugcugcaucuagucaguuguucaaugga
Gpa-miR-994	cacaguugcuguuucuuuuagau	3p	KK499896.1_Scaffold132:422742-422809:+	Cuaaggaaauaguagccgugauuuuauuuaacaacuuacaccaucacaguugcuguuucuuuuagau
Gpa-miR-2507b	aauucggugagcuaugaucauu	5p	KK499803.1_Scaffold39:1396535-1396487:-	Aauucggugagcuaugaucauugcgaucuagcauauucaccguauuga
Gpa-miR-184	ccuuaucauucucucgccccg	5p	KK499975.1_Scaffold213:299027-299089:+	Gguuggccggugcauucguacccuuaucauucucucgccccgugugcacuuaaagacaacuggacggagaacugauaagggcucgucaucaccaauucauc

Gpa-miR-210	agcugcuggccacugcacaagau	5p	KK499913.1_Scaffold150:167176:167269:+	Aaaggugcuauugcagcugcuggccacugcacaagauuagacuuaagacucuugugcgugugacagcggcuauuguaagaggccauagaagcaacagcc
Gpa-let-7	cuauacaaugugcuagcuuucu	3p	KK499824.1_Scaffold60:1061873-1061940:+	Ucuggcaaauugagguaguagguuguauaguaguaauuacacaucauacuauacaaugugcuagcuuucuuugcuuga
Gpa-miR-100	caagaccggcauuaugggaguc	3p	KK499824.1_Scaffold60:1061479-1061540:+	Ccauuaacagaaacccguaaauccgaacugugcuguuuuauaucuguuacaagaccggcauuaugggagucugucaaugcaaacaacugguuuuuggca
Gpa-miR-7	caauaaaucccuugucuucuua	3p	KK499852.1_Scaffold88:1014371-1014433:-	Uggaagacuagugauuuuguuguuuagguauuuguaauaacaauaaaucccuugucuucuua
Gpa-miR-h	aggaugucgugagggcuucc	-	KK500030.1_Scaffold269:179208-179149:-	Uauagcauccacgacaauauagaucauauuuuucaacuaggaugucgugaaugcuguagc

Gpa-miR-1	ccaugcuuccuugcauucaaua	5p	KK499803.1_Scaffold39:433989-434050:+	Uucagccuuugagaguuccaugcuuccuugcauucaauaguuauauucaagcauauggaauguaaagaaguauggagcgaaaucuggcgag
Gpa-miR-252	uccugcugcccaagugcuuauu	3p	KK499942.1_Scaffold180:524238-524306:+	Accaaguucgcuuuccuaaguacuagugccgcaggaguuagguucguguccgcaauaccuccugcugcccaagugcuuauuaaagcggcgagu
Gpa-miR-263a	cgugaucucuuaguggcaucua	3p	KK499809.1_Scaffold45:994005-994084:+	Aauggcacuggaagaauucacgggcuuuuaaaaguuauuucuuauauaauaucuaccgugaucucuuaguggcaucua
Gpa-miR-190	cccaggaaucaaacauauuauu	3p	KK499764.1_Scaffold0:471096-471159:+	Cgaacuaauugaugguuccagugagauauguuugauauucuugguuguuucauucaaaaguucacccaggaaucaaacauauuauuacugugacccucgc
Gpa-miR-iab-8	uuacguauacugaagguaua	5p	KK499939.1_Scaffold177:289201-289261:+	Uuacguauacugaagguauaccggauaggaagcacucaggauacauucaguauacguaua

Gpa-miR-993	gaagcucgucucuacagguaucu	5p	JMRQ01006282.1_sequence:139948-140029:+	Gaagcucgucucuacagguaucucacaggguagaaguaaucaaccuuaagugaaguuuccaaaguaguuuugugucuuuuu
Gpa-miR-5	uaucacagugauuuuccuuuaua	3p	KK499790.1_Scaffold26:90895-90953:+	Gcuaaaaggaacgaucguugugauaugaguuguuuccuaacauaucacagugauuuuccuuuauaacgc
GPA-miR-a	aaaacuuaaaaauuugcucagg	-	KK499786.1_Scaffold22:772187-772276:-	Gugaguugaaauuucuaaguuuuaagauaugaaguuuuagccuaucccgguguaaaagacucugauuaaaacuuaaaaauuugcucagg
Gpa-miR-137	uauugcuugagaauacacguag	3p	KK499774.1_Scaffold10:2419385-2419446:+	Acgcguauucuuagguuauuaacaugucuacgaaauuguuauugcuugagaauacacguag
Gpa-miR-2521	gauaauuuacguuuugucggucu	-	KK499969.1_Scaffold207:165518:165471:-	Acacugacgaugugaauuuuccuauugauaauuuacguuuugucggucu
Gpa-miR-282	acauagccuauaagagguuagg	3p	KK499772.1_Scaffold8:2954604-2954683:-	Uagccucuacuaggcuuugucugugucccacaugcgauguggaguuauuaaaauguuca

				gacauagccuaaaagagguuag
Gpa-miR-2788	uaaugcccuuagaaauuucaa	3p	KK499793.1_Scaffold29:988131-988195:-	Ugaaauuuuuaagcggcacuaacucgcaacauaacugauuugauaaugcccuuagaaauuucaa
Gpa-miR-968	uaaguaguauccuagacgaaguu	5p	KK49956.1_Scaffold194:173272-173208:-	Uaaguaguauccuagacgaaguuaaaaaucuguauuaaaaaggcuucuucaggauacugcucuau
Gpa-miR-e	ugaaguacgaugauucauucuug	-	KK499919.1_Scaffold157:371861-371929:+	Uagaaugaaacaaucguuuucagcguuuaauuuuacaacauauggcugaaguacgaugauucauucuug
Gpa-miR-g	uaccgaucucguaugaaagauu	-	KK500030.1_Scaffold269:148730-146772:+	Uccaucauacuugcuauugguaaucaaaaaaaaaacauguuaccgaucucguaugaaagauu
Gpa-miR-318	ggauacacacaguucaguuuug	3p	KK499896.1_Scaffold132:422970:42294:+	Uuuaugggauacacacaguucaguuuugucacacuucaagcaucacugggcuuuguuuaucucaugag
Gpa-miR-284	gucagcaacuugauuccagca	3p	KK499800.1_Scaffold36:918591-	Guugcaguuccuggaauuaaguugacug

