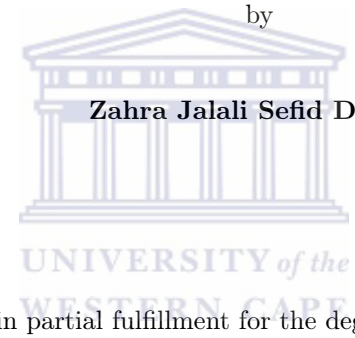
Computational characterization of *IRE*-regulated genes in *Glossina morsitans*

by

Zahra Jalali Sefid Dashti



A thesis submitted in partial fulfillment for the degree of Doctor Philosophiae at the South African National Bioinformatics Institute, Faculty of Science, University of the Western Cape

Supervisor: Professor Alan Christoffels

May 2013

Declaration

I, **Zahra Jalali Sefid Dashti**, declare that this thesis titled**, "Computational characterization of *IRE*-regulated genes in *Glossina morsitans*"** and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for the degree of **Doctor of Philosophy (PhD)** at the **University *of the* Western Cape** and has not previously in its entirely or in part been submitted to any other university.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

Signed:

_____

Date:

_____

Acknowledgments

First and foremost, praises and thanks to God, the almighty, for the wisdom and perseverance that has been bestowed upon me during this research project and throughout my life.

I would like to express my appreciation to all those without whom this dissertation would not have been possible. My deepest gratitude goes to my supervisor, Prof. Alan Christoffels for his valuable guidance, support and patience throughout this study. His enthusiasm for research, and immense knowledge has guided me throughout this project.

I would like to extend my sincere appreciation to Dr. Junaid Gamieldien for his relentless advice and encouragements. I would also like to acknowledge the support of all SANBI members.

Also my appreciations go to the National Research Foundation (NRF) for financially supporting this project.

Last but not least, I would like to thank my parents Susan and Iraj, my sister Mahjoubeh and my beloved husband Jahanshah for their unwavering love and support.

Abstract

Blood feeding is a habit exhibited by many insects. Considering the devastating impact of these insects on human health, it is important to focus research on understanding the biology behind blood-feeding, disease transmission and host-pathogen interactions. Such knowledge would pave the way for developing efficient preventative measures.

Iron an important element for species survival, is at the center of events controlling tsetse's fitness and reproductive success. Hence, targeting genes involved in iron trafficking and sequestration would present possible means of preventing disease transmission. Considering the dynamic and multi-factorial nature of iron metabolism, a well-coordinated regulatory system is expected to be at work. Despite extensive literature on the mechanism of iron regulation and key factors responsible in maintaining its homeostasis in human, less attention has been given to understand such system in insects, especially the blood-feeding insects. The availability of the genome sequences for several insect disease vectors allows for a more detailed analysis on the identification and characterization of events controlling and preventing iron-induced toxicity following a blood-meal.

The International Glossina Genome Initiative (IGGI) has coordinated the sequencing and annotation of the *Glossina morsitans* genome that has led to the identification of 12220 genes. This knowledge-base along with current understanding of the *IRE* system in regulating iron metabolism, allowed for investigating the UTRs of Glossina genes for the presence of these elements.

Using a combination of motif enrichment and *IRE*-stem loop structure prediction, an *IRE*-mediated regulation was inferred for 150 genes, among which, 72 were identified with 5'-*IRE*s and 78 with 3'-*IRE*s. Of the identified *IRE*-regulated genes, the ferritin heavy chain and *MRCK-alpha* are the only known genes to have *IRE*s, while the rest are novel genes for which putative roles in regulating iron levels in tsetse fly have been assigned in this study. Moreover, the functional inference of the identified genes further points to the enrichment of transcription and translation. Furthermore, several hypothetical proteins with no defined functions were identified to be *IRE*-regulated. These include TMP007137, TMP009128, TMP002546, TMP002921, TMP003628, TMP004581, TMP008259, TMP012389, TMP005219, TMP005827, TMP007908, TMP009332, TMP01-3384, TMP009102, TMP010544, TMP010707, TMP004292, TMP006517, TMP014030, TMP009821 and TMP003060 for which an iron-regulatory mechanism of action may be inferred.

We further report 26 *IRE*-regulated secreted proteins in Glossina, that present good candidates for further investigation pertaining to the development of novel vector control strategies.

Using the predicted data on the identified *IRE*-regulated genes and their functional classification, we derived at 29 genes with putative roles in iron trafficking, where several unknown and hypothetical proteins are included. Thus a novel role is inferred for these genes in cellular binding and transport in the context of iron metabolism. It is therefore possible that these genes may have evolved in Glossina, such that they compensate for the absence of an *IRE*- regulated mechanism for transferrin.

Additionally, we propose 14 *IRE*-regulated genes involved in immune and stress response, which may indeed play crucial roles at the host pathogen interface through their possible mechanisms of iron sequestration.

Using the subcellular localization analysis, we further categorized the putative *IRE*-

regulated genes into several subcellular localizations, where the majority of genes were found within the nucleus and the cytosol.

The detection of the conserved motifs in a set of genes, is an interesting yet sophisticated area of research, that allows for identifying either co-regulated or orthologous genes, while further providing support for the putative function of a set of genes that would otherwise remain uncharacterized. This is based on the notion that co-regulated genes are often co-expressed to carry out a specific function. As such, 14 regulatory elements were identified in the 5'- and 3'-UTRs of *IRE*-regulated genes, involved in embryonic development and reproduction, inflammation and immune response, signaling pathways and neurogenesis as well as DNA repair. This study further proposes several *IRE*-regulated genes as targets for micro-RNA regulation through identifying micro-RNA binding sites in their 3'UTRs. Using a motif clustering approach we clustered *IRE*-regulated genes based on the number of motifs they share. Significantly co-regulated genes sharing two or more motifs were determined as critical targets for future investigation.

The expression map of *IRE*-regulated genes was analyzed to better understand the events taking place from 3 hours to 15 days following a blood meal. Re-analysis of Anopheles microarray chip showed the significant expression of three cell envelope and transport genes as early response and six as late response to a blood meal, which could indeed be assigned a putative role in iron trafficking. Genes identified in this study with implications in iron metabolism, whose timely expression allows for maintaining iron homeostasis, represent good targets for future work.

Considering the important role of evolution in species adaptation to habits such as Hematophagy, it is of importance to identify evolutionary signatures associated with these changes. To distinguish between evolutionary forces that are specific to iron-metabolism in blood-feeding insects and those that are found in other insects, the *IRE*-regulated genes were clustered into orthologous groups using several blood feeding and non-blood feeding insect species. Assessment of different evolutionary scenarios using the Maximum Likelihood (ML) approach, points to variations in the evolution of *IRE*-regulated genes between

the two insect groups, whereby several genes indicate an increased mutation rate in the BF-insect group relative to their non-blood feeding insect counterparts. These include TMP003602 (phosphoinositide3-kinase), TMP009157 (ubiquitin-conjugating enzyme9), TMP010317 (general transcription factor IIH subunit1), TMP011104 (serine-pyruvate mitochondrial), TMP013137 (pentatricopeptide Transcription and translation), TMP013886 (tRNA(uridine-2-o-)-methyl-transferase-trm7) and TMP014187 (mediator 100kD).

Additionally, we have indicated the presence of positively selected sites within seven blood-feeding IRE-regulated genes namely TMP002520 (nucleoporin), TMP008942 (eukaryotic translation initiation factor 3), TMP009871(bruno-3 transcript) , TMP010317 (general transcription factor IIH subunit1), TMP010673 (ferritin heavy-chain protein), TMP011104 (serine-pyruvate mitochondrial) and TMP011448 (brain chitinase and chia).

Thus the results of this study provides an in depth understanding of iron metabolism in Glossina morsitans and confers important targets for future validations based on which innovative control strategies may be designed.

Contents

List of Figures

List of Tables

Abbreviations

ABC	Adenosine triphosphate binding cassette
AMA	Average motif affinity
ARE	AU-rich element
ART	Arrestin-related trafficking adaptor
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BBH	Bidirectional best hit
BF	Blood-feeding
Btk	Bruton's tyrosin kinase
CD	Conserved domain
cDNA	Complementary DNA
CP/Cp	Ceruloplasmin
CPE	Cytoplasmic polyadenylation element
D	Day
DCT	Divalent cation transporter
Dcytb	Duodenal cytochrome b
df	Degree of freedom
DMT	Divalent metal transporter
dN	Non-synonymous substitution
DNA	Deoxyribonucleic acid
dS	Synonymous substitution
eALAS	Erythroid-specific isoform of 5-aminolevulinate synthase
eIF3	Eukaryotic translation initiation factor
EST	Expressed sequence tags
Fer	Ferritin
FIE	Ftz instability element
FIMO	Find individual motif occurrences
FRDA	Frataxin
GAIT	Gamma interferon inhibitor of translation
GEO	Gene expression omnibus
GO	Gene ontology
GOMO	Gene ontology for motifs
GPCR	G-protein-coupled receptor
GPI	Glycosylphosphotidylinositol
H/h	Hours
HAT	Human african trypanosomiasis
HCP	Heme carrier protein
Heph	Hephaestin
HFE	Hemochromatosis
HFE3	Hemochromatosis type3
HIF2	Hypoxia-inducible transcription factor
HIRIP	Human ischemia-reperfusion-inducible protein
IGGI	International Glossina genomics initiative
IgSF	Ig-super family
IRE	Iron response elements
Ireg	Iron regulated transporter
IRP	Iron regulatory protein
ISC	Iron sulfur cluster
ISU1	Cytosolic isoform of ISC scaffold
kDa	Kilodalton
LRT	Likelihood ratio test
m-acon	Mitochondrial aconitase
mCRD	Major coding-region determinant
MEGA	Molecular evolutionary genetics analysis
MEME	Multiple EM for motif elicitation
ML	Maximum likelihood
MNS	Meiosis-specific nuclear structural
MRCK-alpha	Serine/threonine-protein kinase
MSL	Male-specific-lethal
MT	Microtubule
NBF	Non-blood-feeding
NECT	Nifurtimox-eflornithine combination therapy
NO	Nitric oxides
Nramp	Natural-resistance associated macrophage protein
OMA	Orthologous matrix
ORF	Open reading frame
PABP	Poly(A)-binding protein
PAS	Polyadenylation signal
PBM	Post blood meal
PCA	Principle component analysis
PTM	Peritrophic membrane
PWM	Position weight matrix
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SAM	Sterile alpha motif
SIRE	Search for iron-responsive elements
SIT	Sterile insect technique
SMC	Structural maintenance of chromosome
SXL	Sex-lethal
SXL-bs	Sex-lethal-binding site
Tf	Transferrin
TfR	Transferrin receptor
TS	Tumor suppressor
TSG	Tumor suppressor gene
uORF	Upstream open-reading frame
URN	Upstream of N-ras
UTR	Untranslated region
VSG	Variable surface glycoprotein
WHO	World health organization

Chapter 1

Literature Review

1.1 Hematophagy in insect disease vectors

Insect disease vectors are of immense significance to human health, with blood-feeding being the major habit exhibited by these insects. Blood-feeding have evolved several times during the course of insect evolution (Balashov, 1984), with Dipterans constituting the most diverse and abundant order (Balashov, 1984). Hematophagy is apparent in a small percentage of insect species, including the lice, fleas, mosquitoes, sand flies, black flies, tsetse flies and bugs and is suggested to be an alternative mechanism adopted by insects in the absence of sufficient nutrients. Another school of thought hold the close and prolonged association of insects and vertebrates responsible for such behavioral adaptations. The behavioral adaptation of insects to direct feeding on the host itself, would have entailed morphological and physiological adaptations. Thus, blood-feeding specialized mouthparts would appear to have been derived from the chewing mouth parts (Azar and Nel, 2012).

In the context of disease transmission, it is important to understand the physiological adaptations of hematophagous insects. To this end, efforts in understanding vector-borne disease biology have resulted in the emergence of several genome sequences. Considering the urgent need for devising new strategies in disease control, due to the insufficiency of current methods, the genomic information pertaining to insect disease vectors provide the ground for identifying novel targets for vector control strategies. Furthermore, the genomic knowledge of these insects, when compared to their non blood-feeding counterparts can provide compelling evidence of the physiological implications of blood feeding, with respect to their defense against host's immune response, ingestion and digestion of the blood meal, as well as counteracting the deleterious effects of iron overload accompanied by such feeding habit. Among the sequenced hematophagous insects, tsetse flies pose important health problems and economic losses through transmitting Trypanosomes, hence affecting both human and livestock.

Trypanosomes, *i.e.* the causative agent of African sleeping sickness (Trypanosomiasis), have become one of the biggest killers in certain parts of Africa. According to the World Health Organization (WHO) estimates, there are 300 000-500 000 cases of Human African

Trypanosomiasis (HAT) with 60 million people at risk in 37 countries (40%) of Africa (Abd-Alla *et al.*, 2008). The disease is characterized by the onset of malaise and irregular fever and if left untreated the parasite invades the central nervous system, which leads to coma and eventually death (Laxman and Beavo, 2007). Besides its destructive impact on human health, through infecting livestock, Trypanosomes affect economy and agriculture, thus further aggravate the burden of poverty in Africa.

Disease transmission by insects generally involves two host contacts, whereby the pathogen is acquired through the first contact. What follows are a series of steps through which the pathogen penetrates the vector's midgut, infect several tissues, and replicate. Infection of the salivary gland by the pathogen is considered as the step where the pathogen is transmitted to another susceptible host through probing (Day, 2005). Trypanosomes undergo divisions in the midgut as large procyclic trypomastigotes, following the entry into the insect vector's bloodstream (Figure 1.1). The parasite then penetrates the peritrophic membrane (PTM) of the gut and reach the ectoperitrophic space, then migrate to the pro-ventriculus and eventually stop dividing (Vickerman *et al.*, 1988). At this point the parasite becomes elongated mesocyclic trypomastigotes, which then migrates via the esophagus, pro-basic lumen and hypopharynx to the vector's salivary glands (Vickerman *et al.* 1988; Hao *et al.* 2001). An important stage for the transmission of trypanosome infection is the development of the metacyclic trypanosome in the insect vector. The metacyclic trypanosomes, which have gone through the process of maturation, lie within the salivary gland lumen until they get discharged in the saliva as the fly bites a mammal. Following entry into the mammalian bloodstream, the slender form (proliferative) of the trypanosome is replaced by the non-proliferative stumpy form. While, the parasite multiplies in the blood stream it can be detected in the lymph nodes, spleen and liver (Laxman and Beavo, 2007). The survival mechanism of parasite involves the expression of antigenic variable surface glycoproteins (VSGs), which are connected to the surface membrane via a glycosylphosphotidylinositol (GPI) anchor (Matthews, 2005).

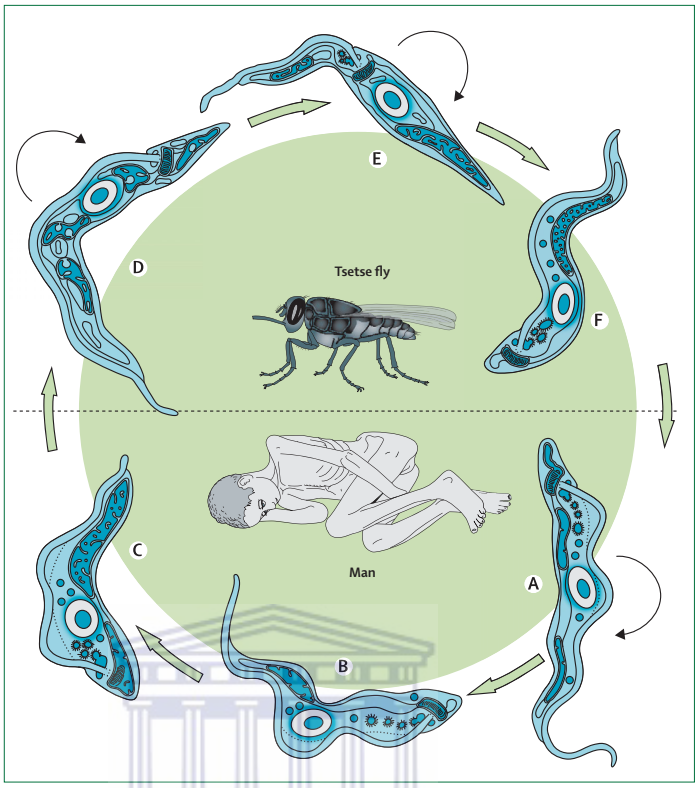


Figure 1.1: Schematic representation of the life cycle of trypanosomes whereby (A) shows the dividing slender form, (B) the intermediate forms and (C) the stumpy forms, all of which points to the cycle within the human host. Furthermore, the transition in tsetse flies occurs from (D) dividing midgut form, to (E) the migrating epimastigote form (to the salivary gland), and finally (F) the infective metacyclic form that is transferred to the human host during the next blood meal. The black arrows point to the dividing forms of trypanosomes (Adapted from Brun *et al.*, 2010).

The surface coat is a barrier that prevents the underlying plasma membrane from being exposed to non-specific immune responses (MacLeod *et al.*, 2001). The acquisition of the surface coat occurs during the differentiation from the epimastigote to metacyclic form (Caljon *et al.*, 2006). The variable antigen coat disappears from the surface as the parasite is ingested by the tsetse fly. However, as the trypanosomes proliferate and the metacyclic trypanosomes are formed in the salivary gland of tsetse fly, the parasite re-acquires the surface coat and inoculates a new mammalian host (Tetley *et al.*, 1987; Matthews, 2005).

1.2 Insect vector: Tsetse fly

Tsetse flies belong to the superfamily Hippobascoidea. There are 33 species and subspecies of tsetse flies (Glossina, *Glossinidae*), while the major vectors of human African trypanosomiasis and animal African trypanosomiasis include *G.m.morsitans*, *G.m.centralis*, *G.pallidipes*, *G.palpalis palpalis*, *G.fuscipesfuscipes* and *G.tachinoides* (Gooding and Krafsur, 2005).

1.2.1 Vector competence and immune response

Tsetse flies are generally refractory to trypanosome infection and less than half the fly population may become infected under ideal conditions (Aksoy *et al.*, 2005). On the other hand many of the infected tsetse flies fail to produce mature parasites and never become infective, hence cannot transmit the parasite. There are many factors that contribute to fly's refractoriness, including lectin levels in the gut (Hao *et al.*, 2003). Studies suggest that tsetse flies produce lectin in their midgut, which is capable of killing trypanosomes (Hao *et al.*, 2001). Symbiotic bacteria also play a vital role in vector competence. Several symbionts have been identified in tsetse fly namely: *Sodalis glossinidius*, *Wigglesworthia glossinidia* and *Wolbachia pipentis*. These symbionts have been shown to reduce vector competence of the fly, hence provide important inhibiting factors against trypanosoma infection (Gooding and Krafsur, 2005). The composition of the ingested blood may also have an effect on vector competence. The coagulation of blood within the midgut prevents the pathogens from migrating out of this environment, thus preventing its development inside the vector. Since the parasite must bypass the gut wall, survive and develop (some cases) in arthropod tissues and penetrate the salivary gland for injection into a new host, the proteolytic enzymes secreted within the midgut (blood digestion) can also have an effect on vectorial competence (Beerntsen *et al.*, 2000).

Besides immune response mechanisms protecting the fly against the invading pathogens, bloodsucking arthropods are constantly faced with two types of host defense mechanisms: (1) defense against blood loss, which includes hemostasis, platelet aggregation and vasoconstriction and (2) host's immune response triggered by inflammation. These

host responses can destroy the insect or interrupt a blood meal acquisition (Gooding and Krafsur 2005; Alves-Silva *et al.*). However, the saliva of tsetse fly contains specialized compounds that protect the insect against host responses during probing and ingestion phases of feeding (Drennan *et al.*, 2005). Salivary gland plays a pivotal role during blood meals as it contains a variety of compounds that prevent platelet and clotting function (anti-coagulant enzymes) as well as modifying inflammatory and immunological reactions in the vertebrate host (Arca *et al.*, 2005). Salivary gland is an important organ in the blood feeding insects, where the acquisition of blood, maturation and transmission of parasite takes place. Several studies including the recent work of Alves-Silva *et al.*, have focused on the characterization of the salivary transcriptome. Furthermore, previous studies on the sialome of several blood-feeding insect vectors including *A. gambiae* (Valenzuela *et al.*, 2002), *A. funestus* (Calvo *et al.*, 2006), *A. aegypti* (Valenzuela *et al.*, 2002) and *I. scapularis* (Valenzuela *et al.*, 2003) have provided a rich resource from which comparative analysis can be done to identify blood-feeding signatures. Transcriptome analysis of the salivary gland has indicated several groups of proteins in Anopheles with a role in blood feeding. These include transcription control and nuclear regulation which is associated with digesting a blood meal and converting it to eggs (*e.g.* histones, methyltransferase, importin and lamin), signal transduction (*e.g.* trypsin-like serine protease), odorant binding protein (*D7*), messages associated with DNA damage checkpoint-mitosis control, messages associated with digestive proteins (*e.g.* mucin, peritrophin I, salivary protein *SG1* family, salivary *gsg1 b* protein, trypsin precursors and *ANTRYP1* trypsin), oogenesis associated proteins (*e.g.* vitellogenin, vitellogenin receptor, apo-lipophorin-III and seprin clade B), and messages associated with immunity and melanization pathways *(e.g.* tyrosin hydroxylase, peptidoglycan recognition, lactase and ovo-peroxidase) (Christophides *et al.*, 2002; Valenzuela *et al.*, 2003; Arca *et al.*, 2005). To complement these findings Alves-Silva *et al.*, identified 250 proteins associated with blood-feeding in the sialome of Glossina. These may be classified as components of immune response (*e.g.* Alpha2 macroglobulin, pattern recognition, defense response peptides and stomoxyn antimicrobial family), mucins, antigen 5 family, yellow protein family, enzymes (*e.g.* endonucleases, serine proteases, esterases and lipases), protease inhibitor domain (*e.g.* anti thrombin peptide, *kunitz* domain of serine protease inhibitor and serpins) as well as several

putatively secreted proteins, which include proteins of unknown families. Additionally, the sialome of Glossina have revealed a number of new multigenic polypeptide families unique to this insect. The richness of data available for Anopheles and the information on the Glossina sialome, assist in identifying many new gene products that may be associated with the fundamental aspects of Glossina biology. The integration of such knowledge with evidences from other insect vectors, provide interesting clues for the development of better interventions and saving lives of those affected by these ravaging diseases.

1.2.2 Tsetse reproductive biology

The reproductive system of tsetse fly is different from other insects as they reproduce by adenotrophic viviparity, where the female carries and nourishes her offspring for the entire developmental period. Furthermore, female tsetse flies can only develop a single larva at a time, which correlates with their low reproductive capacity. This characteristic provides a good target for vector control strategies (Attardo *et al.*, 2010). The life cycle of tsetse begins with the larval stages, whereby the development of a single fertilized egg retained in the uterus is directed such that in the third larval stages the tsetse larva leaves the uterus and begins its independent life cycle as a pupae. During the pupal stage the final morphological transformations into an adult fly take place. Within a month after pupation, the fully developed adult emerges. The energy source for all the stages of tsetse life cycle are provided by the female parent through feeding on blood meal to provide the developing larvae with adequate nutritional stores before the adult fly emerges (Attardo *et al.*, 2006).

A key component of tsetse's viviparous reproductive biology is the accessory gland, also referred to as the milk gland (Gooding and Krafsur, 2005; Guz *et al.*, 2007). The accessory gland comprises of tubules that are interwoven throughout the fat body. The primary function of the accessory gland is to provide the developing larva with enough nutrients (milk). Fats (˜50%) and proteins (˜50%) are considered as major constituents of the milk. Unlike in tsetse, the Anopheles accessory gland secretions are transferred from male to female reproductive tract to trigger several behavioral changes. These may include refractoriness to mating and the propensity for egg laying (Dottorini *et al.*, 2012).

With regards to understanding the biology behind reproductive behavior of tsetse flies, several lines of work have been carried out including the transcriptome analysis of the reproductive tissues (Attardo *et al.*, 2010), and the expression analysis of reproductive-associated genes and their products synthesized during the first and second gonotrophic cycles (Attardo *et al.*, 2006). The results of such studies have provided important insight into the function of several reproductive genes, such as yolk proteins, gonadal proteases, cuticular proteins and several others (Attardo *et al.*, 2010). Exploring these genes and further research may reveal potential means to control vector population, hence disease prevention.

1.2.3 Mechanisms of iron metabolism

Hematophagy or blood feeding is a behavior exhibited by tsetse fly (Vickerman *et al.*, 1988; Attardo *et al.*, 2006). The blood meal is broken down into amino acids, carbohydrates and lipids, which are processed by the fat body into proteins that are required by the growing oocyte within the ovaries. It has been shown that the entire blood meal is converted to over 100 eggs within two days (Ribeiro, 2003). Extensive research has been carried out to understand the effect of blood meal on the expression of genes in insects. One such study is the work of Ribeiro (2003), who assessed the expression of Anopheles transcripts in response to blood meal. Approximately 435 protein sequences were represented as messages that were increased or decreased following blood meal. These include the components of transcription control, digestion, protein synthesis machinery, oogenesis and immunity. Considering that mosquitoes are faced with increased metabolic needs to digest the blood meal and direct it towards egg development, observing up-regulated levels of transcription control associated proteins is expected.

The host's blood meal is a rich source of iron for tsetse's developmental requirements, however excess iron can be detrimental (Attardo *et al.*, 2006). As such, successful reproduction, symbiotic fitness and parasite survival entails a strict control of iron metabolism within the blood-feeding insect (Guz *et al.*, 2007). Iron is an important component of cuticle formation, tanning, melanization and healing (Locke and Nichol, 1992), while its role in respiration, energy metabolism, oxygen transport, gene regulation, DNA biosyn-

thesis and immunity is well-established (Dunkov and Georgieva, 2006; Ong *et al.*, 2006; Muñoz *et al.*, 2009). Despite its crucial role in species survival, iron-induced reactive oxygen species generated through the Fenton reaction can have deleterious impacts on the insect survival (Equation 1.1) (Weinberg and Miklossy, 2008; Hower *et al.*, 2009; Kell, 2009). This is driven by the reduction of iron by superoxides, which is a favored process under mitochondrial oxidative stress (Equation 1.2). Under normal physiological conditions iron Fe^{2+} is oxidized to Fe^{3+} , which results in the formation of reactive oxygen species. The resultant HO can therefore cause significant biological damage to proteins, lipids and DNA (Thomas *et al.*, 2009). However, to prevent such damage species have set mechanisms in place that allow iron uptake and usage, while avoiding iron toxicity.

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^{\cdot} + OH^{-} \tag{1.1}$$

$$Fe^{3+} + H_2O_2 \rightarrow Fe^{2+} + OOH^{\cdot} + H^{+} \tag{1.2}$$

Mammalian blood contains iron in two forms: heme and non-heme. Hence, through a blood meal, hematophagous insects acquire a high load of iron in the form of hemoglobin and ferric-transferrin. The iron from these sources is absorbed and transported to the ovaries for the process of oogenesis (Winzerling and Pham, 2006). Assessment of the fate of blood meal iron in *A. aegypti*, Zhou *et al.* (2007) have shown that, approximately 87% of heme-derived iron is excreted, while 13% is distributed to eggs (7%) and other organs (6%). Moreover it has been shown that only 8% of the transferrin-bound iron is excreted, while 71% is transferred to eggs and 15% distributed to other tissues (Figure 1.2).

Despite extensive literature on the mechanism of iron regulation in human, less is known about insects. Human iron metabolism requires the function of several genes, the majority of which have no known homologs in insects. These include iron regulatory proteins, transferrin receptors, transferrin, ferritin, divalent metal transporter (*DCT1*), ferroportin, hephaestin, ceruloplasmin, iron sulfur cluster proteins (*ISC*), frataxin, ferrochelatase, adenosine triphosphate (*ATP*)-binding cassette (*ABC*) and several others.

Ferroportin, a basolateral membrane protein, involved in human iron uptake and transfer (McKie *et al.*, 2000), has no known homolog in many insects. Studies have suggested that ferroportin mediates iron efflux across membranes by ferroxidase enzyme activity (McKie and Barlow, 2004). Furthermore, McKie *et al.* (2000) have proposed that ferroportin functions in accordance with hephaestin (membrane-associated ferroxidase) and serum ceruloplasmin.

Ceruloplasmin (*CP*) and hephaestin (*Heph*), a homologue of *CP*, are multi-copper oxidases that catalyze the Fe^{2+} oxidation to Fe^{3+} (Winzerling and Pham, 2006). Ceruloplasmin is a plasma protein synthesized by liver cells and is widely distributed throughout the body. It plays a role in the efflux of iron from cells, however it has not been shown to play a role in iron efflux from small intestine. Recent studies have shown that hephaestin is involved in iron transport from the intestine. This further emphasizes the correlation between iron and copper metabolism. It has also been shown that like ceruloplasmin, hephaestin possess a ferroxidase activity (Petrak and Vyoral, 2005). Unlike ceruloplasmin, hephaestin is expressed in a limited number of tissues, however it is highly expressed in the intestine, which accentuates its role in intestinal iron transport (Anderson *et al.*, 2002; Jeong and David, 2003).

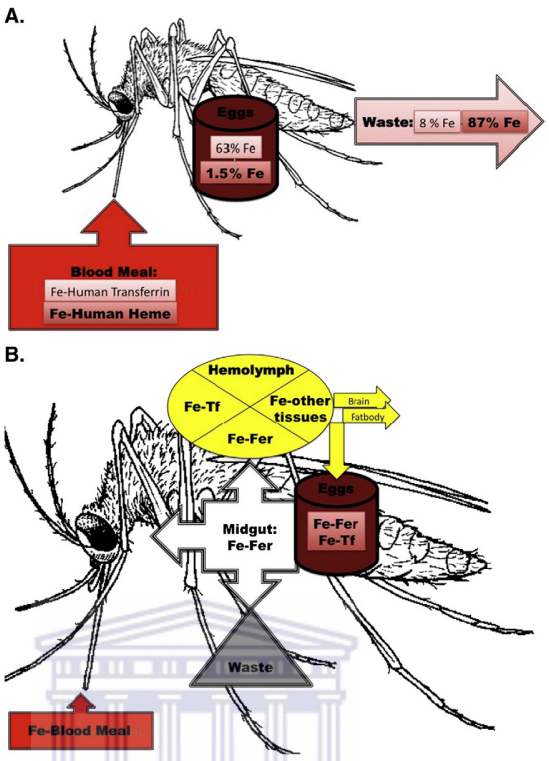


Figure 1.2: Fate of blood meal iron. A) Shows the main path for iron absorption. Fe-human Transferrin refers to transferrin-bound iron from the human blood meal (not bold), while Fe-Human Heme, refers to heme-derived iron from hemoglobin in the human blood meal (bold). B) Shows the proteins known to transfer iron to different tissues in the mosquito. Accordingly Fer, represent insect ferritin and *Tf*, the insect transferrin. (Adapted from Pham and Winzerling, 2010)

Adenosine triphosphate (*ATP*)-binding cassette (*ABC*) is another member of the human iron metabolism network that has no significant homolog in the majority of insect species. In human it is located in the mitochondria and regulates iron transport and iron-containing proteins (Napier *et al.*, 2005). Mitochondria exports both heme and Fe through specific transporters. Heme transporters are unknown, however [Fe-S] clusters are exported to the cytosol through the mitochondrial inner membrane transporter, *ABCB7*. Studies in yeast have shown that *ABC* transporter of mitochondrion1 protein (*Atm1p*) a member of Adenosine triphosphate (*ATP*)-binding cassette is responsible for [Fe-S] cluster transport. Studies suggest that mutations in the human *ABCB7* gene result in mitochondrial Fe accumulation and the formation of ring sideroblast, pointing to the important

role of *ABC* in maintaining iron homeostasis.

Though such knowledge in the context of insect iron metabolism is at its infancy, the emergence of genome sequences for a number of disease-vectors provide the ground for addressing this topic in more depth. To date all experimental studies in hematophagous insect's iron metabolism have addressed iron-regulated proteins including ferritin (Pham and Chavez, 2005; Dunkov *et al.*, 2002), transferrin (Harizanova *et al.*, 2005; Yoshiga *et al.*, 1997), iron regulatory protein 1 (*IRP1*) (Zhang *et al.*, 2002) as well as the divalent metal transporter 1 (*DMT1*) (MartŠnez-Barnetche *et al.*, 2007). In light of current literature the role of these genes in regulating insect's iron levels will be addressed.