			918684:+	uguagccugugagggcaaggcuugaauaaugcuccugaagucagcaacuugauuccagcaauugcggcccga
Gpa-miR-f	uauaccgaucaauucagaauc	-	KK499963.1_Scaffold201:126348-126377:+	Uauaccgaucaauucagaaucauuaucugaauugaucuaggcc
Gpa-miR-929	cucccuaacggagucagauug	3p	KK499792.1_Scaffold28:1414131-1414190:-	Aguccugguggagcucaaauugacucuaguagggaguccuuuaaugagcgacucccuaacggagucagauugagcugcaaaggagcga
Gpa-miR-79	uaaagcuagauuaccaaagcau	3p	KK499824.1_Scaffold60:382325-382387:-	Gcuuuggcgcuuuagcuguaugauagauuugauaccuucauaaagcuagauuaccaaagcau
Gpa-miR-994	cuaaggaaauaguagccgugau	5p	KK499896.1_Scaffold132:422742-422809:+	Cuaaggaaauaguagccgugauuuuauuuaacaacuuacaccaucacaguugcuguuucuuuuagau
Gpa-miR-988	gugugauuuguagcaaagugau	3p	KK499825.1_Scaffold61:660040-660128:+	Gacggcgguaccgggcauuuugggugugugauuuguagcaaagugauauguauuugaucaucccccuuguugcaaaccucacgccaa

				agaugaucugcga
Gpa-miR-219	ugauuguccaaacgcaauucu	5p	KK499772.1_Scaffold8:1087955-1088027:-	Uaauucgauuuuuagcuaugauugucca aacgcaauucuuguugauauucaauauuc aaggguugcgacugggcaucgcggcucga aauaagaauacaac
Gpa-miR-317	ugggauacacccugugcucgcu	5p	KK499952.1_Scaffold190:146208-146286:+	Auacacccugugcucgcuuugaaauuuu cuauuaauaaaauuuuauaugauaucagu gaacacagcuuggugguaucu
Gpa-miR-87	uugagcaaaauuucaggugug	3p	KK499824.1_Scaffold60:202317-202378:+	Caccuguaucuugcucaaccguuacuuau auaaaauccgguugagcaaaauuucaggu gug
Gpa-miR-993	gaagcucgucucuacagguaucu	3p	JMRQ01006282.1_sequence:139948-140029:+	Gaagcucgucucuacagguaucucacagg guagaaguaaucaaccuuaagugaaguuu ccaaaguaguuuugugucuuuuu
Gpa-miR-929	aaauugacucuaguagggagu	5p	KK499792.1_Scaffold28:1414131-1414190:-	Aguccugguggagcucaaauugacucua guagggaguccuuuaaugagcgacucccu aacggagucagauugagcugcaaaggagc

				ga
Gpa-miR-6	uaucacaguggcuguucuuuuu	3p	KK499790.1_Scaffold26:93729-93811:+	gagaacaguuauugcagauauguuuuccauuugagaaauucaucaacauuuucauauuuauaucacaguggcuguucuuuuu
Gpa-miR-iab-4	acguauacugaauguauccuga	5p	KK499939.1_Scaffold177:289199-289258:-	Acguauacugaauguauccugagugcuuccuauccgguauaccuucaguauacguaaca
Gpa-miR-980	uagcugccuucugaagggcuucc	3p	KK500133.1_Scaffold376:118867-18888:+	Aguugauuguaugucaguuuuucauuuggccuggcuagcuuacuccuuuuuaaaauaugcuagcugccuugugaagggcuuacguguaauugcaguuc
Gpa-miR-285	acugagaucgauuggugcauaga	5p	KK499768.1_Scaffold4:437703-437790:-	Ucgaaucgaagaacugagaucgauuggugcauagauaucaggagaacccacucaauuuaacucuagcaccauucgaaaucagugcuuuugauaagaaac
Gpa-miR-286	ggcgaaugucgguauggucucu	5p	KK499790.1_Scaffold26:90597-90695:+	Uuaaaauugaauggcgaaugucgguauggucucuuuuucaaagaaagguuucgauuaagcgaagugacuagaccgaacacucgugc

				uauaauuuuaaaauc
Gpa-miR-275	cgcgcuaaucagugaccggggcu	5p	KK500075.1_Scaffold315:241100-241168:+	Uguaaagucuccuaccuugcgcgcuaaucagugaccggggcugguuuuuuauauacagucagguaccugaaguagcgcgcguggugcagacauauau
Gpa-miR-125	acaaguuuugaucuccgguau	3p	KK499824.1_Scaffold60:1062703-1062763:+	Ucccugagacccuaacuugugacuuaucacuugaguuucacaaguuuugaucuccgguau
Gpa-miR-927	caaagcguuuggauucugaaac	3p	KK499855.1_Scaffold91:322630-322691:-	Uuuagaauuccuacgcuuuaccuuaugaucugaaauuggcaaagcguuuggauucugaaac
Gpa-miR-124	gguauccacuguaggccucuaug	5p	KK499862.1_Scaffold98:1265604-1265662:-	Ugugggagcgagacggggacucacugugcuuauuaaauagucagucuuuuucucucuccuaua
Gpa-miR-b	uccacuggauaaaguuugucuug	-	KK499790.1_Scaffold26:89500-89567:+	Cgauaaacuuugugcaggacuggugaccauuuuugguaacgaucauccacuggauaaaguuugucuug

Gpa-miR-9b	uagagcuuuauuaccaaaaacc	3p	KK499824.1_Scaffold60:382082-382145:-	Ucuuuggugauuuuagcuguaugauuuacaauuauauuccauagagcuuuauuaccaaaaacc
Gpa-miR-971	uugguguuacuucuuacaguga	3p	KK499780.1_Scaffold16:347671-347733:+	Gcuguaagaugguaacaacaaagcugaguuauguuauugcuugguguuacuucuuacaguga
Gpa-miR-996	ugacuagauuucaugcucgucu	5p	KK499818.1_Scaffold54:1375134-1375199:+	Gcgaacauggaucuagugcacgguuuaucauaaucaaguucgugacuagauuucaugcucgucu
Gpa-miR-193	uugggauuuuuuagaucagcag	5p	KK499793.1_Scaffold29:9844-98342:-	Ugugugcccuauuaugguugggauuuuuuagaucagcaguuauugcuauauagccauauuuauaaaaucuucuacuggccuacuaagucccaacauaaugagaguaaa
Gpa-miR-c	acacuggaccaaagugugucucg	-	KK499790.1_Scaffold26:89646-89709:+	Agauacaccuuguccaguuuauguuuccgacgauuggacaaacacuggaccaaagugugucucg
Gpa-miR-2a-2	ucacagccagcuuugaugagcu	5p	KK499782.1_Scaffold18:2343307-	Ccucaucaaggguugugauauggauuaucaacgcauaucacagccagcuuugauga