1.2.3.1 Iron regulatory proteins

Iron regulatory proteins (*IRP*s) are central regulators of iron homeostasis. Proteins involved in iron metabolism are regulated post-transcriptionally by the interaction of iron response elements (*IRE*s) and *IRP1*. *IRE*s are RNA stem-loop structures, located in the untranslated regions (UTRs) of mRNAs involved in cellular iron homeostasis, that were first identified in the 5'-UTR of human ferritin H-chain messenger RNA (Hentze *et al.*, 1987). The regulatory effect of *IRE* takes place through the binding of *IRP*s to *IRE*s in the 5'- and 3'-UTRs, which results in the expression initiation or suppression of the associated genes (Yamanaka *et al.*, 1999). Under low cellular iron levels this regulatory process takes place through the binding of *IRP* to *IRE*s, which consequently prevents the assembly of the ribosomal apparatus. Hence, the expression of genes such as ferritin is blocked (Eisenstein, 2000; Pantopoulos, 2004; Muckenthaler *et al.*, 1998). Alternatively, when cellular iron levels increase, the *IRP-IRE* interaction declines and synthesis increases. Furthermore, increased intracellular iron levels result in the conversion of *IRP1* from an RNA binding protein to an Fe-S cluster containing cytoplasmic aconitase, which is no longer capable of binding *IRE*s. Mutations that block the [4Fe-4S] cluster insertion at the three cysteine residues, promote RNA binding activities of *IRP*s, hence make it active in *IRE* binding (Hentze *et al.*, 2004). Such alterations between the enzyme and the iron binding functions of *IRP* occur through an "iron-sulfur switch" mechanism, controlled by the extracellular frataxin (Condo *et al.*, 2010). Frataxin is conserved among species

and studies have shown conservation of the B1-B2 strands as well as B6-B7 residues in Anopheles. Frataxin is localized in the mitochondria and binds Fe^{2+} , hence promotes cluster formation and prevents cluster dis-assembly. The binding of iron to frataxin results in the reduction and lowering of iron availability due to protein aggregation (Winzerling and Pham, 2006). When cells are depleted of iron the [4Fe-4S] cluster disassembles and *IPR1* binds *IRE*s with high affinity (Kuriyama-Matsumura *et al.*, 2001). Molecular characterization of *IRP1* in several insects including *A. aegypti* and *A. gambiae* have shown substantial similarities to the human *IRP1* and Drosophila's *IRP1*A and *IRP1*B. The transcription of *IRP1* in mosquitoes is fairly constant throughout various developmental stages. However, an immune response have been shown to be implicated in increased *IRP-IRE* binding activities, which may further point to the role of this system in insect immune response (Zhang *et al.*, 2011).

1.2.3.2 Insect transferrin and ferritin

One of the mechanisms of preventing the deleterious effects of iron is through its transport and storage. One such regulator is the transferrin (serum glycoprotein), which has been found in abundance in the hemolymph (Guz *et al.*, 2007). Based on the mammalian system of iron metabolism, iron uptake from transferrin involves the binding of transferrin (*Tf)* to the transferrin receptor (*TfR*). However, the lack of iron-delivery to insect organs through the receptor-mediated action of *Tf,* poses the question about the possible conservation of the role of this gene in insects as transporters of iron. This however, could suggest the presence of a different form of transferrin receptor or that insects have adopted different means of transporting iron (Nichol and Winzerling, 2002). Furthermore, despite the common perception of the role of transferrin in iron transport, other functions such as immune response and vitellogenesis have been attributed to this gene (Guz *et al.*, 2007; Yoshiga *et al.*, 1997).

Comparative assessment of transferrin across several insect species including *A. aegypti*, *C. quinuefasciatus*, *G. morsitans*, *D. silvestrisi*, *A. mellifera* and several others have indicated that most insect transferrins have both the N and the C lobes as seen in the mammalian transferrin. However, some are only capable of binding iron in the N-terminal

region (Geiser and Winzerling, 2012). One such example is the tsetse fly's transferrin that has lost its iron regulatory abilities on the C-terminal lobe along with the loss of large sequence blocks. However, transferrin has retained three of the five residues necessary for iron-binding in the N-terminal lobe (Strickler-Dinglasan *et al.*, 2006). It may therefore be extrapolated that in spite of transporting iron, insect transferrin is not the predominant protein in performing iron transportation in these insects. In line with this, insect ferritin has been suggested as an alternative factor (Aisen *et al.*, 2001; Liu and Theil, 2005; Nichol and Winzerling, 2002). Ferritin is a polymer involved in detoxification and iron storage. It can store proximately 4 500 atoms of iron as ferrihydrite in a soluble and non-reactive form in the cell. The heavy chain subunits of ferritin are characterized by the presence of seven conserved amino acid residues that form a ferroxidase center, required for the oxidation of ferrous iron to ferric iron. The light chain subunit on the other hand, does not constitute any of these residues; instead it exposes a basic side chain to the cavity that allows for the crystallization and stabilization of the iron core. Studies in *Drosophila melanogaster* have indicated that the ratio between the heavy and light chain ferritin is fixed and that the over-expression of either subunits does not affect its protein levels. Considering that ferritin heavy chain is post-transcriptionally regulated by an *IRP-IRE* system, it has been speculated that such regulation is not necessary for the light chain, as regulation of the heavy chain will lead to the co-regulation of the light chain ferritin (Missirlis *et al.*, 2007). To contradict this scenario both the heavy and light chain ferritin of Lepidoptera are post-transcriptionally regulated by *IRE*s, which may be indicative of other reasons for the lack of *IRE* in the light chain subunit of Dipteran's ferritin.

Based on the mammalian models, the perception of ferritin is mainly as a cytoplasmic iron storage protein, however, insects have evolved such that ferritin is utilized as both an iron storage and an iron transporter (Nichol and Winzerling, 2002). Even though vertebrate's ferritin mainly lacks signal peptides, insects ferritin including that of the tsetse fly contains secreted signals. Secreted insect ferritin has been found in high concentrations in the hemolymph while cytoplasmic ferritin is significantly lower (Ong *et al.*, 2006; Yoshiga *et al.*, 1997). Insect's ferritin is generally expressed in the midgut and the fat body to prevent the toxic effects of excess iron by binding to it. To compensate for the high levels

of iron, the midgut excretes ferritin into the gut lumen and eventually excretes it in the feces. The midgut has been shown to secrete iron rich ferritin, whereas the fat body secretes iron poor ferritin (Attardo *et al.*, 2006). Furthermore, studies have illustrated that ferritin is expressed throughout all developmental stages of mosquito's life, though in response to blood meal the expression increases in the ovaries, gut and hemolymph (Dunkov *et al.*, 2002). As such ferritin heavy chain has been suggested as a major defense mechanism against intracellular iron overload (Geiser *et al.*, 2006).

1.2.3.3 Natural Resistance Associated Macrophage Protein2 (*NRAMP2*)

In human, natural resistance associated macrophage protein2 (*Nramp2*), previously named divalent metal transporter (*DMT1*) is responsible for the intestinal uptake of non-heme iron (McKie and Barlow, 2004). *Nramp* protein families are involved in the transportation of Fe, Mn, Zn, Co and Ni (Gunshin *et al.*, 1997). Genes sharing similarities to *Nramp*2 have been identified in insects (*e.g.* Drosophila's *malvolio*, Anopheles' *AnaNramp* and in *C.elegans*) (Rodrigues *et al.*, 1995; MartŠnez-Barnetche *et al.*, 2007; Bandyopadhyay *et al.*, 2009). Assessment of *malvolio* in *Drosophila melanogaster* have suggested that mutations in this gene can induce abnormal taste behavior, which was suppressed by metal supplementation. As in vertebrates, a role in immune response has been suggested for *malvolio* (Rodrigues *et al.*, 1995). Furthermore, this gene is expressed in the nervous system, macrophages and malpighian tubules of Drosophila (Rodrigues *et al.*, 1995; Folwell *et al.*, 2006). Also, *AnaNramp* is down-regulated in the midgut and up-regulated in the malpighian tubules (MartŠnez-Barnetche *et al.*, 2007). Furthermore, the expression of *Nramp* is regulated through the *IRP-IRE* system in vertebrates, however such regulatory mechanism has not yet been validated for insects homologs of this protein (Petrak and Vyoral, 2005).

In conclusion, iron homeostasis in insects, significantly relies on the *IRP-IRE* system, while transcriptional regulation is also of importance in maintaining a balanced iron level. Furthermore, in spite of its importance in survival, our knowledge of iron metabolism in insects is still obscure, while the exact mechanism of action and factors at play in determining the fate of blood meal in hematophagous insects remain elusive. Hence,

identifying genes specific to the iron regulatory and metabolism pathways could reveal unique aspects of insect biology and provide clues to guide future research, hence lead to developing innovative means of control mechanisms.

1.3 Research rational

Iron is a double-edged sword in that, while being indispensable for the life of organisms it can be a potent toxin (Guz *et al.*, 2007). Iron can form a variety of complexes with organic ligands, which directly contributes to its biological function. Since the last comprehensive review of iron metabolism in insects by Nichol and Winzerling (2002), a number of insect genomes have been successfully sequenced, which have provided a more complete perspective to better study iron uptake and regulation in insects.

Most species, specifically blood-feeding insects are constantly faced with the toxic effects of iron from blood meal that causes intense damage to various biological systems through the production of reactive oxygen species. However, it is evident that an intricate system is in place, which involves the function and interaction of various molecules and enzymes. A more comprehensive look at insect pathways compared to that of other vertebrates could give a clearer view of how insects have evolved to meet their blood feeding needs by looking at the comparative aspects of iron metabolism and regulation. Iron-uptake, transporters, receptors and regulators are implicated in pathogen virulence and immunogenicity making them favorable drug targets (MartŠnez-Barnetche *et al.*, 2007; Guz *et al.*, 2007; Geiser and Winzerling, 2012; Strickler-Dinglasan *et al.*, 2006). Furthermore, iron plays an integral role in eggshell formation in these insects and is required for optimal egg development, thus altering iron levels can be directed towards reducing the reproductive health of the insect. Although the amount of data on iron assimilation is growing, there is still a significant gap in our knowledge, hence iron uptake, metabolism and regulation in blood feeding insects presents novel areas for research, by providing numerous putative targets for innovative therapies (Glanfield *et al.*, 2007).

As part of the International Glossina Genomics Initiative (IGGI), the genome of *Glossina*

morsitans was sequenced by the Sanger Institute and annotated using the MAKER pipeline. Furthermore, an extensive set of expressed sequence tags (ESTs) and bacterial artificial chromosome (BACs) libraries have been generated from tissue-specific normalized libraries, which along with NGS illumina reads were aligned to the genome as part of the pipeline to assist in the description of gene models. These include ESTs from 11 cDNA libraries, namely head, midgut, reproductive, salivary gland, larvae, pupae, fat body and whole body. In summary, the gene models were generated using a sequence similarity approach and gene annotations are extrapolated from the annotation of the matching homologs. In the context of identifying iron regulatory proteins, there are other features such as *IRE* motifs that provide additional evidence for the location of iron regulatory genes. As an important vector of human and animal disease, the complete genome sequence of Glossina will foster identification and development of novel means of controlling the devastating disease of sleeping sickness in Africa. As iron plays a vital role in the oocyte development and vitellogenesis as well as parasite development, targeting specific proteins involved in iron regulatory pathway in tsetse fly can lead to developing innovative means of controlling Tsetse fly population from transmitting the trypanosoma infection.

1.4 Aims and Objectives

Iron plays an important role at the host-pathogen interface, while affecting the vector's reproductive success and survival. Hence, identifying genes associated with iron regulation in blood-feeding insects contributes towards developing innovative preventative measures for the vector-derived transmission of disease. Iron-metabolism is well-characterized in human, however more research needs to focus on the identification and characterization of it's components in insects and more specifically blood-feeding insects. To this end the major aims of this study can be summarized as follows:

- Expand the collection of iron-responsive genes through the identification of *IRE*-stem loop structures.
- Functional characterization of *IRE*-regulated genes in *Glossina morsitans*.
- Computational elucidation of co-regulation governing *IRE*-regulated genes by identifying other regulatory elements in their UTRs.
- Determine the transcriptional responses of *IRE*-regulated genes in the presence of iron using microarray data available for *Anopheles gambiae*.
- Identify evolutionary mechanisms governing *IRE*-regulated genes in blood-feeding insects.

Chapter 2

Identification of *IRE*-regulated genes

ABSTRACT

Background:

Iron metabolism and regulation is an indispensable part of species survival specifically the blood feeding insects. Iron response proteins are central regulators of iron homeostasis, whose binding to iron response element (*IRE*) stem loop structures within the UTRs of genes regulate expression at the post-transcriptional level. Despite extensive literature on the IRE-regulated genes in human less attention has been given to insect and more specifically the blood feeding insects, where research has focused mainly on the characterization of ferritin and transferrin. To this end, a more focused approach is necessary to give insight into the *IRE*/*IRP* mechanism of regulation in blood feeding insects. Therefore, this study examined the mechanism of iron homeostasis in blood feeding insects through the genome-wide computational identification and characterization of motifs enriched in addition to *IRE*s in the UTRs of *Glossina morsitans* with the view of identifying new *IRE*-regulated genes.

Methods:

Glossina UTRs were screened for patterns of canonical, non-canonical and several SELEX-based *IRE* patterns using the RSAT algorithm. Putative *IRE* patterns were filtered using *IRE* stem loop structures and the resulting *IRE*-regulated genes were functionally categorized. MEME and

UTRScan was used to identified additional motifs enriched in the UTRs of the *IRE*-regulated genes. A motif clustering approach was implemented to shed light on co-regulation among the *IRE*-regulated genes.

Results:

A total of 150 *IRE*-regulated genes were identified in this study, of which two are known to contain *IRE*s, namely the ferritin heavy chain and *MRCK-alpha.* The remainder of the identified *IRE*-regulated genes are considered novel containing 11 hypothetical proteins, for which a putative role was assigned in this study namely regulating iron levels in tsetse fly. The findings of this work gives insight into mechanisms by which tsetse flies overcome the toxic effects of iron from the blood meal through the interplay of genes involved in several biological aspects of survival such as transport and binding, metabolism, transcription and translation, and immune response. In the context of iron-sequestration, key players of tsetse's immune defense against trypanosomes have been introduced namely 14 stress and immune response genes, which provides a ground for future research aiming at identifying alternative approaches to vector control, hence disease prevention.

Iron trafficking is an important mechanism in maintaining a balanced iron level. The regulated function of genes involved in iron trafficking, have greater implications in the survival of blood-feeding insects, due to the enriched iron content of the ingested blood meal. Thus, 28 cell-envelope, transport and binding genes identified in this study may have profound effects in ensuring insect's survival by regulating iron levels. The motif discovery approach led to the identification of a set of 5'- and 3'-elements enriched in the UTRs of the *IRE*-regulated genes. Among these motifs were three microRNA-binding sites namely *Gy-box*, *Brd-box* and *K-box* motifs.

The clustering of genes based on their motif sharing points to a tight co-regulation between genes involved in transcription and translation, where phosphoinositid 3-kinase, *spire*, *MCM7*, mediator, integrator4 and 40S-ribosomal protein were found to have more than three motifs shared (co-regulation) with other *IRE*-regulated genes.

Conclusion:

In conclusion, the significance of the identified *IRE*-regulated genes may be highlighted as alternative mechanisms by which iron regulation is maintained in tsetse flies. Besides the new perspective on iron metabolism in blood-feeding insects, the findings of this study further provides

putative candidates for future development of vector control strategies by targeting iron homeostasis in tsetse fly.

2.1 Introduction

2.1.1 Iron metabolism and regulation

Gene regulation occurs at several levels including transcriptional and translational regulation, which involves the binding of the regulatory machinery to the regulatory elements within the promoter and the UTRs respectively. Thus, understanding how complex regulatory elements respond to signals and functionally integrate to regulate the extent and the timing of translation of iron responsive genes, is of great importance. Furthermore, targeting mRNA translation is emerging as a mechanism to control gene expression and represents a promising area for the development of control interventions targeting insect disease vectors. Several factors affect the translation efficiency of a gene including the length of the 5'-UTR, start site consensus sequences, upstream AUGs, upstream open reading frames (*uORF*s), internal ribosomal entry sites (*IRES*) along with sequences that function as binding sites for regulatory proteins. Additionally, the 3'-UTR contains various binding sites for regulatory factors as well, that are often proteins, while several trans-acting RNAs have also been reported (Gray and Wickens, 1998; Mignone *et al.*, 2002). Studies have identified a number of motifs present in the 3'-UTR whose roles are crucial in the regulation of mRNAs, namely the AU-rich elements (*ARE*s), cytoplasmic polyadenylation element (*CPE*), *K-box* and *PAS*. UTR sequences involved in regulation can be classified into short sequence motifs that function as binding sites for RNA binding proteins as well as non-coding RNAs. Repetitive elements such as CUG repeats have also been documented to function as targets for RNA binding proteins (Yoon *et al.*, 2008).

In the context of iron metabolism, despite existing literature on the regulatory mechanisms of iron responsive elements (*IRE*s), no in depth analysis has been carried out to identify other elements in the UTRs of these genes. The knowledge of UTR elements may provide clues about the regulatory machinery controlling iron metabolism through their possible synergistic effects with *IRE*s. Such knowledge is even more sparse with

respect to insects and more specifically the blood-feeding insects. Cellular iron homeostasis is post-transcriptionally controlled with iron regulatory proteins (*IRP*s) as its key regulators. This is achieved through their binding to *IRE*s in the untranslated regions of mRNA encoding proteins. Under iron deprivation, *IRP*s bind to *IRE*s and block 43S ribosomal recruitment, hence preventing mRNA translation (Muckenthaler *et al.*, 2008). To date six mammalian mRNAs have been identified to be *IRE*-regulated through the binding of *IRP* to their 5' *IRE*s, namely the erythroid-specific isoform of 5-aminolevulinate synthase (*eALAS*), mitochondrial aconitase (*m-acon*), ferroportin, the hypoxia-inducible transcription factor (*HIF2a*), as well as the H and L subunits of ferritin (Leipuviene and Theil, 2007; Muckenthaler *et al.*, 2008).