			2343367:-	gcu
Gpa-miR-989	ugugaugugacguaguggaaca	3p	KK500228.1_Scaffold477:52361-52445:+	ugugaugugacguaguggaacauuccugacguuuugcaaacuacauguacuucuuuauaaugcgcacacauacacacacacaca
Gpa-miR-286	ugacuagaccgaacacucgcgcu	3p	KK499790.1_Scaffold26:90608-90682:+	Cgcgauugucggcuuagucucucuuucguuuuaucauauuuccauauugagugacuagaccgaacacucgcgcu
Gpa-miR-263b	cuuggcacugggagaauucacag	5p	KK499881.1_Scaffold117:224424-224487:+	Cuuggcacugggagaauucacagucgauacauuagaauucugugguucuuugggugccaaaac
Gpa-miR-193	uacuggccuacuaagucccaac	3p	KK499793.1_Scaffold29:9844-98342:-	Ugugugcccuuauuaugguugggauuuuuuagaucagcaguuauugcuauauagccauauuuauaaaaucuucuacuggccuacuaagucccaacauaaugagaguaaa
Gpa-miR-5	aaaggaacgaucguugugauaug	5p	KK499790.1_Scaffold26:90895-90953:+	Gcuaaaaggaacgaucguugugauaugaguuguuuccuaacauaucacagugauuuuccuuuauaacgc

Gpa-miR-j	uccagcggacggugcugcuccu	-	KK500152.1_Scaffold395:107143-107069:-	Uccagcggacggugcugcuccuccagcggaaggggcugcuccuccagcggaaggagcugcaccuccagcggaagg
Gpa-miR-i	uagcugccuucugaagggcuucc	-	KK500097.1_Scaffold338:26390-26450:+	Caguccuucaauuggcaauuuagcuguuuuuuuauuguuagcugccuucugaagggcuucc
Gpa-miR-965	uaagcguauagcuuuuccccu	3p	KK500130.1_Scaffold373:109288-109379:-	Ggggugacgcuguacgcuguauguguuguuuucauagauucuuuggaaagacaaauucuuuuaauaaucauaagcguauagcuuuuccccu
GPA-miR-2760	guuuuggguuuaaggaguu	-	KK499941.1_Scaffold179:417870:417929:+	Ccuuuuaaaaacaagauuguuggcaauguguugcgccgucuuguuuuugguuuaaggaguu
Gpa-miR-932	ucaauuccguagugcauugcagu	5p	KK499866.1_Scaffold102:630600-630664:-	Ucaauuccguagugcauugcaguuugucgauuuaauaauuacugcaagcauuguggaaaugaug
Gpa-bantam	ccgguuuucgauuugguuugacu	5p	KK499768.1_Scaffold4:3165912-	Ccgguuuucgauuugauuugacuuauuuucacaaagugagaucauuuugaaagcuga

			3165971:-	uu
Gpa-miR-100	aacccguaaauccgaacuugu	5p	KK499824.1_Scaffold60:1061479-1061540:+	Ccauuaacagaaacccguaaauccgaacuugugcuguuuuauaucuguuacaagaccggcauuaugggagucugucaaugcaaacaacuggguuuuuggca
Gpa-miR-981	uucguugucgacgaaaccugca	3p	KK499853.1_Scaffold89:68720-68782:-	Cgggguuucguaaacagcgcacuguauacgauaaaacuaaguucguugucgacgaaaccugca
Gpa-miR-285	uagcaccauucgaaaucagugc	3p	KK499768.1_Scaffold4:437703-437790:-	Ucgaaucgaagaacugagaucgauuggugcauagauaucaggagaacccacucaauuuaacucuagcaccauucgaaaucagugcuuuugauaagaaac
Gpa-miR-124	uaaggcacgcggugaaugcc	3p	KK499862.1_Scaffold98:1265604-1265662:-	Ugugggagcgagacggggacucacuguggcuuauuaaauagucagucuuuuucucucuccuaua
Gpa-miR-276a	agcgagguauagaguuccuacg	5p	KK499833.1_Scaffold69:779760-779821:-	Agcgagguauagaguuccuacguuauaauauuaucuguaggaacuucauaccgug

				cucu
Gpa-miR-276b	cagcgagguauagaguuccuacg	5p	KK499833.1_Scaffold69:987273-987339:-	Aaaaccgaagucuuuuuaccaucagcgaggguauagaguuccuacguuccuauauucagucguaggaacuuaauaccgugcucuuggaggacugucgacc
Gpa-miR-284	ccuggaauuaaguugacuguguagcc	5p	KK499800.1_Scaffold36:918591-918684:+	Guugcaguuccuggaauuaaguugacuguguagccugugagggcaaggcuugaauaaugcuccugaagucagcaacuugauuccagcaauugcggcccga
Gpa-miR-998	uagcaccaugagauucagcuc	5p	KK500148.1_Scaffold391:34784-34867:-	acugaauucucgugggucugcauagcauaccauauaucguguguacaguaauugaaaauuuguagcaccaugagauucagcuc
Gpa-miR-305	auuguacuucaucaggugcucu	5p	KK500075.1_Scaffold315:241260-241319:+	Auuguacuucaucaggugcucuggugcauuaaauacccggcacauguugaaguacauuc
Gpa-miR-34	uggcaguguggguuagcugguu	3p	KK499952.1_Scaffold190:182043-	Uggcaguguggguuagcugguuguguaauuuuauuuaaaaucacagccacuaucuuca