Both *IRP1* and *IRP2* are involved in iron homeostasis, however their modes of action differ (Ke *et al.*, 2000). *IRP1* is a stable protein and has been suggested to play a bifunctional role in iron regulation. The activities of *IRP1* are controlled through the iron-responsive formation of a Fe-S cluster, transforming its activities from an *IRE*-binding protein to the cytoplasmic isoform of aconitase (*c-acon*). Studies have suggested extensive structural rearrangements of the four *IRP1* domains upon loss of the Fe-S clusters (Williamson, 2000; Dupuy *et al.*, 2006; Walden *et al.*, 2006). Like *IRP1*, *IRP2* functions similarly, though in the presence of iron it rapidly degrades (Yamanaka *et al.*, 1999). Furthermore, studies on murine pro-B cell lines, Ba/F3, have shown that *TfR* and ferritin are fully regulated by *IRP2* as these cells lack *IRP1*. This suggest that even though both *IRP1* and *IRP2* are involved in iron homeostasis the presence of one alone may be sufficient for *IRE*-mediated iron regulation (Kuriyama-Matsumura *et al.*, 2001).

Novel *IRE*s involved in diverse cellular pathways have been identified that is suggestive of the extended network of *IRP-IRE* to processes beyond iron homeostasis (Anderson *et al.*, 2012) (Table. 2.1). Initial reviews addressing the topic of *IRE* regulation were focused on two *IRE*-regulated genes *i.e.* ferritin and *TfR* known at the time (Theil, 1990). Since then many mRNAs have been identified to be *IRE*-regulated, while many more remain to be characterized. As a result, great amount of details are now known about the *IRE* tertiary structure and the multiple signaling pathways known to converge on the

IRP/IRE interactions (Rouault and Klausner, 1996; Hentze and kuhn, 1996; Eisenstein, 2000; Theil, 1998; Ke *et al.*, 1998, 2000; Fleming *et al.*, 1997; Gunshin *et al.*, 1997). Most of the known *IRE*s to date have initially been identified in mammalian mRNAs. The first functional *IRE* for insects on the other hand was identified in the 5'-UTR of Drosophila's *SDHB* mRNA (Kohler *et al.*, 1995), while the *IRE*-regulation of this gene in human and other mammals is not evident. Piccinelli and Samuelsson (2007) through phylogenetic analysis of *IRE*s have suggested *FTH1/FTL IRE*s to occur in the majority of metazoans, while *IRE* of *DMT1* (*SLC11A2*) is confined to mammals. Such differing evidence on the type and nature of *IRE*-regulated genes necessitates a more focused analysis of these regulators in species other than mammals.

Table 2.1: A list of known iron metabolism genes, across various species, that have *IRE*s present in their 5'- and 3'-UTRs (Adapted from Stevens *et al.*, 2011; Balashov, 1984).

Gene	Species	Alternative name	Name	Location	Ref	Date
FTH1	*Homo sapiens*	*FHC; FTH; PLIF; FTHL6; PIG15; MGC104426; FTH1*	ferritin, heavy polypeptide 1	5' UTR	1	1987
FTL	*Homo sapiens*	*NBIA3; MGC71996; FTL*	ferritin, light polypeptide	5' UTR	25	1987
TFRC	*Homo sapiens*	*TFR; CD71; TFR1; TRFR; TFRC*	transferrin receptor (p90, CD71)	3' UTR	3	1989
ALAS2	*Homo sapiens*	*ASB; ANH1; XLSA; ALASE; XLDPP; ALAS-E; FLJ93603; ALAS2*	aminolevulinate, delta-, synthase 2	5' UTR	26	1991
SdhB	*Drosophila melanogaster*	*CG3283; Dmel\CG3283; Ip; SDH; SDH-Ip; SDH-IP; sdhB; SDHb*	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	5' UTR	14	1995
ACO2	*Bos taurus*		aconitase 2, mitochondrial	5' UTR	27	1996
Ferritin	*Pacifastacus leniusculus*			5' UTR	28	1999
Hao1	*Mus musculus*	*GOX; Gox1; Hao-1; MGC141211; Hao1*	Hao1 hydroxyacid oxidase 1, liver	3' UTR	29	1999
qoxD	*Bacillus subtilis*		cytochrome aa 3-600 quinol oxidase (subunit IV)	3' UTR	30	1999
SLC11A2	*Homo sapiens*	*DCT1; DMT1; NRAMP2; FLJ37416; SLC11A2*	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	3' UTR	11	2001
Ferritin	*Manduca sexta*		*Manduca sexta* ferritin heavy chain-like protein precursor	5' UTR	31	2001
NDUFS1	*Homo sapiens*	*CI-75k; CI-75Kd; PRO1304; MGC26839; NDUFS1*	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75 kDa (NADH-coenzyme Q reductase)	5' UTR	32	2001
Ferritin	*Calpodes ethlius*		Calpodes ethlius fat body secreted ferritin S subunit precursor.	5' UTR	33	2002
Slc40a1	*Mus musculus*	*MTP; Ol5; Pcm; Dusg; Fpn1; MTP1; IREG1; Slc11a3; Slc39a1; Slc40a1*	solute carrier family 40 (iron-regulated transporter), member 1	5' UTR	34	2003
alas2	*Danio rerio*	*sau; alas-e; cb1063; sauternes; alas2*	aminolevulinate, delta-, synthase 2	5' UTR	35	2005
CDC42BPA	*Homo sapiens*	*MRCK; MRCKA; PK428; FLJ23347; KIAA0451; DKFZp686L1738; DKFZp686P1738; CDC42BPA*	CDC42 binding protein kinase alpha (DMPK-like)	3' UTR	12	2006
CDC14A	*Homo sapiens*	*cdc14; hCDC14; CDC14A*	CDC14 cell division cycle 14 homolog A (*S. cerevisiae*)	3' UTR	13	2006
EPAS1	*Homo sapiens*	*HLF; MOP2; ECYT4; HIF2A; PASD2; bHLHe73; EPAS1*	endothelial PAS domain protein 1	5' UTR	15	2007

Some of the important structural features of *IRE* that allow for its recognition by *IRP* are, the presence of a terminal loop of sequences CAGUGH (H = U, C or A), a downstream stem of five base pairs that form an alpha-helix followed by a midstream C-bulge (C8) created by the conserved G-C bps, as well as sufficient base pairings before C8 to allow for *IRE* stability (Addess *et al.*, 1997; Dandekar *et al.*, 1991; Leipuviene and Theil, 2007; Volz,

2008). Additionally, the C14 and G18 nucleotides of the terminal loop, pair and form a pseudo-triloop (AGU) (Addess *et al.*, 1997; Henderson *et al.*, 1994; Sierzputowska-Gracz and Theil, 1995; Walden *et al.*, 2006). Several factors affect *IRE-IRP* binding affinity such as increasing the length of the upper stem, the size of the terminal loop, altering the identity of the conserved unpaired loop nucleotides and disrupting the secondary structure of *IRE* (Barton *et al.*, 1990; Leibold *et al.*, 1990; Bettany *et al.*, 1992; Kikinis *et al.*, 1995). A number of *IRE-IRP* interactions have been identified through which *IRP* recognizes and binds to a non-consensus bulged nucleotide on the 3' side of the *IRE* stem loop structures such as those in *DMT1* and *HIF2a*. Additionally, several SELEX-based studies have suggested *IRP* interactions with RNA stem loops that do not resemble the canonical forms of *IRE* sequences or structures (Henderson *et al.*, 1994; Butt *et al.*, 1996; Meehan and Connell, 2001). Hence, taking into account the known canonical and non-canonical structures associated with *IRE*s as well as the variations observed in the secondary structures of several newly identified ones, patterns defining *IRE*s remain elusive, while future studies may reveal a wider rang of *IRE* stem loop structures.

The knowledge of iron regulation at the post-transcriptional level remains poorly understood, while identification of new *IRE*-regulated genes provide the ground for more extensive work. Relating these findings to aspects of blood feeding behavior allows for better understanding the evolution of these insects and how through implementing regulatory mechanisms, they have compensate for iron overload and the consequent toxicity. To this end the focus of this chapter is to interrogate the UTR regions of the newly annotated *Glossina morsitans* genes for *IRE* signals to derive at signatures of regulation. This is achieved through the identification of *IRE* stem loop structures and further characterization of other motifs elements co-existing and co-regulating the putative iron metabolism genes. Such knowledge would pave the way for further experimental validation of iron metabolism genes in Glossina, while representing putative targets for future development of control strategies.

2.2 Materials and Methods

2.2.1 *IRE* prediction

2.2.1.1 Data retrieval

In the absence of a well annotated UTR dataset for 12220 Glossina genes and for the purpose of this study the 5'- and 3'-UTRs were defined as a 1000 bps up- and down-stream of a gene respectively. As such, a perl script was written to retrieve the UTR sequences. The script scans through the .GFF file (glossina_morsitans_core_v0_0_3_oct9_All.gff3) for the word "CDs" and extracts a 1000 bps up- and down-stream using the genomic sequence. Additionally, the contiguous genes were excluded from the study, due to their overlapping 5'- and 3'-UTRs. The UTR sequences are saved in a fasta-formatted file, which can be used for further analysis.

2.2.1.2 Pattern matching analysis

The program "dna-pattern" as part of the RSAT tools (Thomas-Chollier *et al.*, 2008) was used to scan UTR sequences with string-based patterns of *IRE*s. These patterns were provided to RSAT as regular expressions using IUPAC-IUB symbols, covering a wide range of canonical, non-canonical and SELEX-based *IRE* patterns (Table 2.2). RSAT pattern matching analysis was used to refine the putative list of UTR sequences containing *IRE*-like patterns.

2.2.1.3 Identification of *IRE* stem loop structures

The UTRs of *Glossina morsitans* were screened for the presence of *IRE* stem loop structures using Search for *IRE*s (SIRE) (Campillos *et al.*, 2010), whereby the 3'- and 5'-UTR sequences were separately provided as inputs. The SIRE prediction utilizes structure analysis, predicted RNA fold, and folding energy data, which are then used to flag the identified *IRE* based on its similarity with well-characterized *IRE*s. It further provides information about the nucleotide composition of the apical loop and the presence of possible mismatches and or 3' bulged nucleotide in the upper stem, presence of a guanine (G) at n25 as well as the number of wobble base pairs (Campillos *et al.*, 2010). The predicted

IRE is then further folded using "RNAfold" as part of the Vienna RNA package (Gruber *et al.*, 2008).

Table 2.2: Canonical and non-canonical *IRE*-associated patterns.

Pattern	*Regulatory expression*
C1	CNNNNNCAGTGN
C2	CNNNNNCAGAGH
S1	CNNNNNCTGTGY
S2	CNNNNNCTTAGC
S3	CNNNNNCAATGC
S4	CNNNNNCAGGGN
S5	CNNNNNTAGTAY
S6	CNNNNNTAGGAT
S7	CNNNNNTAGAAY
S8	CNNNNNTAGCAG
S9	CNNNNNGAGTCR
S10	CNNNNNGAGCCR
S11	CNNNNNGAGAGK
S12	CNNNNNGAGTGW

*The first two patterns represent canonical forms of *IRE*, while the rest are the result of SELEX experiments (non-canonical *IRE*s).

The results of SIRE predictions are assessed based on three levels of stringency, which is presented as the sum score value with "High" scoring from 8 to 6, "Medium" from 5.9 to 4 and "Low" scoring from 3.9 to 0. The "High" quality motifs include all known and well characterized *IRE*s *in vitro* and *in vivo*. Furthermore, the "Medium" and "Low" *IRE* motifs include *IRE*-like motifs that do not meet most *IRE* prediction criteria. It is notable that *IRE*s present in the mouse *Hao1* and the human *CDC42BPA* mRNAs whose, functionalities beside the SELEX-evidence have not yet been verified *in vivo*, are predicted by SIRE as "Medium" quality. As such for the purpose of this study, we only considered *IRE*s with High confidence as well as medium-scored *IRE*s containing no mismatches for further analysis.

The output of the program is in tabular format as well as schematic representation of the important features of the identified *IRE*s. The apical loops are presented in red,

C8 bulge in green, G.U or U.G wobble base in orange, and 3' bulge or upper stem mismatch in yellow (Figure 2.1). Furthermore, information such as *IRE* start and end positions, free energy and quality flag, as well as loop type are stored in GFF format. The filtering step was carried out and *IRE*s with "High" score were retrieved. To account for genes with IRE-like elements and considering that ferritin *IRE* was ranked "Medium" by SIRE as part of the filtering step, genes ranked as "Medium" with no mismatch were also considered for further analysis. The energy score of the predicted RNA folds were assessed and those with zero folding energy were discarded to reduce false positives. The identified *IRE* sequence patterns were collected in fasta format file for further analysis. Additionally the protein sequences associated with the identified genes were extracted from the available Glossina peptide file for further analysis.

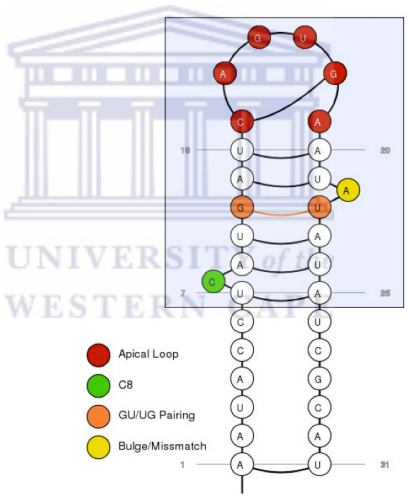


Figure 2.1: Stem loop structure of *IRE*. The circles in red represent the apical loop sequence, and the green refers to the conserved C-bulge. This pattern may vary for different genes, and can be observed in the patterns of canonical and SELEX-based *IRE*s.

2.2.2 Motif discovery and analysis

2.2.2.1 Background model determination

Some upstream sequences are known to have compositional biases that are often captured by the background model. In line with this, the retrieved UTR sequences were assessed

using fasta-get-markov method as implemented in the MEME suite (Bailey *et al.*, 2009). To predict the background model, the program estimates a Markov model from the fasta sequence file, while ignoring ambiguous characters prior to model computing. As such the background models contribute significantly to the accuracy of results in a motif discovery approach.

2.2.2.2 Database collection of UTR elements

A MEME-readable database of UTR elements is required to later be used for the annotation of the identified motifs. A database of all known UTR elements was created, through retrieving their sequences from UTRsite (`http://utrsite.ba.itb.cnr.it/`) and RFAM database (`http://rfam.sanger.ac.uk/`). The gathered sequences for each of the known elements were searched for an overrepresented pattern using MEME, by specifying the minimum and maximum width associated with each element. This was carried-out for 51 5'-UTR and 79 3'-UTR elements. The resultant matrix files were then summarized into a single meme file using meme2meme program.

2.2.2.3 Motif identification and comparison

Motif discovery was carried out using MEME v4.9.0 (Bailey *et al.*, 2009). MEME uses expectation maximization methods to identify over-represented motifs in the query sequence. As part of the parameters, the distribution of motif occurrences was set to zero or one per sequence. Furthermore, a maximum of ten motifs were specified to be identified by the program. Additionally, the minimum and maximum motif width were set to six and thirty respectively. The p-value cut-off to consider a motif significant was specified as 10^{-4}. Parameter specification was based on the visual inspection of the alignment.

2.2.2.4 Motif comparison

The annotation of the identified motifs were carried out through their comparison with the database of known UTR elements (Section 2.2.2.2). The matrix representation of the identified motifs, were used as inputs to the Find Individual Motif Occurrences (FIMO), which then compares the DNA motifs to the elements of the database. The output of the program includes a list of matching motifs that are ranked by q-value (minimal

false discovery rate at which the results are considered significant). Accordingly the best matching motif (based on the q-value) is then used to annotate each of the identified motif.

2.2.2.5 Functional prediction of novel motifs

To assign roles to the identified motifs the Gene Ontology for MOtifs (GOMO) program was used which requires the scoring file and a GO-term database, for which the *Drosophila melanogaster* database (from the available precomputed databases implemented in MEME) was used. The scoring file is generated using the Average Motif Affinity (AMA), which scores the query sequence given a DNA binding motif and treats each position as a possible binding site. As a result, AMA outputs an average motif affinity score for the identified elements, which is then used as inputs for GOMO. Accordingly, GOMO provides a list of GO-terms that are significantly associated with the target genes of the identified motifs. The putative GO-terms are sorted based on the q-value, which is calculated using the Benjamini and Hochberg method (Benjamini *et al.*, 2001). Hence, the best scoring GO is assigned to each of the corresponding motif identified in this study.

2.2.2.6 UTRScan prediction of regulatory elements

To complement the results of MEME, the UTR sequences containing putative *IRE* stem loop structures were further analyzed using UTRScan. The program searches for the previously identified UTR patterns that are available in UTRsite (Grillo *et al.*, 2010). UTRsite provides details of various regulatory elements present in the UTR regions that are experimentally validated. The UTR sequences of Glossina genes, identified to be *IRE*-regulated were provided as input, while 5'- and 3'-UTR sequences were searched in separate runs to differentiate between patterns that are specific to each of these regions and/or those that are commonly found in both the 5'- and 3'-UTRs.

2.2.2.7 Motif clustering analysis

To cluster genes into families based on their motif sharing, MotifCluster tool (Hamady *et al.*, 2008) was used. The algorithm allows for visualizing motifs mapped to trees,

sequences and three-dimensional structures. MotifCluster makes use of six distance matrices, some of which cluster sequences that share motifs retained in the same order within the sequences, while others mainly focus on motifs that are shared. Furthermore, motifs that are longer and more significant *i.e.* shared among several sequences, are assigned more weight by the program, while the reverse applies to shorter and less significant motifs i.e shared by a few sequences. As such to gain perspective on the identified motifs, the matrix representations of the discovered motifs as retrieved by MEME (Section 2.2.2.3) along with the UTR sequences were provided as inputs to the MotifCluster package. The program was run using default parameters *i.e.* edge threshold of 2, conservation threshold of 0.8, weight motifs set to "yes" and the ID mapping file was specified which included GO assignments for each gene. Furthermore, as part of the generated results a weighted graph showing the relationship between all sequences in terms of the motifs they share is generated. Vertices in the graph represent sequences in the input data, while edges show the relationship between sequences that share two or more motifs. As a result sequences are connected to each other if at least two common motifs are shared. Following the analysis, edges with weights less than the specified "edge threshold" (default is 2) are removed from the graph.

2.2.3 ProtFun

The genome sequence of *Glossina morsitans* has led to the identification of 12220 genes, a significant proportion of which, are functionally uncharacterized. It is therefore of interest to use algorithms designed to assign these genes to functional categories. As such for the purpose of this study ProtFun v.2.2 (Jensen *et al.*, 2003) was used, which is an *ab-initio* prediction tool that uses sequence-based features such as localization, secondary structure and post-translational modification predictions to assign functions. Furthermore, the prediction of functional categories is achieved using neural networks trained to recognize features associated with a specific function (Appendix A).

ProtFun classifies genes into 12 cellular role categories, namely amino acid biosynthesis, biosynthesis of cofactors, cell envelope, cellular process, central intermediary metabolism, energy metabolism, fatty acid metabolism, purines and pyramidines, regulatory function,

translation as well as transport and binding. Furthermore, proteins are assigned to 14 gene ontology categories including signal transducer, receptor, hormone, structural protein, transporter, ion channel, voltage-gated ion channel, cation channel, transcription, transcription regulation, stress response, immune response, growth factor and metal ion transport. Additionally, as part of the ProtFun analysis the protein is classified as enzyme or non-enzyme with the former being further classified into six enzyme groups namely oxidoreductase, transferase, hydrolase, isomerase, ligase and lyase.

Furthermore, secreted proteins play critical roles in several biological processes such as cellular immunity and communication, hence contribute to discovery of novel biomarkers (Hathout, 2007; Xue *et al.*, 2008). Most eukaryotic secreted protein consists of a signal peptide typically 15-30 amino acids long at the N-terminus. To this end, SignalP-NN v.4.1 (Petersen *et al.*, 2011) and TMHMM v.2.0 (Jensen *et al.*, 2003) as implemented in ProtFun allowed for the inference of putative secreted proteins, which were further classified as signal peptides or signal anchors.

2.2.4 Predicting sub-cellular localization

Eukaryotic cells contain several sub-cellular compartments, which make the annotation of proteins sub-cellular localization a major contributor to functional annotation. Several approaches for predicting sub-cellular localization as well as the presence of secretary signal peptides have been introduced including the SignalP scheme (Bendtsen *et al.*, 2004), LOCTree (Nair and Rost, 2005), BaCelLo (Pierleoni *et al.*, 2006), TargetP (Emanuelsson *et al.*, 2007) and WoLF PSORT (Horton *et al.*, 2007). PSORT is one of the most robust tools available for the prediction of sub-cellular localization. PSORT is capable of distinguishing between, 17 different sub-cellular localization in eukaryotes. Several prediction programs as well as statistical methods are implemented in PSORT. WoLF-PSORT is an updated version of PSORT program for the prediction of eukaryotic sequences. WoLF-PSORT uses amino acid sequence features and converts them to numerical vectors that are weighted with k-nearest neighbor classifier (for details on the method refer to Horton *et al.*, 2007). WoLF-PSORT has been ranked with high accuracy by several studies (Klee and Ellis, 2005; Min, 2010). As such, WoLF-PSORT v.0.2 was used to determine

sub-cellular localization of Glossina proteins. The fasta-formatted protein sequences were provided as input to WoLF-PSORT. The results were captured in a localization feature table and localizations with highest probability (indicated by the numbers representing the number of nearest neighbors *i.e.* proteins in WoLF training data that have the most similar localization features) were considered (Appendix A).

2.2.5 Orthology detection

The cDNA sequences for six blood feeding and five non-blood feeding insects were retrieved from Ensembl (`http://metazoa.ensembl.org/info/data/ftp/index.html`). A perl script named longORF.pl was used to identify the longest ORF for each sequence. For the purpose of this study, we specified a "notstrict" parameter so that the script not only allows for the canonical ATG but also, searches for longest ORF from the start of the sequence. For further analysis the identified ORF sequences were translated into proteins using "transeq" as part of the EMBOSS suite (Rice *et al.*, 2000). The resultant peptide sequences were then used as inputs for the program QuartetS (Yu *et al.*, 2011).

In essence orthology detection involves the differentiation between orthologs and paralogs that evolutionary share common ancestors, though undergoing different evolutionary events *i.e.* speciation and duplication respectively (Koonin, 2005). To establish orthology relationships for the identified 12220 genes of *Glossina morsitans* the QuartetS method was applied (Appendix A). According to this approach an approximate phylogenetic analysis of quartet gene trees are carried out to identify duplication events and distinguish paralogs from orthologous genes (Yu *et al.*, 2011). As an initial step QuartetS performs sequence similarity search using BLAST program with parameters being set at blast_evalue = $1e^{-5}$, blast_matrix = BLOSUM80, blast_score = 50, while the default setting was used for other parameters. According to the authors this method outperforms the highly specific OMA (Orthologous MAtrix) method (for details about the method refer to Roth *et al.*, 2008), while predicting 50% more orthologs with 50% lower false positives rate compared to the commonly used BBH (Bidirectional Best Hit) method (for details about the method refer to Zhang and Leong, 2012). This method is suitable for large-scale genome-wide orthology detection. QuartetS extracts the evolu-

tionary evidence from the quartet gene trees that are formed by two genes of interest and two genes from the third genome, and is carried out for all available genomes. Hence, possible duplication events are identified which allows for discriminating between orthologs and paralogs. Finally, as part of the post-processing analysis clustering of the pair-wise orthologs is carried-out. This is achieved using the MCL clustering program implemented in QuartetS (`http://micans.org/mcl`). MCL carries-out an unsupervised clustering step following which genes are clustered into orthologous groups. Furthermore, the size and the number of the clusters are controlled through setting the inflation parameter. As also suggested by the authors, the inflation value of three was used for the purpose of this study as evaluation of other values proved insignificant on the results (Yu *et al.*, 2011).

2.3 Results and Discussion

2.3.1 Identification of *IRE*-regulated genes

By definition UTR is the untranslated region on either sides of the coding mRNA sequence, which starts at the transcriptional start site and ends before the start codon with respect to the 5'-UTR and starting immediately following the stop codon for 3'-UTR. Though it is desirable to use a set of well-annotated UTR sequences for investigating the presence of *IRE*s, this annotation is not always available. Hence, in the absence of a curated UTR dataset for *Glossina morsitans*, a thousand base pairs up- and down-stream of the start and stop codons were extracted for 12220 annotated Glossina genes. The use of 1000 bp is consistent with several previous studies (Ohler *et al.*, 2004; Jareborg *et al.*, 1999).

Several patterns have been associated with *IRE*s some of which have been experimentally validated and some are the results of SELEX experiments. There are two classical/canonical patterns for *IRE*s (5'-CNNNNNCACNNNNNCAGUGN-3', and 5'-CNNNNNCAGAGN-3') and 16 SELEX motifs (Henderson *et al.*, 1994; Butt *et al.*, 1996). These patterns were used in a pattern matching analysis using the regulatory sequence analysis tools RSAT. From 24440 5'- and 3'-UTRs that were analyzed (for 12220 Glossina genes), 1233 (5.04%) were identified to contain canonical forms of *IRE*, while 5383 (22.02%) were shown to have the non-canonical form. The use of RSAT pattern matching tool allowed for the identification of a list of genes with canonical and non-canonical *IRE* patterns. The folding energies of the 6616 UTRs containing the canonical and non-canonical *IRE* motifs were assessed using the SIRE tool. Among the 6616 UTR sequences provided as input to SIRE, 902 were identified to have *IRE*s (canonical and non-canonical) in their untranslated regions. Additionally, SIRE provides a score based on the quality of the predicted motifs, according to which, 72 *IRE*s had "High" quality, 299 "Medium" quality and 531 "Low" qualities (Figure 2.2).

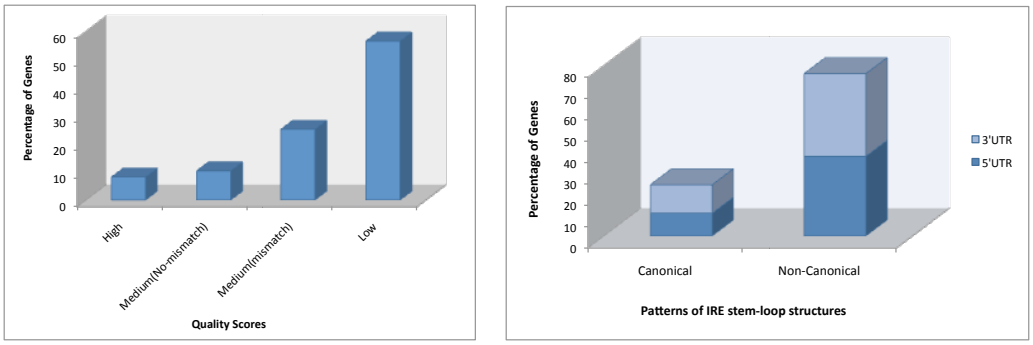


Figure 2.2: Pattern and score distributions of the identified *IRE* stem loop structures. Left: is a representation of the identified *IRE*s with high, medium and low scores, from which only the high and medium scores with no mismatches were accepted as putative *IRE* stem loop structures. Right: from the putative *IRE* structures identified, the majority resemble the non-canonical forms, while all identified structures are similarly distributed between the 5'- and 3' -UTRs.

The "High" quality motifs have the confidence score of 8 to 6 and include all known and well characterized *IRE*s *in vitro* and *in vivo*. Furthermore, the "Medium" and "Low" *IRE* motifs have the confidence scores of 5.9 to 4 and 3.9 to 0.0 respectively, and include *IRE*-like motifs that do not meet most *IRE* prediction criteria. It is notable that *IRE*s present in the mouse Hao1 and the human *CDC42BPA* mRNAs whose, functionalities beside the SELEX-evidence have not yet been verified *in vivo* are predicted by SIRE as "Medium" quality. As such, only *IRE*s with "High" confidence as well as medium-scored *IRE*s containing no mismatches were considered for further analysis. Based on the score filtering a total of 150 putative *IRE*-containing genes were retrieved among which, 72 genes were identified with 5'-*IRE*s and 78 with 3'-*IRE*s (Appendix B).

As a result 150 genes with known *IRE*-elements were identified in the genome of *Glossina morsitans* (AppendixB). While the majority are novel *IRE*-regulated genes, two genes know to have *IRE*s in their UTRs were also identified, namely ferritin heavy chain and serine/threonine protein kinase (*MRCK-alpha*). Though literature suggests the linkage of some of these genes to mechanisms responsive to iron, no exact information on their *IRE*-associated regulation exists. Furthermore, several hypothetical proteins with no defined function were also identified to be *IRE*-regulated. These include TMP007137, TMP009128, TMP002546, TMP002921, TMP003628, TMP004581, TMP008259, TMP012389, TMP005219, TMP00-5827, TMP007908, TMP009332, TMP013384, TMP009102, TMP010544,

TMP010707, TMP004292, TMP006517, TMP014030, TMP009821 and TMP003060 for which an iron-regulatory mechanism of action may be inferred.

2.3.2 Co-regulators of the putative *IRE*-regulated genes

Considering that functionally related genes are co-regulated by specific cis-regulatory RNA motifs controlling translational efficiency (Iengar and Joshi, 2009; Segal *et al.*, 2005), it is of interest to identify such motifs to help understand the regulatory mechanisms governing iron metabolism in tsetse fly. The availability of genome sequences for several blood feeding and non-blood-feeding insects have paved the way for more detailed studies of gene evolution in the context of transcriptional and translational regulation. Furthermore, the availability of a detailed database of transcription factors (TFs) for Drosophila along with their associated PWMs provide a good resource for the identification and characterization of newly identified regulatory elements.

With the aim of characterizing the post-transcriptional regulatory elements governing iron metabolism in the blood-feeding insect, *Glossina morsitans*, the UTRs of *IRE*-regulated mRNAs *i.e.* putative iron metabolism genes were further analyzed for the presence of other enriched regulatory motifs. This was achieved through the use of Multiple EM for Motif Elicitation (MEME) program as part of the MEME suite (Bailey *et al.*, 2010), for *de novo* motif discovery as well as the UTRScan for searching for the presence of experimentally validated elements. As a result, seven sequence motifs were identified by MEME in the 5'- and 3'-UTRs of Glossina genes, while UTRScan found twelve that share similarities with the experimentally validated elements (Tables 2.3, 2.4).

The MEME-predicted motifs were annotated through their mapping to the database of known UTR elements, collected from the UTRsite and Rfam using TOMTOM. The TOMTOM program as part of the MEME suite compares the results of the motif discovery step to the elements of the database of known UTR motifs. As a result a set of matching motifs are reported that are ranked based on q-values. Figure 2.3 shows the best-ranked motifs resembling significant similarities with the input motif sequences. Additionally

Gene Ontology for MOtifs (GOMO) was used to determine putative roles for the identified motifs, through GO annotations that were assigned to target genes (Figure 2.3).

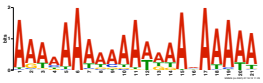
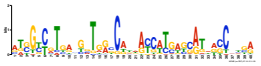
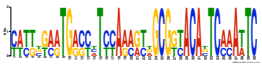
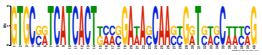
Motif No	UTR	Motif name	GOMO	LOGO
1	3'	PAS	MF triplet codon-amino acid adaptor activity BP translation CC cytosol CC protein kinase CK2 complex BP synapse organization	
2	3'	FIE3	NA	
3	3'	GAIT	BP chromatin assembly or disassembly CC nucleosome MF DNA binding CC Cajal body	
4	3'	K-Box	NA	
5	5'	UNR-bs	MF transcription factor activity CC protein kinase CK2 complex BP regulation of transcription	
6	5'	crcB	NA	
7	5'	Bacteroid-trp	MF protein binding BP response to heat MF transcription factor activity BP pericardial cell differentiation BP cell development	

Figure 2.3: Patterns of motif sharing in the UTRs of the *IRE*-regulated genes. The identified motifs as a result of MEME analysis were found to share similarities with known motifs listed in column three. The last column presents the consensus sequence for the identified motifs. Some of the motifs have no assigned GO term as a result of GOMO analysis.

The results of this study may be summarized as a set of 5' and 3' elements identified by either MEME or UTRScan, while five elements were commonly identified by both methods. The observed difference between the results obtained from these methods is mainly due to the fact that MEME carries out a motif discovery analysis without any prior knowledge about these elements. UTRScan on the other hand searches for patterns specific to known UTR-element. As such, the results of both analysis can provide insight into novel and known co-regulatory elements controlling iron-responsive genes in *Glossina morsitans*. To further understand the significance of the identified elements a brief review of their roles will be provided.

One of the elements identified in the 5'-UTRs of 76 *IRE*-regulated genes is the *uORF* (Table 2.3). This element can be characterized by a start codon, a short open reading frame (ORF) and a stop codon, located upstream of the major start codon that initiates translation. When encountering a *uORF*, a ribosome can either scan through it until the major ORF is recognized, hence translate the *uORF* and then re-initiate translation of the major ORF, or it may translate the *uORF* and then stall. As a result, the latter can decrease translation of the major ORF and shorten mRNA half-life (Morris and Geballe, 2000). According to literature *uORF* has been identified in an estimated 50% of all annotated human and mice transcripts, while its mutations have been confirmed in several human-disease-associated genes (Calvo *et al.*, 2009).