			182108:+	cugccgcc
Gpa-miR-190	agauauguuugauauucuugguug	5p	KK499764.1_Scaffold0:471096-471159:+	Agauauguuugauauucuugguuguugcauuuugucuuucacccaggaaucaaacauauuauu
Gpa-miR-316	ugucuuuuuccgcuuacuggcg	5p	KK499926.1_Scaffold164:76825:76899:+	Aaauucuagucgauuugucuuuuuccgcuuacuggcguuucaauuccacaacgacaggaaagggaaaaaggcguauuuacuaugaguuu
Gpa-miR-92a	cauugcacuugucccggccuau	3p	KK499822.1_Scaffold58:1279274:1279355:+	Aauaugaauuucccguaggacgggaaggugucaacguuuugcauuucgaauaaacaugcacuugucccggccuaugggcgguuuguaauaaaca
Gpa-miR-308	aaucacaggauuauacug	3p	KK500010.1_Scaffold249:70402-70459:-	guauauuuuuguguuuuuguguaaauuuaacaaaagucaaaucacaggauuauacug
Gpa-miR-996	gcgaacauggaucuagugcacg	3p	KK499818.1_Scaffold54:1375134-1375199:+	Gcgaacauggaucuagugcacgguuuaucauaaucaaguucgugacuagauuucaugcucgucu

Gpa-miR-1000	auauuguccugucacagcagu	5p	KK499902.1_Scaffold138:468597-468663:+	Auauuguccugucacagcaguaugugauuucuuucaauauauuacugcugggccgggggcauuaaca
Gpa-miR-998	acugaauucucgugggucugca	3p	KK500148.1_Scaffold391:34784-34867:-	acugaauucucgugggucugcauagcauaccauauaucgugugguacaguaauugaaaauuuguagcaccaugagauucagcuc
Gpa-miR-9b	ucuuuggugauuuuagcuguau	5p	KK499824.1_Scaffold60:382082-382145:-	Ucuuuggugauuuuagcuguaugauuuacaauuauauuccauagagcuuuauuaccaaaaacc
Gpa-miR-282	uagccucuacuaggcuuugucu	5p	KK499772.1_Scaffold8:2954604-2954683:-	Uagccucuacuaggcuuugucugugucccacaugcgauguggaguuauuaaaauguucagacauagccuaaaagagguuag
Gpa-miR-125	ucccugagacccuaacuuguga	5p	KK499824.1_Scaffold60:1062703-1062763:+	Ucccugagacccuaacuugugacuuaucacuugaguucacaaguuuugaucuccgguau
Gpa-miR-317	ugaacacagcuuggugguaucu	3p	KK499952.1_Scaffold190:146208-146286:+	Auacacccugugcucgcuuugaaauuuucuauuaauaaaauuuuauaugauaucagu

				gaacacagcuuggugguaucu
Gpa-miR-278	ucggugggacuuucguccguuu	3p	KK499814.1_Scaffold50:555936-555999:+	Ccggacgaugguucacaacgaccguuuuuucuaacauuggucgguggacuuucguccguuu
Gpa-miR-283	aaauaucagcugguaauucug	5p	KK499783.1_Scaffold19:1470952-1471009:+	aaauaucagcugguaauucugggauauuuauaacucggaauuucaguugguauaaau
Gpa-miR-l	uccucguuugcuaacggacgccu	-	KK500272.1_Scaffold524:73940-74019:-	Cuaguauuuuccgggcuagcuacuugacggcuauucauaguuguuucgauuguugguccucguuugcuaacggacgccu
Gpa-miR-3479	uauugcacuugucccggccu	3p	KK499822.1_Scaffold58:1279285-1279344:+	Gugggaaaggugucaauguuuuagugaauaccaauuaauauugcacuugucccggccu
Gpa-miR-278	ccggacgaugguucacaacgacc	5p	KK499814.1_Scaffold50:555936-555999:+	Ccggacgaugguucacaacgaccguuuuuucuaacauuggucgguggacuuucguccguuu
Gpa-miR-307a	ucacaaccuccuugagugagc	3p	KK499832.1_Scaffold68:383020-383082:+	Acucacucaaccuggugugauguuuuagaucugguauccaucacaaccuccuugag

				ugagc
Gpa-miR-281	aagagagcuaucugucgacagu	3P	KK500053.1_Scaffold293:205739-205815:+	Aagagagcuaucugucgacagucugguuaaagccuguuccacuacagaauacaacugucauggaauugcucucuuu
Gpa-miR-7	uggaagacuagugauuuuguu	5p	KK499852.1_Scaffold88:1014371-1014433:-	Uggaagacuagugauuuuguuguuuagguauuuguaauaacaauaaaucccuugucuucuua
Gpa-miR-970	ucauaagacacacgcggcuau	3p	KK499806.1_Scaffold42:273920-273989:+	Agccagcguauguuuuauuugguaguuaaggaacauauuaauauccuaucauaagacacacgcggcuau
Gpa-miR-956	guguuuggaaugucucguuagcu	3p	KK499805.1_Scaffold41:1675730-1675697	Gaucguuaucguguuuggaaugucucguuagcuaacggaugagcaagugcucggcucacuggcccaaaugcaguuugcccggagacgccgguuaacccagcacugaaauguguaguuucgagaccacucuaauccauugcagcauuu
Gpa-miR-957	ugaaaccguccaaaacugaggc	3p	KK499771.1_Scaffold7:2697992-	Cuuagcuuugggcgggguuuuggugugu auggguuugaaaccauugaaaccguccaa

			2698057:-	aacugaggc
Gpa-miR-279	ugacuagauccacacucauu	3p	KK499818.1_Scaffold54:1350148-1350213:+	Agugagugaggguccaguguuucacaugguuuuuuucagauauugugacuagauccacacucauu
Gpa-miR-13a	uaucacagccauuuugaugaguu	3p	KK500136.1_Scaffold379:73448-73515:-	Ucuucaagauguugugaaaugucgcuuuggaauuuuucaauucauaucacagccauuuugaugaguu
Gpa-miR-133	uuggucccuucaaccagcug	3p	KK499803.1_Scaffold39:807488-807550:+	Agcugguugauaucgggucagaucuguuuuauugcuagcauuuggucccuucaaccagcug
Gpa-miR-274	uuugugaccgacacuaacgggu	5p	KK499768.1_Scaffold4:2050283-2050374:-	Uuugugaccgacacuaacggguaacugaaauauccuuuuauguguuuguuucuucauaacaaagauuuacucguuucugcgaucacaaauu
Gpa-miR-306	ucagguacuuagugacucucaa	5p	KK499824.1_Scaffold60:383894-383837:-	guccacucgauggcucagguacuuagugacucucaaugcuuuugacauuuugggggucacucugugccugugcugccagugggac