Another overrepresented 5'-UTR elements identified is the *IRES*, present in 32 *IRE*-regulated genes (Table 2.3). This suggests an *IRES*-specific mode of translation for these genes. As expected a considerable number of these genes were involved in transcription and translation, and amino acid biosynthesis (Section 2.2.3). Several cell envelope, transport and binding, and energy metabolism genes were also identified among the *IRES*-regulated genes. Furthermore, various cell growth, proliferation and apoptosis-related mRNAs have structured 5'-UTRs that often comprise of internal ribosome entry sites (*IRES*s) and upstream open-reading frames (*uORF*s) (Pickering and Willis, 2005).

Another identified regulatory element is the *crcB* RNA motif located in the 5'-UTR of *IRE*-regulated mRNAs (Table 2.3). Many genes including those involved in DNA repair, K^+/Cl^- ion transporters, formate hydrogen lyases and *nifU* iron-sulfur proteins have been proposed to be regulated by *crcB* RNA elements. The additional motif identified in the 5'-UTR of *IRE*-containing genes is the Bacteroid-trp-like RNA motif, which is a conserved element detected in the 5'-UTR of genes that encode enzymes used in the synthesis of the amino acid tryptophan (Weinberg *et al.*, 2010).

The results of this study have further led to the identification of several 3'-UTR elements including *PAS*, *FIE3*, *GAIT*, *GY-box*, *Brd-box*, *K-box*, *ARE* and *CPE* (Table 2.4). The polyadenylation signal (*PAS*) is believed to be involved in controlling mRNA stability and

degradation. Most post-transcriptional cis-regulatory sequences are known to be located between the stop codon and *PAS*.

Ftz instability element 3' (*FIE3*) is another element identified in the 3'-UTR of several *IRE*-regulated genes. *FIE3* has been shown to localize in the 3'-UTR of the fushi tarazu mRNA, involved in the establishment of Drosophila embryonic body plan (Riedl and Jacobs-Lorena, 1996). This RNA motif contains a GU-rich sequence and is located upstream of the polyadenylation signal (Riedl and Jacobs-Lorena, 1996; Vlasova-St Louis and Bohjanen, 2011). Insect embryonic development depends on precise expression of maternal and zygotic pattern-forming genes, but more importantly they are dependent on sufficient amount of iron for such processes, which provides support for the iron-regulated expression of these genes. As such it could be postulated that *IRE* regulates the expression optimization or inhibition of these genes based on the insect's iron levels, the absence of which could negatively affect the fly's reproductive health and embryonic development.

The gamma interferon inhibitor of translation elements (*GAIT* element) is another cis-acting RNA motif identified in the 3'-UTRs of the *IRE*-regulated target genes (Sampath *et al.*, 2003). This motif is known to be located in the ceruloplasmin (*Cp*) mRNA. The regulatory effects of *GAIT* element is in selective translational silencing of the *Cp* as well as other possible transcripts and is achieved through the binding of *GAIT* inhibitor complex to the *GAIT* element in the 3'-UTR (Mazumder *et al.*, 2005). *Cp* is a multifunctional copper protein that is known to play an important role in iron homeostasis and inflammation. The ferroxidase activities of *Cp* has allowed for the inference of its role in iron metabolism. Furthermore, *Cp*-stimulated iron loading into apo-transferrin and iron efflux from the liver provides support for its iron regulatory-related activities (Mazumder *et al.*, 2006). Based on these results an interplay between iron homeostasis and inflammation has been drawn which is evident in our findings as well.

GY-box, *Brd-box*, and *K-box* motifs are conserved sites in 3'-UTR that are complementary to the 5'-end of miRNAs. These motifs are often found in the UTR of *Notch* target genes that allow for sufficient miRNA-mediated regulation. *Notch* signaling pathway

is involved in the proliferative signaling during neurogenesis, being present in the cell membrane the *Notch* proteins allow for intracellular *Notch* signal to be transferred from cell to cell, hence allowing for an influential effect of groups of cell on one another (Lai *et al.*, 1998, 2005; Artavanis-Tsakonas *et al.*, 1999; Schweisguth, 2004). As a result if one cell expresses a specific trait, this may be switched off in neighbor cells (Oswald *et al.*, 2001). Furthermore, the *Notch/Lin-12/Glp-1* receptor family in Drosophila and *C.elegans* were identified in the specification of cell fate during development (Singson *et al.*, 1998). Furthermore, the involvement of Notch signaling in various biological processes has been well established which include, embryogenesis, development of central nervous system and function, cardiovascular and endocrine development (Apelqvist *et al.*, 1999; Lammert *et al.*, 2000; Jensen, 2004). Additionally, their associated miRNAs are highly conserved in insects and a number of the above mentioned elements including *GY-box*, *BRD-box* and *K-box* have been identified in Drosophila, mosquitoes, bees, moth and several other insect species (Lai *et al.*, 2005). Based on the findings pertaining to the regulation of these genes in an *IRE*-mediated mechanism, a putative role for Notch signaling pathway may be elucidated in iron metabolism.

AU-rich elements (*ARE*s) identified in the 3'-UTR of *IRE*-regulated genes either determine the degradation or mediate the repression of translation for the corresponding mRNA. Studies have proposed a link between *p38alpha MAPK* and *AREs*, by introducing *AREs* as targets for *P38alpha* (Zhao *et al.*, 2008). *P38alpha MAPK* is crucial for post-transcriptional regulation of most pro-inflammatory cytokines. Furthermore, the cytoplasmic polyadenylation element (*CPE*) identified in the 3'-UTR regulate translation through binding *CPE*-binding protein, which results in the extension of the polyadenine tail and subsequent activation of translation. One of the major characteristics of *CPE* is its involvement in oogenesis and spermatogenesis (de Moor and Richter, 1999; Luitjens *et al.*, 2000; Charlesworth *et al.*, 2004).

Additionally, two elements were identified in both the 5'- and 3'-UTRs of *IRE*-regulated genes in Glossina, namely the *UNR*-bs and *SXL*-bs. In human, *UNR*-bs is involved in *c-Fos* protein destabilization as well as translational repression of the poly(A)-binding

protein (*PABP*) (Chang *et al.*, 2004; Patel *et al.*, 2005) (Table 2.3, 2.4). Furthermore, *UNR* functions as a translational regulator of several mammalian transcripts through the *IRES* (Hunt *et al.*, 1999; Boussadia *et al.*, 2003; Brown and Jackson, 2004; Dormoy-Raclet *et al.*, 2005; Tinton *et al.*, 2005; Schepens *et al.*, 2007). *In vivo* and *in vitro* studies have identified *UNR* as a critical factor in major coding-region determinant of instability (*mCRD*)-mediated mRNA turnover due to its function as a *mCRD*-binding protein as well as a *PABP*-interacting protein. As a result *mCRD/UNR* complex is considered as the responsible unit in the formation of deadenylation/decay mRNP complex (Chang *et al.*, 2004). In Drosophila the translational repression of male-specific-lethal 2 (*MSL2*) mRNA by Sex-lethal (*SXL*) requires the functioning of *UNR* (Abaza *et al.*, 2006). Hence, the explanation of the putative role of iron in the reproductive aspects of insects. *MSL2* is a component of Drosophila dosage compensation complex that regulates the expression of X-linked genes between males (XY) and females (XX). This is achieved through promoting hyper-transcription of the single male X chromosome (Gelbart and Kuroda, 2009). Hence, the presence of *UNR*-bs in the UTRs of *IRE*-regulated genes point to the implications of iron in the reproductive aspects of insect biology.

The *SXL* binding site (*SXL*-bs) is evident in both the 5'- and 3'-UTRs of *IRE*-regulated genes, which is where *SXL* binds to inhibit *MSL2* translation (for review refer to Graindorge *et al.*, 2011). It has been suggested that the binding of *SXL* to both UTRs allows for tight translational repression whereas a partial inhibition can be achieved through binding to each UTR alone (Bashaw and Baker, 1997; Gebauer *et al.*, 1999). The binding of *SLX* to 3'-UTR inhibits the 43S ribosomal complex recruitment to the mRNA, while its binding to the 5'-UTR prevents the scanning of complexes that have escaped the 3'-mediated inhibition (Gebauer *et al.*, 2003; Beckmann *et al.*, 2005). Furthermore, the *uORF* upstream of the *SXL*-bs has been recognized to be important for 5'-UTR-mediated translational repression (Medenbach *et al.*, 2011).

Table 2.3: A list of motifs identified in the 5'-UTRs of *IRE*-regulated genes in Glossina, using MEME and UTRScan.

Gene ID	Gene Name	Motifs	
		MEME	UTRscan
TMP002406	Cuticle protein	-	IRES,UNRbs,uORF
TMP002546	hypothetical protein	Bacteroidtrp	UNRbs,uORF
TMP002587	alternative splicing factor SRP20/9G8	-	SXL_BS,uORF
TMP002921	hypothetical protein	-	uORF
TMP003021	kelch-like protein diablo	-	uORF
TMP003297	TRP channel protein	-	SXL_BS,uORF
TMP003577	st7-like protein	UNRbs	IRES,uORF
TMP003628	hypothetical protein	Bacteroidtrp,UNRbs	uORF
TMP004029	reverse transcriptase	UNRbs	IRES,uORF
TMP004047	GJ14155	Bacteroidtrp	IRES,uORF
TMP004240	elongation factor1 alpha	Bacteroidtrp	SXL_BS,UNRbs,uORF
TMP004517	importin9	Bacteroidtrp,crcB	uORF
TMP004746	myosin heavy chain type II	-	IRES,uORF
TMP004763	spire	Bacteroidtrp,UNRbs	uORF
TMP004803	GJ11047	UNRbs	uORF
TMP005026	beta-carotene dioxygenase	Bacteroidtrp,UNRbs	UNRbs,crcB,uORF
TMP005160	homeobox protein	UNRbs	IRES,UNRbs,uORF
TMP005167	arp 23 complex subunit arpc5	UNRbs	IRES,UNRbs,uORF
TMP005219	Unknown	-	IRES,uORF
TMP005259	gly-rich protein	UNRbs,crcB	IRES,uORF
TMP005357	LIX1-like protein	-	UNRbs,uORF
TMP005511	atg5	-	IRES,uORF
TMP005554	proton-associated sugar transporter a-like	-	IRES,uORF
TMP005765	heterochromatin-associated protein	-	uORF
TMP006019	insulin-like peptide 1	Bacteroidtrp	uORF
TMP006162	shaking B	Bacteroidtrp,UNRbs	IRES,SXL_BS,UNRbs,uORF
TMP006242	nmda receptor glutamate-binding chain	UNRbs	IRES,uORF
TMP006315	acylphosphatase	-	UNRbs,uORF
TMP006377	coenzyme Q-binding protein	-	IRES,uORF
TMP006517	GI16026	-	uORF
TMP006591	saccharopine dehydrogenase domain-containing protein	Bacteroidtrp,UNRbs	IRES,uORF
TMP006617	20s proteasome regulatory subunit beta	Bacteroidtrp	uORF
TMP006651	phenylalanyl-tRNA synthase subunit beta	-	IRES,SXL_BS,UNRbs,uORF
TMP006853	visual system homeobox 1	UNRbs	uORF
TMP007030	ganglioside-induced differentiation-associated protein 1	UNRbs	uORF
TMP007122	pdgf-and vegf-related factor isoform e	-	IRES,uORF
TMP007350	defective proboscisextension response	Bacteroidtrp	uORF
TMP007441	mitochondrial processing peptidase beta subunit	-	uORF
TMP007537	integrator 4	UNRbs	SXL_BS,uORF
TMP007548	retinoid-and fatty acid-binding glycoprotein	UNRbs	IRES,UNRbs,uORF
TMP007550	GH23986	UNRbs	IRES,uORF
TMP007593	barren	UNRbs	uORF
TMP007653	Abdominal-B protein	UNRbs,crcB	IRES,uORF
TMP007741	meiosis-specific nuclear structura l	UNRbs	IRES,uORF
TMP007858	Semaphorin	-	uORF
TMP007908	Unknown	-	uORF
TMP008000	nucleosome remodeling factor - isoform a	UNRbs	UNRbs,uORF
TMP008259	hypothetical protein	Bacteroidtrp	IRES,uORF
TMP008274	CG15118	UNRbs	uORF
TMP008329	gustatory receptor isoformc	UNRbs	IRES,uORF
TMP008384	LIM domain protein	Bacteroidtrp,UNRbs	uORF
TMP008448	autophagy-related protein	UNRbs	SXL_BS,UNRbs,uORF
TMP008747	stoned isoform e	-	uORF
TMP008854	receptor tyrosine kinase	-	IRES,UNRbs,uORF
TMP009102	GG15559	UNRbs	IRES,SXL_BS,uORF
TMP009128	hypothetical conserved protein	UNRbs	SXL_BS,uORF
TMP009134	150 Kda dynein-associated polypeptide	UNRbs	uORF
TMP009157	ubiquitin-conjugating enzyme9	UNRbs	uORF
TMP009332	GF15465	-	uORF
TMP009552	synaptotagmin	UNRbs	UNRbs,uORF
TMP009821	GJ22290	Bacteroidtrp,UNRbs	UNRbs,uORF
TMP010544	GG21264	UNRbs	IRES,uORF
TMP010651	trna delta-isopentenyl pyrophosphate transferase	-	uORF
TMP010673	ferritin heavy chain-like protein	UNRbs	uORF
TMP011448	brain chitinase and chia	Bacteroidtrp	IRES,uORF
TMP012368	GJ11024	UNRbs	IRES,uORF
TMP012387	cytochrome P450	UNRbs	IRES,SXL_BS,uORF
TMP012391	cytochrome B561	Bacteroidtrp	uORF
TMP012414	Sticks and stones, isoform A	-	uORF
TMP013100	sox box protein isoform a	-	IRES,uORF
TMP013254	btk family kinase at isoform a	Bacteroidtrp,UNRbs	IRES,uORF
TMP013701	zinc finger protein 106	-	uORF
TMP013886	trna(uridine-2-o-)-methyl transferase trm7	UNRbs	uORF
TMP014187	mediator 100kd	UNRbs	IRES,uORF

Table 2.4: A list of motifs identified in the 3'-UTRs of *IRE*-regulated genes in Glossina, using MEME and UTRScan.

Gene ID	Complete Name	Motifs	
		MEME	UTRscan
TMP002103	cg11159-likeprotein	PAS	SXL_BS
TMP002357	dynein lightchain	PAS	KBOX
TMP002520	nucleoporin	PAS	-
TMP002718	shaggy	PAS	SXL_BS
TMP002821	Glucose-methanol-choline oxidoreductase	FIE3	-
TMP003059	serine/threonine-protein kinase	FIE3,GATE,KBOX,PAS	-
TMP003060	GK22170	FIE3,GATE,KBOX,PAS	-
TMP003203	DNA replication licensing factor mcm7	PAS	PAS
TMP003336	Ig-like domains containing protein	PAS	-
TMP003369	Serine/threonine-protein kinase	PAS	BRDBOX,KBOX,PAS
TMP003500	beta-hexosaminidase fdl	PAS	-
TMP003602	phosphoinositide3-kinase	PAS	-
TMP004121	zinc finger protein2	PAS	PAS
TMP004201	acetyl cholinesterse	FIE3	-
TMP004289	f-box protein	PAS	SXL_BS
TMP004292	GH20979	-	CPE,PAS,SXL_BS
TMP004413	Wnk,isoformA	PAS	BRDBOX
TMP004552	DNA polymerase epsilon	PAS	PAS
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	PAS	-
TMP004581	hypothetical protein	PAS	PAS
TMP004987	cuticular protein 67B	PAS	-
TMP005243	40S ribosomal protein	PAS	UNRbs
TMP005335	zinc carboxypeptidase a1	FIE3	-
TMP005511	atg5	PAS	SXL_BS
TMP005595	ATP-dependent RNA helicase dhx 36-like	PAS	PAS,UNRbs
TMP005827	unknown	PAS	PAS,UNRbs
TMP005901	neuronal pentraxin-1-like	FIE3	-
TMP006363	fork head box transcription factor	-	PAS
TMP006826	ring-box protein2	PAS	GYBOX,
TMP006868	5-hydroxy tryptamine receptor1	-	BRDBOX
TMP006998	saccharopine dehydrogenase domain containing protein	PAS	PAS
TMP007137	hypothetical conserved protein	PAS	SXL_BS
TMP007216	kinesin heavy chain	PAS	BRDBOX,PAS,UNRbs
TMP007308	folylpolyglutamate synthase, mitochondrial like	-	BRDBOX
TMP007388	hmg coenzyme a isoform a	FIE3,PAS	PAS
TMP007507	topoisomerase 3beta	PAS	-
TMP007537	integrator 4	PAS	PAS,SXL_BS
TMP007621	26s proteasome regulatory complex subunit psmd5	PAS	PAS,SXL_BS
TMP008441	gamma-tubulin ring complex	PAS	-
TMP008943	Glucosidase	FIE3	-
TMP009871	bruno-3 transcript	PAS	PAS
TMP009934	rhomboid	PAS	SXL_BS
TMP010016	tpr repeat-containing protein	PAS	PAS
TMP010134	clock work isoformb	PAS	PAS
TMP010317	general transcription factor IIH subunit1	PAS	-
TMP010821	Abdomnal A	PAS	SXL_BS
TMP010846	n-methyl-d-aspartate receptor	FIE3,PAS	GYBOX,UNRbs
TMP010987	sam-motif-containing protein	PAS	PAS,UNRbs
TMP011104	serine-pyruvate mitochondrial	PAS	PAS
TMP011371	laccase-2	PAS	-
TMP011421	tm2d1_drome	PAS	GYBOX
TMP011617	E3 ubiquitin-protein ligase	FIE3	-
TMP011786	disulfide isomerase	PAS	KBOX,SXL_BS
TMP011993	transposase	PAS	-
TMP012225	phd finger protein	FIE3,PAS	BRDBOX
TMP012389	hypothetical protein	PAS	PAS
TMP012534	ataxin-2 binding protein	PAS	PAS,UNRbs
TMP012772	reverse transcriptase	PAS	-
TMP013269	UBX domain-containing protein 7	PAS	-
TMP013384	GF18386	PAS	ARE2,KBOX,PAS
TMP013948	cch-cch-domain-containing protein2	FIE3,PAS	-
TMP013958	GYF domain-containing protein	PAS	UNRbs
TMP014133	unc-50	PAS	PAS,SXL_BS
TMP014204	60s ribosomal protein l23	PAS	UNRbs

2.3.2.1 Motif clustering analysis

All biological processes depend on the coordinated activity of a selected group of proteins. Before a given biological process takes place, it is necessary to synchronize the expression of genes that code for the set of implicated proteins. This synchronization can be achieved at the post-transcriptional level through the action of specific RNA binding proteins and non-coding RNAs that recognize UTR sequences shared by a group of genes. Considering the notion that co-regulated genes share similarities in functions (Allocco *et al.*, 2004), clustering of *IRE*-regulated genes based on their shared motifs allow for drawing a link between certain cis-acting elements and specific functions. In doing so MotifCluster tool was used (Hamady *et al.*, 2008), where genes presented in Table 2.5 and their assigned GO-terms were used to draw the network. The program assigns an identifier to each gene, which is used in the network illustration (Table 2.5).

MotifCluster searches for the presence or absence of all the identified motifs and clusters genes into families based on the motifs they share. Clustering of the sequences based on the shared motifs further points to the complexity of interaction among the iron metabolism genes. Figure 2.4 represents a network view of such clustering, whereby the connected components are determined by the edge threshold that was set to two (*i.e.* sequences that share two or more motifs). The circles represented in Figure 2.4 illustrate the number of genes with partners sharing two or more motifs, while the larger circles refer to those with the most motif-sharing with other genes present in the network. Furthermore, the colored circles represent genes corresponding to a specific GO category, suggesting that genes involved in transcription and translation have greater number of genes sharing more than two motifs. This indicates that there is a sophisticated interdependence between iron metabolism and transcription/translation of genes involved in a variety of processes. Additionally, several co-regulated genes were identified that share similar motifs with the majority of *IRE*-regulated genes present in the network (indicated by larger circles). These include, the phosphoinositid 3-kinase, *spire*, *MCM7*, mediator, integrator4 and 40S-ribosomal protein.

The nature and distribution of the regulatory motifs in the UTR sequences of *IRE*-regulated genes identified in this study may represent a valuable resource for the future validation of the co-regulated networks controlling iron metabolism.

Table 2.5: GO term assignments to *IRE*-regulated genes sharing two or more motifs, followed by an ID representing each gene in the MotifCluster network.

Gene ID	GO term	MotifCluster	Gene ID	GO term	MotifCluster ID
TMP004552	Biosynthesis	4	TMP013701	Metabolism	145
TMP009332	Biosynthesis	6	TMP007030	Purines & Pyrimidines	13
TMP003297	Biosynthesis	16	TMP009134	Purines & Pyrimidines	23
TMP007247	Biosynthesis	30	TMP004517	Purines & Pyrimidines	80
TMP005335	Biosynthesis	34	TMP004047	Purines & Pyrimidines	86
TMP012412	Biosynthesis	50	TMP004763	Purines & Pyrimidines	93
TMP014187	Biosynthesis	65	TMP008441	Purines & Pyrimidines	110
TMP005511	Biosynthesis	81	TMP008854	Purines & Pyrimidines	127
TMP007441	Biosynthesis	82	TMP003602	Purines & Pyrimidines	148
TMP006162	Biosynthesis	87	TMP004121	Transcription & Translation	0
TMP003577	Biosynthesis	117	TMP004553	Transcription & Translation	5
TMP007350	Biosynthesis	119	TMP007137	Transcription & Translation	7
TMP013269	Biosynthesis	121	TMP008942	Transcription & Translation	15
TMP007621	Biosynthesis	131	TMP010544	Transcription & Translation	20
TMP005554	Biosynthesis	133	TMP010707	Transcription & Translation	21
TMP006591	Biosynthesis	137	TMP007122	Transcription & Translation	24
TMP003203	Biosynthesis	142	TMP014204	Transcription & Translation	27
TMP010673	Cell envelope	8	TMP004803	Transcription & Translation	31
TMP008943	Cell envelope	14	TMP008259	Transcription & Translation	36
TMP002863	Cell envelope	19	TMP012491	Transcription & Translation	44
TMP009821	Cell envelope	33	TMP012225	Transcription & Translation	47
TMP003336	Cell envelope	46	TMP013384	Transcription & Translation	48
TMP006019	Cell envelope	59	TMP003060	Transcription & Translation	49
TMP005026	Cell envelope	63	TMP013958	Transcription & Translation	51
TMP012387	Cell envelope	77	TMP009934	Transcription & Translation	52
TMP005713	Cell envelope	118	TMP002718	Transcription & Translation	55
TMP007537	Cell envelope	135	TMP002921	Transcription & Translation	57
TMP011786	Cell envelope	136	TMP013254	Transcription & Translation	66
TMP005901	Cell envelope	138	TMP006363	Transcription & Translation	68
TMP005259	Cell envelope	149	TMP004581	Transcription & Translation	70
TMP012368	Metabolism	9	TMP008448	Transcription & Translation	71
TMP007550	Metabolism	28	TMP004029	Transcription & Translation	78
TMP004240	Metabolism	29	TMP010134	Transcription & Translation	83
TMP013948	Metabolism	32	TMP012534	Transcription & Translation	89
TMP003628	Metabolism	35	TMP006853	Transcription & Translation	97
TMP013886	Metabolism	37	TMP005167	Transcription & Translation	101
TMP011104	Metabolism	38	TMP013100	Transcription & Translation	102
TMP007653	Metabolism	42	TMP011994	Transcription & Translation	103
TMP011371	Metabolism	54	TMP010549	Transcription & Translation	104
TMP007388	Metabolism	58	TMP010821	Transcription & Translation	105
TMP012389	Metabolism	76	TMP011993	Transcription & Translation	106
TMP004201	Metabolism	84	TMP002587	Transcription & Translation	111
TMP005765	Metabolism	96	TMP004289	Transcription & Translation	112
TMP003059	Metabolism	99	TMP002357	Transcription & Translation	115
TMP005160	Metabolism	100	TMP007216	Transcription & Translation	124
TMP010987	Metabolism	107	TMP005243	Transcription & Translation	126
TMP008000	Metabolism	108	TMP007593	Transcription & Translation	129
TMP009552	Metabolism	114	TMP008384	Transcription & Translation	141
TMP007507	Metabolism	116	TMP008274	Transcription & Translation	146
TMP009871	Metabolism	120	TMP008329	Transport & Binding	26
TMP009102	Metabolism	125	TMP012391	Transport & Binding	56
TMP011448	Metabolism	128	TMP005827	Transport & Binding	67
TMP011421	Metabolism	132	TMP006242	Transport & Binding	69
TMP006617	Metabolism	139	TMP002103	Transport & Binding	98
TMP004413	Metabolism	140	TMP003369	Transport & Binding	143

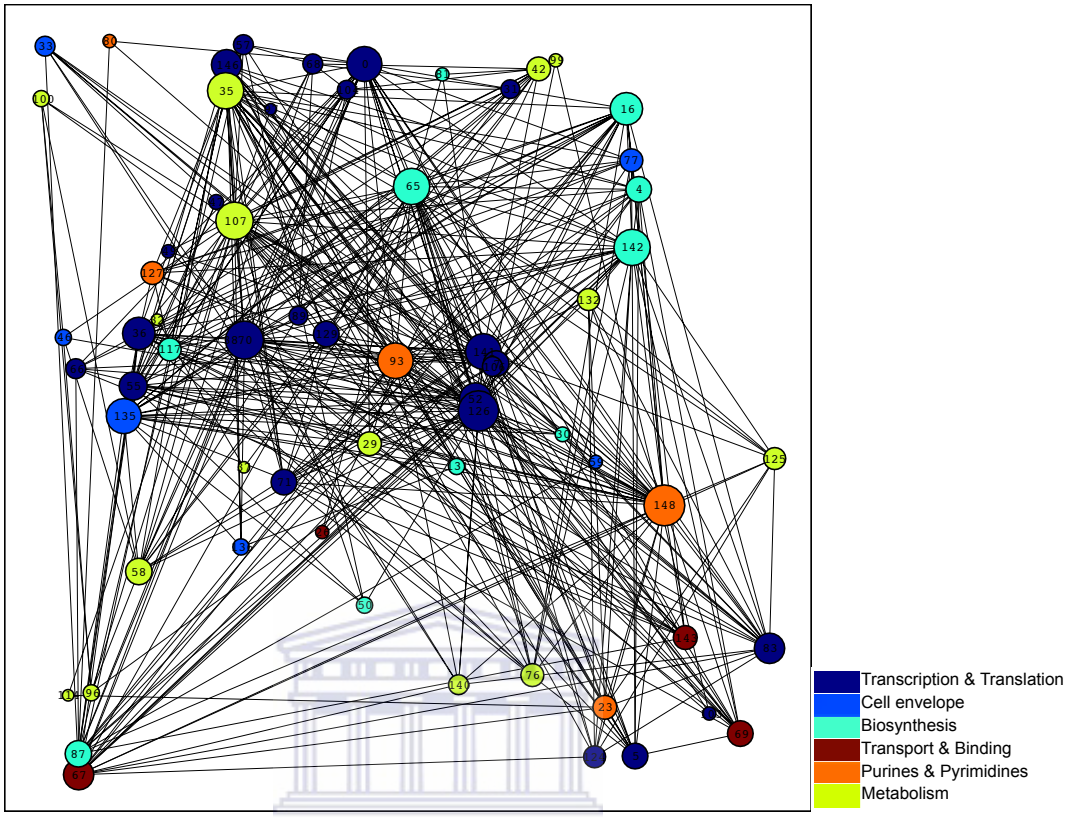


Figure 2.4: Network view of the motif clustering analysis. The larger circles represent genes with greater number of motifs shared with other *IRE*-regulated genes. Genes are color-coded based on the GO term assignments.

2.3.3 Functional assessment of the putative iron metabolism genes

To better understand the function of the putative *IRE*-regulated genes, their GO assignments from previous analysis were retrieved (Section 2.2.3). As a result the identified 150 *IRE*-regulated genes can be categorized into biosynthesis, cell envelope, metabolism, purines and pyrimidines, transcription and translation, and transport and binding (Figure 2.5a). Among these transcription and translation as well as metabolism are the over-represented functions. The biosynthesis and metabolism categories include components of amino acid biosynthesis, and biosynthesis of co-factors, as well as central intermediary metabolism, energy metabolism, and fatty acid metabolism respectively. Furthermore, based on the results obtained from ProtFun, 58.94% of the putative *IRE*-regulated genes are enzymes, which were further classified into ligase (25.84%), lyase (16.85%) and iso-

merase (11.23%) .

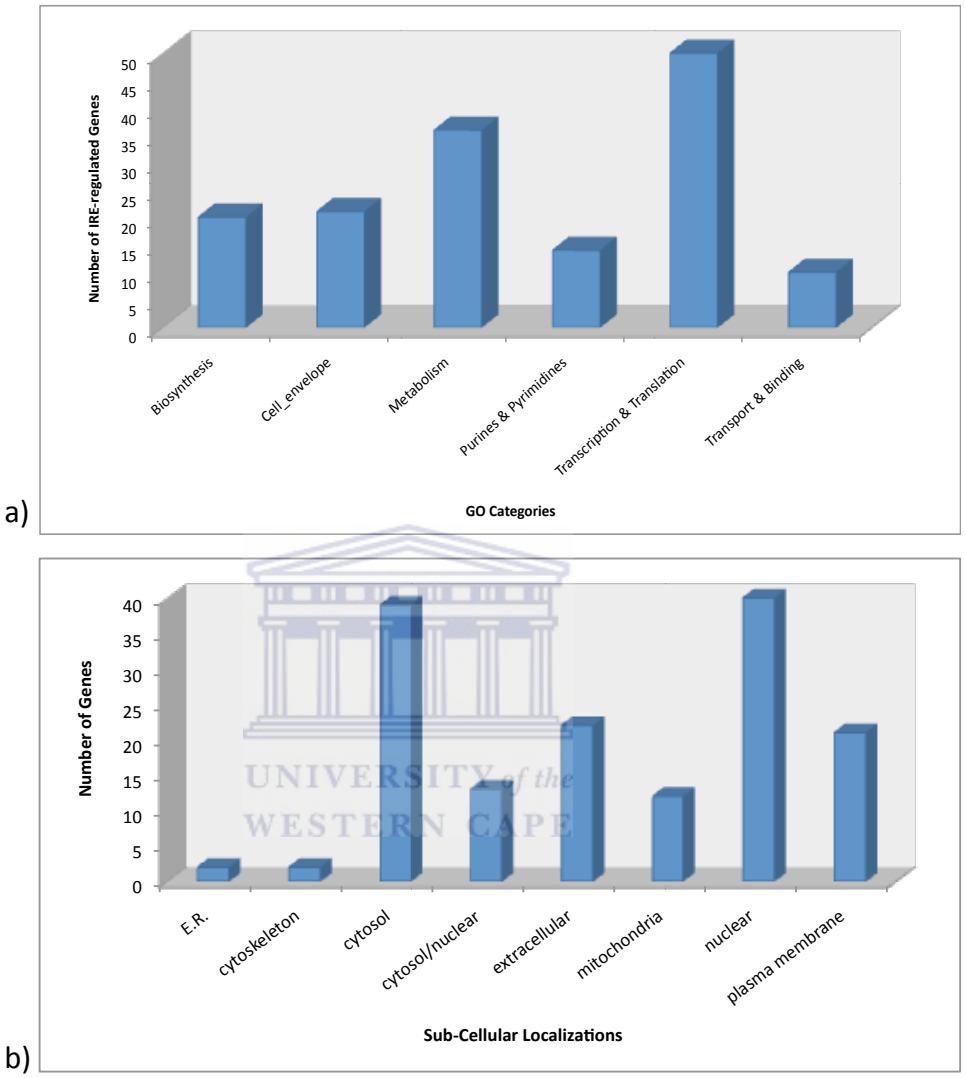


Figure 2.5: GO-category and sub-cellular localization assignments of *IRE*-regulated genes in *Glossina morsitans*. Most of the identified *IRE*-regulated genes are involved in transcription and translation as well as metabolism. Also b) shows the localization of these genes, indicating that they are mostly cytosolic and nuclear-localized.

Based on sub-cellular localization it is evident that most *IRE*-regulated genes reside in the nucleus (26.49%) and the cytosol (25.82%), while others are distributed between E.R., cytoskeleton, extracellular, mitochondria and plasma membrane (Figure 2.5b). It is notable that 8.60% of *IRE*-regulated genes have dual localizations in both the cytosol and the nucleus. In addressing the localization of iron-responsive genes, Ponka (1997) has suggested

the "kiss and run" theory, which assumes a direct transfer of iron from the *Tf*-containing endosome to the mitochondrion by passing through the cytosol (Ponka, 1997; Richardson and Milnes, 1997; Zhang *et al.*, 2011). However, details about mechanism driving the communication between the cytosol and the mitochondria remains to be elucidated. Recent studies have provided clues about the possible roles of myosin Vb (Provance *et al.*, 2004) and cytoskeletal regulatory molecules such as *MRCKalpha* (Iacopetta and Morgan, 1983) as well as vesticular docking, such as *Sec15l1* (Lim *et al.*, 2005; Zhang *et al.*, 2006) in facilitating kiss and run (Sheftel *et al.*, 2007). It is therefore possible that genes identified in this study as being *IRE*-regulated localized in the cytosol and the mitochondria are indeed the key drivers in maintaining mitochondrial and cellular iron metabolism.