Gpa-miR-m	caaucucaguauguaaauguggu	-	KK499783.1_Scaffold19:1473705-1473771:+	Caaucucaguauguaaaugugguuagauuuguuuauauaugaauucacguugcauguggagauuacc
Gpa-miR-927	uuuagaauuccuacgcuuuacc	5p	KK499855.1_Scaffold91:322630-322691:-	Uuuagaauuccuacgcuuuaccuuaugaucugaaauuggcaaagcguuuggauucugaaac
Gpa-miR-995	uagcaccacaugauucggcuu	3p	KK499889.1_Scaffold125:392520-392589:-	Gcccggacccauguuugugcuguguaauuauuuaaacucaauaccucauagcaccacaugauucggcuu
Gpa-miR-13b	uaucacagccauuuugaugaguu	3p	KK499770.1_Scaffold6:536659-536722:+	Ccgccaaaacgguugugaauuauguauuuucuaauagacauaucacagccauuuugacgaguu
Gpa-miR-210	cuugugcgugugacagcggcu	3p	KK499913.1_Scaffold150:167176:167269:+	Aaaggugcuuauugcagcugcuggccacugcacaagauuagacuuaagacucuugugcgugugacagcggcuauuguaagaggccauagaagcaacagcc
Gpa-miR-8	caucuuaccgggcagcauuaga	5p	KK500001.1_Scaffold240:175491-	Caucuuaccgggcagcauuagauuucaugaacuaaauuucuaauacugucagguaa

			75555:+	agauguc
Gpa-miR-275	ucagguaccugaaguagcgcg	3p	KK500075.1_Scaffold315:241145-241186:+	Uguaaagucuccuaccuugcgcgcuaaucagugaccggggcugguuuuuuauauacagucagguaccugaaguagcgcgcguggugcagacauauau
Gpa-miR-10	caaauucgguucuagagagguuu	3p	JMRQ01006282.1_sequence:198125-198183:-	Ccacgucuacccuguagauccgaauuuguuuuauacuagcuuuaaggacaaauucgguucuagagagguuugugugg
Gpa-miR-315	uuuugauuguugcucagaaagcc	5p	KK499778.1_Scaffold14:123470-123535:-	Uuuugauuguugcucagaaagcccuauauuucauaucaguuggcuuucgagcaucaaucaaaaac
Gpa-miR-92b	aauugcacuagucccggccugc	3p	KK499822.1_Scaffold58:1294166-1294231:+	Agaccgugccuagugcuuauuugugacauuuuaauuuaauucaaauugcacuaguccccggccugc
Gpa-miR-2c	ucaucaaaaagggcugaagaaaga u	3p	KK500136.1_Scaffold379:74462:74535:+	Ucguaucuuacuuucaaugucaucaaaaagggcugaagaaagauauuucugcauuugaaucguaucacagccagcuuugaugggcau

				ugcaaugagcagcga
Gpa-miR-281-1	aagagagcuguccgucgacag	5p	KK500053.1_Scaffold293:205475-205539:+	Aagagagcuguccgucgacaguccaguucagacauauuaauacugucauggaauugcucucuuu
Gpa-let-7	ugagguaguagguuguauagu	5p	KK499824.1_Scaffold60:1061873-1061940:+	Ucuggcaaauugagguaguagguuguauaguaguaauuacacaucauacuauacaaugugcuagcuuucuuugcuuga
Gpa-miR-281-2	aagagagcuauccgucgacagu	5p	KK500053.1_Scaffold293:205471-205545:+	Cgaauugugaaaugaagagagcuauccgucgacagucaaguuaagaccgauuguaauacugucauggaauugcucucuuuguauaacauucg
Gpa-miR-12	agaguauuacuuagguacuggu	3p	KK499783.1_Scaffold19:1474287-1474371:+	agaguauuacuuagguacuggugugacuuuguaagaaaauuuucuuucgagaauaacaacagccaguacuuaugucauacuguc
Gpa-miR-9c	ucuuugguauucuagcuguag	5p	KK499824.1_Scaffold60:384436-384498:-	Ucuuugguauucuagcuguagauucauaauaaacuaucaucuaaagcuuuuauaccaaagcu

Gpa-miR-2b	uaucacagccagcuuugaggag	3p	KK499782.1_Scaffold18:2343731-2343802:-	Uucuucaaaguggcugcgaaauguuguacacagcucgauuuaauauucauaucacagccagcuuugaggag
Gpa-miR-2b	uaucacagccagcuuugaggag	3p	KK499837.1_Scaffold73:688955-689018:-	Uucuucagagcuguagugacauguagcuuuuuuaauauucauaucacagccagcuuugaggag
Gpa-miR-987	uaaaguaaaauagucuggauugau	5p	KK500082.1_Scaffold323:87571-87636:+	Uaaaguaaaauagucuggauugaugaaaggauuucaaagauuucaucaccagacauuuacuucaac
Gpa-miR-9a	uaaagcuagcuuaccgaaguua	3p	KK499791.1_Scaffold27:1664993-1665054:-	Ucuuugguuaucuagcuguaugagugauuaauaacgucauaaagcuagcuuaccgaaguua
Gpa-miR-958	aguagaauagcaggcuuaucaca	3p	KK499963.1_Scaffold201:126348-126377	Ggcgugucuauggcaaguagaauagcaggcuuaucacauguuuaauucaaucugcugugagauucuucuauucuacuuucgacaacacccgu
Gpa-miR-307	ucacaaccucuuugagugagcu	3p	KK499814.1_Scaffold50:84499-	Ccucacucaacuuggguuugauguguuaaaaaguuucgauagcaucacaaccucuuu

			84565:+	gagugagcu
Gpa-miR-305	cggcacauguugaaguacauuc	3p	KK500075.1_Scaffold315:241260-241319:+	Auuguacuucaucaggugcucuggugcauuaaauacccggcacauguugaaguacauuc
Gpa-miR-2a	uaucacagccagcuuugaugagc	3p	KK499782.1_Scaffold18:2343516-2343607:-	Aucuaagccucaucaaaguggguugugauauggauacccaacgcauaucacagccagcuugaugagcuaggau
Gpa-miR-2a	uaucacagccagcuuugaugagc	3p	KK500136.1_Scaffold379:74464-74531:-	ucaucaaaaggggcuguuguaugauauuuuugcauucauuaaguaucacagccagcuugaugagca
Gpa-miR-375	uuuguucguuuggcuuaaguu	3p	KK499812.1_Scaffold48:719527-719594:+	acuuaagccaagugaauacaaacauauaaaaauugauuccauacgguuuguucguuuggcuuaaguu
Gpa-miR-11	caagaacuuucucugugacccg	3p	KK500148.1_Scaffold391:1087701-1087675:-	gcacuugucaagaacuuucucugugacccgcguguacuuaaaagccgcaucacagucugaguucuugcugagugc
Gpa-miR-9a	ucuuugguuaucuagcuguauga	5p	KK499791.1_Scaffold27:1664993-	ucuuugguuaucuagcuguaugagugau

			1665054:-	uaauaacgucauaaagcuagcuuaccgaaguua
Gpa-miR-252	cuaaguacuagugccgcaggag	5p	KK499942.1_Scaffold180:524238-524306:+	accaaguucgcuuuccuaaguacuagugccgcaggaguuagguucguguccgcaauaccuccugcugcccaagugcuuauuaaagcggcgagu
Gpa-miR-34	cagccacuaucuucacugccgcc	5p	KK499952.1_Scaffold190:182043-182108:+	aauuggcuaugcgcuuuggcaguguggguagcugguuguguagccaauuauugccguugacaauucacagccacuaucuucacugccgccgcgacaagc
Gpa-miR-33	gugcauuguagucgcauuguc	5p	KK499791.1_Scaffold27:1646313-1646379:+	gugcauuguagucgcauuguccguguuauaaggaagaagucggccaauacuucugcaaugcaaacu
Gpa-miR-999	uguuaacuguaagacugugucu	3p	KK499930.1_Scaffold168:823175-823236:+	acauagucguacagaaaauauuguguuauacgaguccaauguuaacuguaagacugugucu
Gpa-miR-14	ucagucuuuuucucucuccuau	3p	KK500118.1_Scaffold360:74664-	ugugggagcgagacggggacucacugugcuuauuaaaauagucagucuuuuucucucuc

			74589:-	cuaua
Gpa-miR-31b	uggcaagaugucggaauagcuga	5p	KK499897.1_Scaffold133:89436-89512:+	uggcaagaugucggaauagcugaaaauuacaauaaucuauuuuacauauaacggcuaugccucaucuagucaauug
Gpa-miR-1	uggaauguaaagaaguauggag	3p	KK499803.1_Scaffold39:433989-434050:+	uucagccuuugagaguuccaugcuuccuugcauucaauaguuauauucaagcauauggaauguaaagaaguauggagcgaaaucuggcgag
Gpa-miR-31a	uggcaagaugucggcauagcuga	5p	KK499775.1_Scaffold11:1615005-1615097:+	uccguugguaaauuggcaagaugucggcauagcugacguugaaaagcgauuuugaagagcgcuaugcugcaucuagucaguuguucaaugga
GPA-bantam	ugagaucauuuugaaagcugauu	3p	KK499768.1_Scaffold4:3165912-3165971:-	ccgguuuucgauuugauuugacuuauuuucacaaagugagaucauuuugaaagcugauu
Gpa-miR-276b	uaggaacuuaauaccgugcucu	3p	KK499833.1_Scaffold69:987273-987339:-	aaaaccgaagucuuuuuaccaucagcgagguauagaguuccuacguuccuauauucagucguaggaacuuaauaccgugcucuugga

				ggacugucgacc
Gpa-miR-276a	uaggaacuucauaccgugcucu	3p	KK499833.1_Scaffold69:779760-779821:-	agcgagguauagaguuccuacguuauaauauauuaucuguaggaacuucauaccgugcucu
Gpa-mir-277	uaaaugcacuaucugguacgaca	3p	KK499952.1_Scaffold190:179086-179153:+	cguaucaggggugacauuugcacuggguuuucaauaguaguuuguaaaugcacuaucugguacgaca
Gpa-miR-263a	aauggcacuggaagaauucacgg	5p	KK499809.1_Scaffold45:994005-994084:+	aauggcacuggaagaauucacgggcuuuuaaaaguuauuucuuauauaaauaucuacccgugaucucuuaguggcaucua
Gpa-miR-8	uaauacugucagguaaagauguc	3p	KK500001.1_Scaffold240:175491-75555:+	caucuuaccgggcagcauuagauuucauugaacuaaaauuucuaauacugucagguaaagauguc
Gpa-miR-184	uggacggagaacugauaagggc	3p	KK499975.1_Scaffold213:299027-299089:+	gguuggccggugcauucguacccuuaucauucucucgccccgugugcacuuaaagacaacuggacggagaacugauaagggcucguaucaccaauucauc

Gpa-miR-10	acccuguagauccgaauuuguu	5p	JMRQ01006282.1_sequence:198125-198183:-	ccacgucuacccuguagauccgaauuuguuuuauacuagcuuuaaggacaaauucgguucuagagagguuugugugg

The novel miRNA identified in Appendix III:

1. Gpa-miR indicates that miRNA was identified from other insect species
2. Gpa-miR indicates that miRNA was uniquely identified from the species

miRNAs identified by mirdeep software from the *G. pallidipes* species: The table represents the mature ID, mature sequences, the position of the pre miRNA in genome and the premiRNA sequence.

Appendix IV: Arm selection preference of 5p arm and 3p arm miRNAs in G. pallidipes

3p arm count	5p arm count	MicroRNA
0	2430	Gpa-miR-1000
0	6405	Gpa-miR-1889
0	655	Gpa-miR-263b
0	6255	Gpa-miR-274
0	11059	Gpa-miR-281-1
0	14292	Gpa-miR-281-2
0	2650	Gpa-miR-282
0	538	Gpa-miR-2a-2
0	6371	Gpa-miR-306
0	24892	Gpa-miR-307
0	2347	Gpa-miR-308
0	8605	Gpa-miR-315
0	2234	Gpa-miR-316
0	99253	Gpa-miR-31b
0	68254	Gpa-miR-33
0	33206	Gpa-miR-375
0	1056	Gpa-miR-932

0	17485	Gpa-miR-987
0	42	Gpa-miR-iab-8
36412	0	Gpa-miR-11
15048	0	Gpa-miR-12
5772	0	Gpa-miR-133
73	0	Gpa-miR-137
5700	0	Gpa-miR-13a
6906	0	Gpa-miR-13b
95509	0	Gpa-miR-14
11	0	Gpa-miR-2491
198121	0	Gpa-miR-277
91	0	Gpa-miR-2788
5279	0	Gpa-miR-279
4266	0	Gpa-miR-281
28279	0	Gpa-miR-2a
28279	0	Gpa-miR-2a
17312	0	Gpa-miR-2b
17312	0	Gpa-miR-2b
10398	0	Gpa-miR-2c
186	0	Gpa-miR-318
3883	0	Gpa-miR-3479
6	0	Gpa-miR-4968

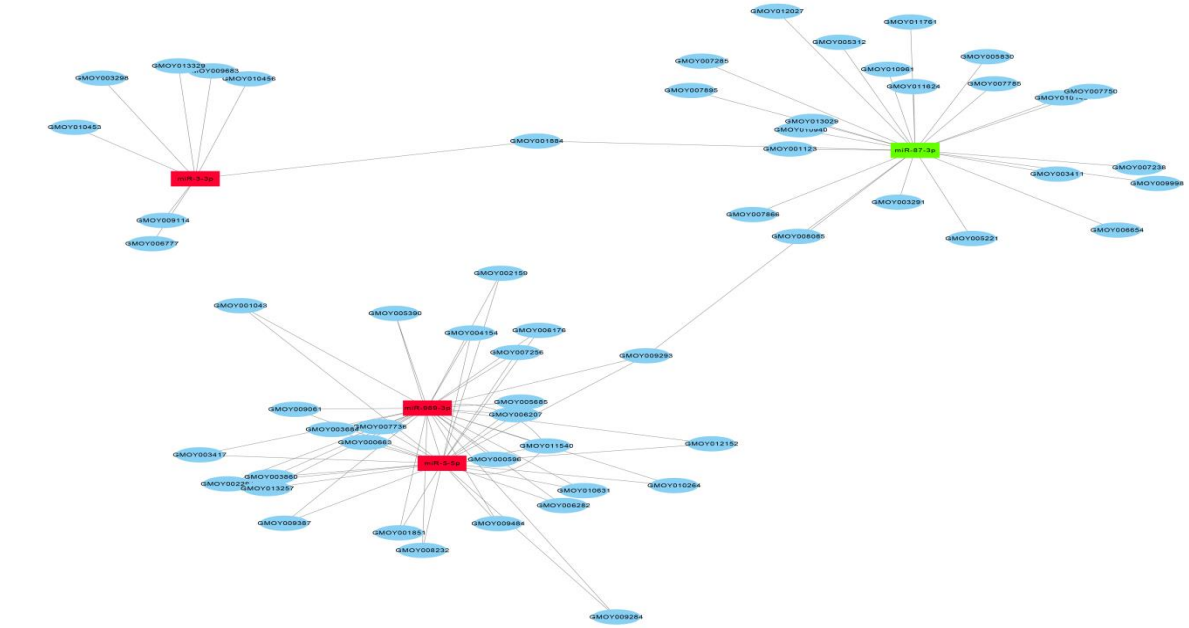
264	0	Gpa-miR-6
4	0	Gpa-miR-8489a
13	0	Gpa-miR-8513
248	0	Gpa-miR-87
2254	0	Gpa-miR-92a
8813	0	Gpa-miR-92b
4669	0	Gpa-miR-956
4685	0	Gpa-miR-957
2254	0	Gpa-miR-958
991	0	Gpa-miR-965
95	0	Gpa-miR-968
4644	0	Gpa-miR-970
400	0	Gpa-miR-971
277	0	Gpa-miR-980
1273	0	Gpa-miR-981
244	0	Gpa-miR-988
552	0	Gpa-miR-989
6729	0	Gpa-miR-995
85550	0	Gpa-miR-999
408456	16	Gpa-miR-184
111681	27	Gpa-miR-1
3936	7	Gpa-miR-307a

7623	21	Gpa-miR-210
163103	1432	Gpa-miR-276a
124775	1135	Gpa-bantam
130633	1432	Gpa-miR-276b
585259	8503	Gpa-miR-10
236	5	Gpa-miR-79
26869	1553	Gpa-miR-305
2974	247	Gpa-miR-317
2411	419	Gpa-miR-996
262	49	Gpa-miR-993
1278	284	Gpa-miR-285
1371	330	Gpa-miR-124
632	284	Gpa-miR-286
2445	1447	Gpa-miR-998
668	511	Gpa-miR-193
230	262	Gpa-miR-929
2991	3924	Gpa-miR-278
21003	36720	Gpa-miR-9a
396	2638	Gpa-miR-9b
212	1437	Gpa-miR-284
311	2815	Gpa-miR-125
50	711	Gpa-miR-5

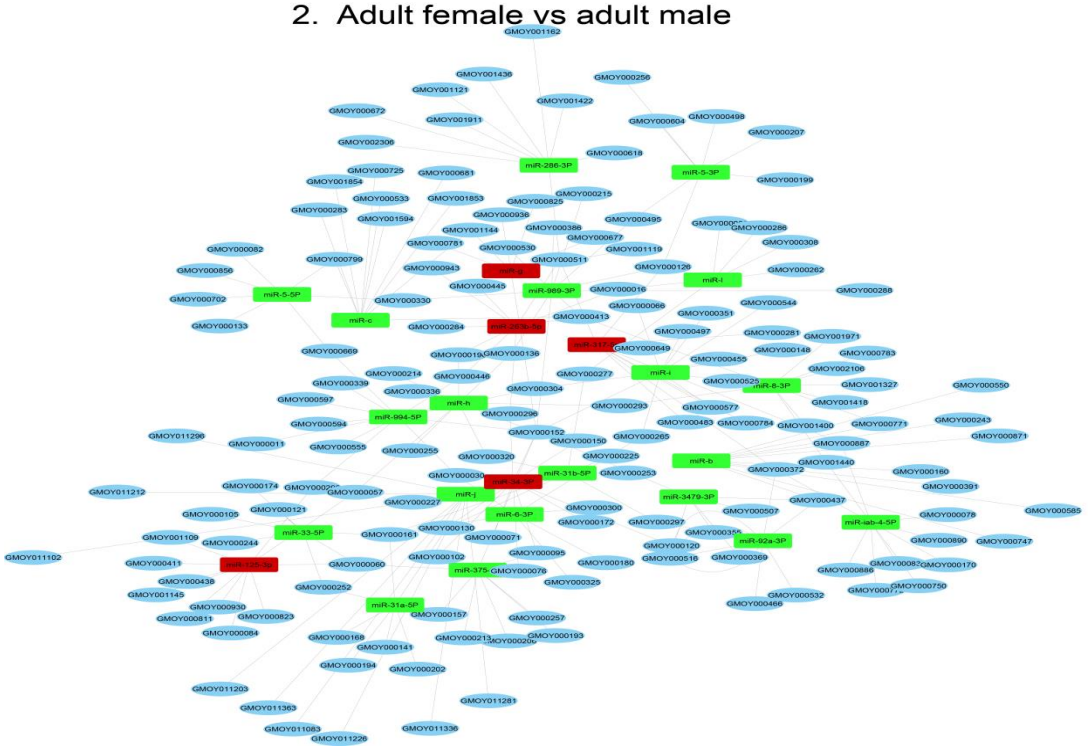
12	236	Gpa-miR-994
321	6528	Gpa-miR-927
11	247	Gpa-miR-219
285	8560	Gpa-miR-275
9	274	Gpa-miR-iab-4
1659	64170	Gpa-miR-34
405443	8029	Gpa-miR-8
24	1227	Gpa-miR-100
38	1980	Gpa-miR-190
24	4574	Gpa-miR-7
24	12072	Gpa-let-7
4	3253	Gpa-miR-283
28	49718	Gpa-miR-252
6	15464	Gpa-miR-9c
12	120581	Gpa-miR-31a
28	303750	Gpa-miR-263a

Appendix V: Sex-specific miRNA: mRNA networks

1. Teneral female vs teneral male



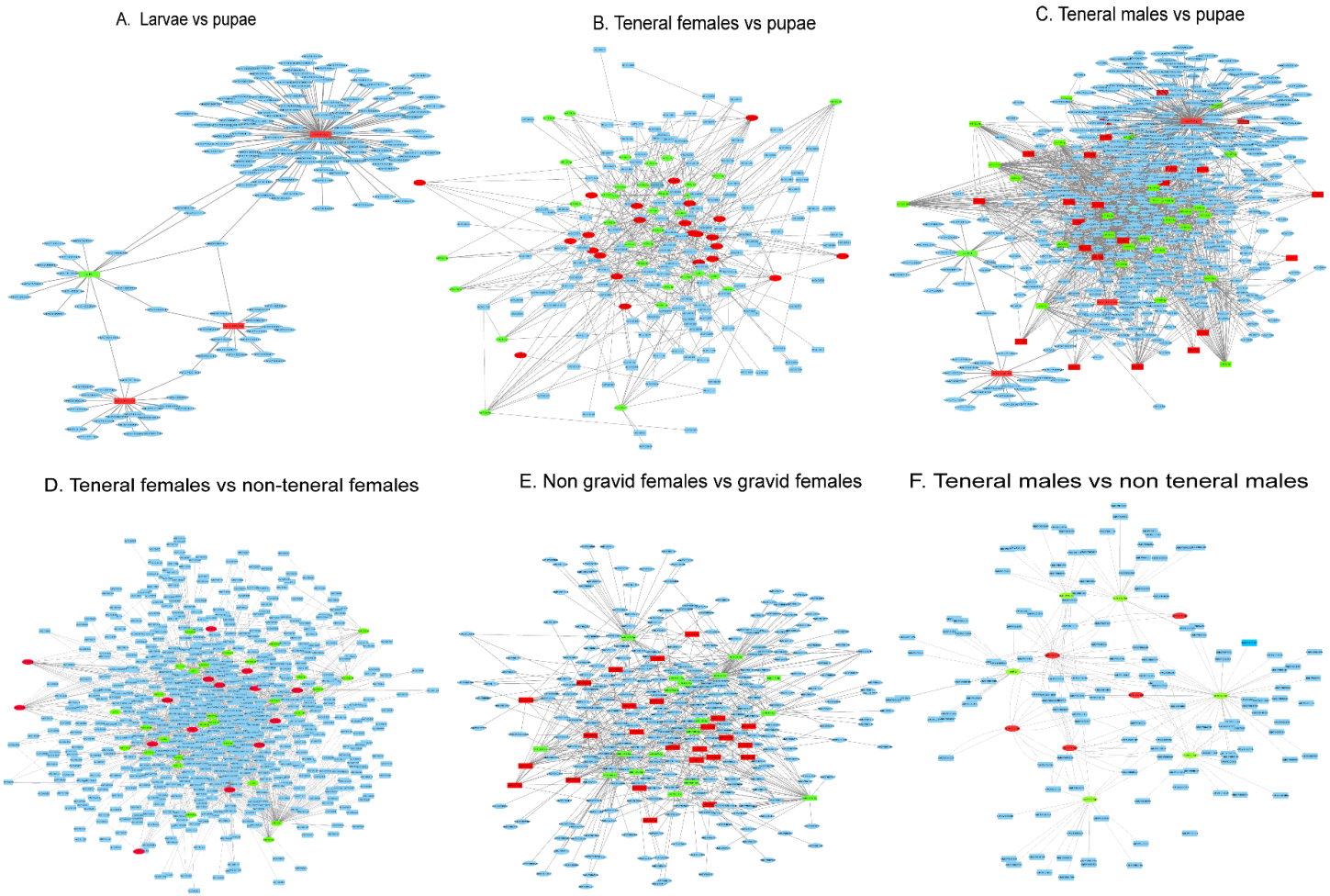
2. Adult female vs adult male



Visualisation of the networks in cytoscope: 1. Teneral females vs. males miRNA: mRNA network, 2. Non-teneral females vs. males miRNA: mRNA network:

The Green represents down regulated miRNA genes in the female specimen and the red represents up regulated miRNA genes in the female specimens, the blue represents the target mRNA genes. The lines connecting the miRNA and mRNA represent their interaction.

Appendix VI: Stage-specific miRNA: mRNA networks



Visualisation of the networks in cytoscope: (A) Larvae vs pupae specific miRNA-mRNA network, (B) Teneral female vs pupae specific MicroRNA-mRNA network, (C) Teneral male vs pupae specific miRNA-mRNA network, (D) Teneral female vs non-teneral female specific miRNA-mRNA network, (E) Non-gravid vs gravid female specific miRNA-mRNA network, (F) Teneral male vs non-teneral male specific miRNA-mRNA network. Green represents downregulated MicroRNA nodes, while red represents upregulated miRNA nodes and blue represents the mRNA transcripts. The lines represent the interaction between miRNA and mRNA.