Figure 2.6 is a schematic representation of *IRE*-regulated genes with different localizations in the cell. Several cell envelope, and transport and binding proteins were identified for which a possible role in iron trafficking may be inferred. The majority of these proteins have plasma membrane and extracellular localizations. Moreover, a novel function in iron metabolism was assigned to five cell envelope, and transport and binding genes with no previously defined biological function. Furthermore, several secretary signals were detected in these genes *i.e.* a total of 17, the majority of which are localized in the extracellular matrix. Furthermore, the results of this study suggested the enrichment of transcription and translation among the *IRE*-regulated genes, for which the localization analysis suggest an over-representation in the cytosol and the nucleus. Additionally, a relatively random distribution of other functional categories is observed across different cellular components.

Additionally, 17.21% of the identified *IRE*-regulated genes are secretary proteins, of which 69.23% are signal peptides and 30.76% are signal anchors. Among the predicted secretary proteins, the majority are localized in the extracellular matrix. Furthermore, 61.53% of the putative secretary proteins are cell envelope (Table 2.6). Secreted proteins play critical roles in several biological processes such as cellular immunity and communication, hence contribute to the discovery of novel biomarkers (Hathout, 2007; Xue *et al.*, 2008). Hence, the identified *IRE*-regulated secreted proteins present good candidates for future control

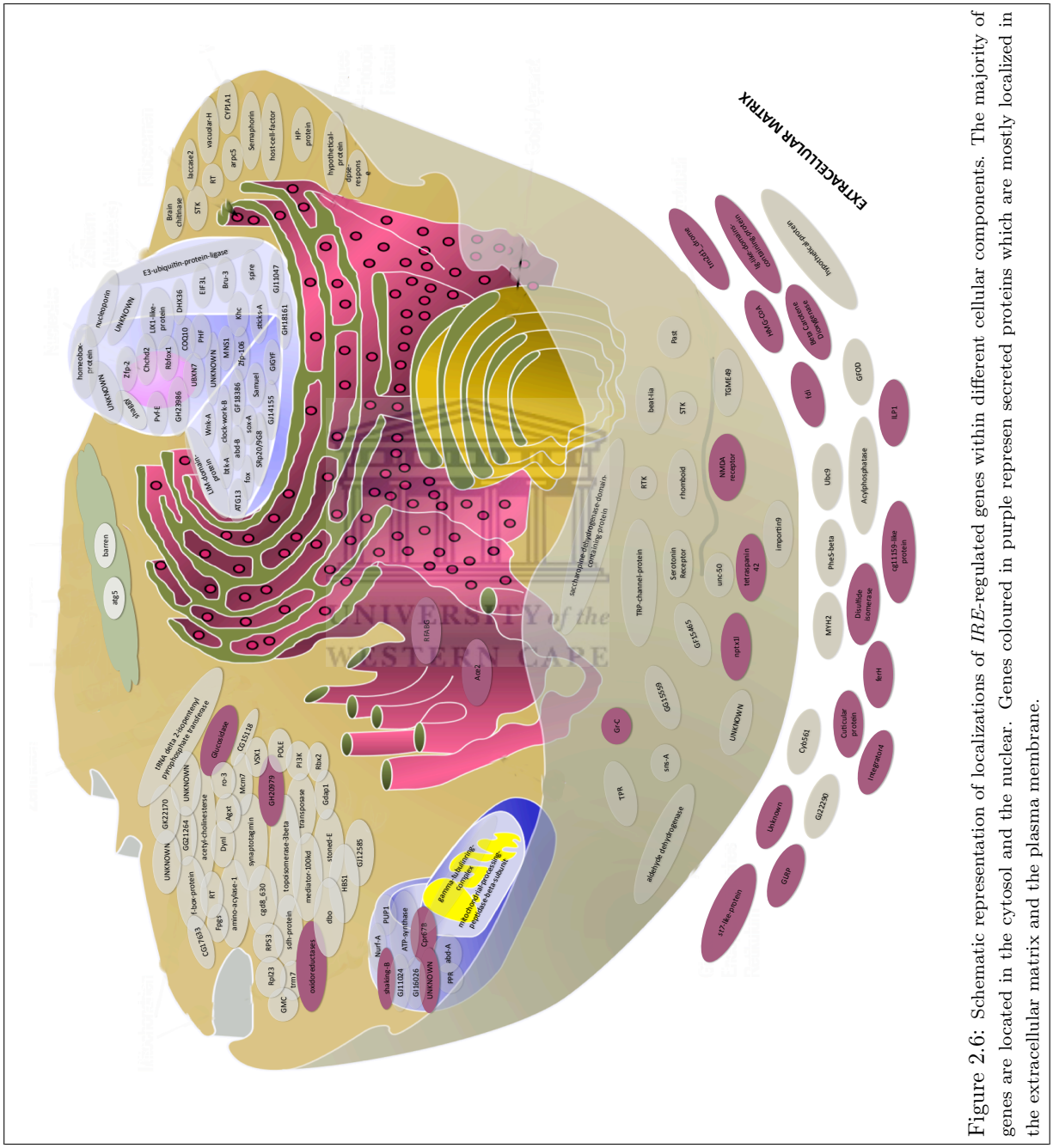


Figure 2.6: Schematic representation of localizations of *IRE*-regulated genes within different cellular components. The majority of genes are located in the cytosol and the nuclear. Genes coloured in purple represen secreted proteins which are mostly localized in the extracellular matrix and the plasma membrane.

strategies targeting iron-regulatory pathways in tsetse fly.

Table 2.6: A list of putative *IRE*-regulated secreted proteins identified in *Glossina morsitans*.

Secreted Proteins	GO	Localization
ShakingB	amino acid biosynthesis	Mitochondria
St7-like protein	Biosynthesis of Co-Factors	Extracellular
Cuticle-protein	Cell Envelop	Extracellular
Ig-like domain-containing protein	Cell Envelop	Extracellular
Beta-hexosaminidase-fdI	Cell Envelop	Extracellular
Beta-Carotene-dioxygenase	Cell Envelop	Extracellular
Gly-rich protein	Cell Envelop	Extracellular
Insulin-like peptide1	Cell Envelop	Extracellular
Integrator4	Cell Envelop	Extracellular
Unknown	Cell Envelop	Extracellular
Ferritin heavy chain	Cell Envelop	Extracellular
disulfide isomerase	Cell Envelop	Extracellular
Glucosidase	Cell Envelop	Cytosol
Cuticular protein 67B	Cell Envelop	Mitochondria
Angiotensin converting enzyme	Cell Envelop	E.R.
Retinoid and fatty acid-binding glycoprotein	Cell Envelop	E.R.
Tetraspanin42	Cell Envelop	Plasma membrane
Neuronal-pentraxin1	Cell Envelop	Plasma membrane
Glucose methanol choline oxidoreductase	Central intermediary metabolism	Cytosol
GH20979	Central intermediary metabolism	Cytosol
Hmg-coenzyme A-isoformA	Energy metabolism	Extracellular
tm2d1	Energy metabolism	Extracellular
Unknown	Energy metabolism	Mitochondria
Lyzosyme-domain-containing protein	Transport & Binding	Extracellular
NMDA-receptor glutamate-binding protein	Transport & Binding	Plasma membrane
Gustatory receptor isoformC	Transport & Binding	Plasma membrane

2.3.4 Tsetse's mechanism of iron sequestration and trafficking

Proliferation is an important aspect of infection (Bullen *et al.*, 2000), in which the role of iron is well-established (Ong *et al.*, 2006). The acquisition of iron is essential for the metabolic processes of the pathogen, that have allowed them to adopt mechanisms to acquire protein-bound iron. To combat pathogen survival and invasion, iron sequestration is an important part of innate immune response that is elegantly employed by insect vectors (Beck *et al.*, 2002). This is achieved through the function of several iron-binding proteins such as those identified in this study (Table 2.7). These include *beaten pathIIa*, *sticks and stones*, *Ig*-like domain containing protein, beta-carotene dioxygenase, *MRCK-alpha,* reverse transcriptase and defective proboscis extension response. The majority of these genes belong to the immunoglobulin superfamily with definitive roles in immune response. Studies in Drosophila and *C.elegans* have suggested the important role of immunoglobulin superfamily in multiple physiological systems (Vogel *et al.*, 2003; Teichmann and Chothia,

2000; Rougon and Hobert, 2003). Among the members of *IgSF,* the major histocompatibility complex class I and II molecules, proteins of the T cell receptor complex, virus receptors and cell surface glycoproteins can be named (Soroka *et al.*, 2008). A number of *IgSF* members are considered as biomarkers for cancer progression (Johnson *et al.*, 1997; Zeng *et al.*, 2011, 2012). Immunoglobulin superfamily proteins are known for their ability to mediate cell surface reception and pathogen recognition in insect vectors such as *Anopheles gambiae,* hence defending against bacteria and Plasmodium parasite (Dong *et al.*, 2006). Additionally, several lines of evidence point at the role of several *Ig*-super family members in regulating iron homeostasis, such as the mammalian neogenin (*IgSF* member), which interacts with hemojuvelin.

Table 2.7: *IRE*-regulated immune response and stress response genes.

Gene Name	GO Category	Sub-Cellular localization
Beaten path	Immune Response	plasma membrane
Sticks and stones	Immune Response	plasma membrane
Unknown	Immune Response	mitochondria
Ig-like domain containing protein	Immune Response	extracellular
Beta-carotene dioxygenase	Immune Response	extracellular
Unknown	Immune Response	extracellular
Serine-threonine protein kinase	Immune Response	cytoplasm, nucleus
Reverse transcriptase	Immune Response	cytoplasm, nucleus
Defective proboscis extension response	Immune Response	cytoplasm, nucleus
Ferritin heavy chain	Stress Response	extracellular
Glucose-fructose oxidoreductase	Stress Response	extracellular
Angiotensin-converting enzyme	Stress Response	E.R.
GH20979	Stress Response	cytoplasm
Ring-box protein 2	Stress Response	cytoplasm

Hence, the findings of this study pertaining to the *IRE*-regulated immune response genes, further points to the importance of these genes at the host-pathogen interface through the possible sequestration of iron from the invading pathogen. Furthermore, several stress-response genes were also identified to be *IRE*-regulated including the ferritin heavy chain, glucose fructose oxidoreductase, angiotensin-converting enzyme, GH20979 and ring box protein2. Considering the implications of iron overload in oxidative stress and the subsequent irreparable cellular damage (Touati, 2000), identifying *IRE*-regulated stress-response genes is expected. These genes could indeed play a role in iron-withholding response to deprive the invading pathogens of iron, hence protecting the insect vector. Though *IRE* regulation of ferritin heavy chain is well established (Hentze *et al.*, 1987) no

knowledge of such regulation exists on the other stress response genes identified in this study, which represent novel targets for future development of control strategies.

Genes responsible for the binding and trafficking of iron are vital for the survival of all organisms especially the insect vectors that are faced with the over-abundance of iron and the subsequent oxidative stress that may follow. To this end, the results of this study have proposed a number of genes with putative functions in iron binding, transport and storage. These include the several cell envelope as well as transport and binding proteins, the majority of which are localized in the extracellular environment and the plasma membrane respectively (Table. 2.8).

Table 2.8: Putative genes involved in iron trafficking in *Glossina morsitans*.

Gene Name	GO Category	Sub-Cellular Localization
Glucosidase	Cell envelop	Cytosol
Cytochrome-P450	Cell envelop	Cytosol/Nuclear
Angiotensin-converting enzyme	Cell envelop	E.R.
Retinoid & fatty acid glycoprotein	Cell envelop	E.R.
Cuticle protein	Cell envelop	Extracellular
Ig-like domain containing protein	Cell envelop	Extracellular
Beta-hexosaminidase-fdI	Cell envelop	Extracellular
Beta carotene dioxygenase	Cell envelop	Extracellular
Gly-rich protein	Cell envelop	Extracellular
Insulin-like peptide1	Cell envelop	Extracellular
Integrator4	Cell envelop	Extracellular
Unknown	Cell envelop	Extracellular
Ferritin heavy chain	Cell envelop	Extracellular
disulfide isomerase	Cell envelop	Extracellular
Hypothetical protein	Cell envelop	Extracellular
Glucose-fructose oxidoreductase	Cell envelop	Extracellular
Cuticular protein 67B	Cell envelop	Mitochondria
Hypothetical protein	Cell envelop	Mitochondria
Tetraspanin42	Cell envelop	Plasma membrane
Neuronal-pentraxin1	Cell envelop	Plasma membrane
Lyzosyme domain-containing protein	Transport & binding	Extracellular
Cytochrome B561	Transport & binding	Extracellular
Unknown	Transport & binding	Nuclear
Hypothetical protein	Transport & binding	Nuclear
NMDA-receptor glutamate binding	Transport & binding	Plasma membrane
Gustatory receptor isoformC	Transport & binding	Plasma membrane
Serine/threonine protein kinase	Transport & binding	Plasma membrane
Hydroxy-tryptamine-receptor	Transport & binding	Plasma membrane
Unc50	Transport & binding	Plasma membrane

Besides their importance in protecting the vulnerable insect from the toxic effects of iron overload, the identified genes can also function as antimicrobial peptides through with-

holding ferric ion from the invading pathogen *i.e.* trypanosomes. This is inline with the fact that iron trafficking genes are generally implicated in virulence and immunogenecity (Forbes and Gros, 2001), which are often surface located making them favorable for identifying preventative measures. Thus the result of this work may present novel targets that have adapted to meet the blood-feeding needs of tsetse fly to accommodate their unique reproduction biology and survival. One of the well-studied genes involved in iron storage and transport, which was also presented in our results is the ferritin heavy chain which was found not only to be confined to the cytosol but also to be present in the extracellular environments.

Additionally, to confirm and complement previous finding, the work presented here have recognized the implication of ferritin heavy chain in secretory pathways. As previously described by Nichol and Winzerling (2002) this observation though common to several insects, is in contrast to vertebrates ferritin heavy chain that is mainly cytosolic. Furthermore, despite the evident role of transferrin in iron transport among various species, lack of evidence on identifying an *IRE* stem loop in the UTRs of Glossina *Tf* in this study may point to the importance of ferritin heavy chain and the other proposed transporters in this study in tsetse fly. This could further imply *Tf*-regulation by mechanisms other than *IREs*. Additionally, the absence of *Tf*-receptor in the genome of Glossina, which is evolutionarily in line with the loss of the C-terminal lobe from insect-*Tfs* (Attardo *et al.*, 2006) further highlights the significance of the identified genes as alternative mechanisms by which iron regulation is maintained.

Additionally, several energy metabolism and mitochondrial genes were identified to be *IRE*-regulated (Table 2.9), which supports the role of iron in modulating the components of energy metabolism hence affecting ATP formation via oxidative phosphorylation (Lenaz *et al.*, 2002).

Table 2.9: *IRE*-regulated energy metabolism genes in *Glossina morsitans*.

Gene Name	Subcellular Localization
Hypothetical protein	plasma membrane
LIX1-like protein	nucleus
Bruno-3 transcript	nucleus
20s proteosome regulatory subunit beta	mitochondria
hmg coenzyme a	extracellular
foly polyglutamate synthase	cytosol
Serine-pyruvate mitochondrial	cytosol
Reverse transcriptase	cytosol
phenylalanyl-tRNA synthase subunit beta	extra cellular
tm2d1_drome	extra cellular
Unknown	mitochondria
serine/threonine-protein kinase	cytosol, Nucleus
coiled-coil-helix-coiled-coil-helix	nucleus
GJ11024	mitochondria
heterochromatin-associated protein	cytosol, Nucleus
hypothetical protein	cytosol
acetyl cholinesterse	cytosol
synaptotagmin	cytosol

2.4 Conclusion

The results of the genome-wide search for *IRE* stem loop structures led to the identification of 150 *IRE*-regulated genes among which, only two are known to have *IRE*s namely the ferritin heavy chain and *MRCK-alpha.* The rest are novel genes with putative roles in regulating iron levels in tsetse fly. To further assess the regulation of these genes their UTRs were analyzed for the presence of other motif elements besides *IRE*s. These results can give insight into the mechanisms of iron regulation in *Glossina morsitans*. More interestingly perhaps is the observation that points to the identification of a number of elements with roles in regulating reproductive genes. This is in line with the fact that iron is an indispensable component of reproduction and the sole reason behind the blood-feeding habit exhibited by these insects. The findings of our work, further adds complexity to our current understanding of iron regulation by proposing a putative role of miRNAs in regulating iron metabolism in concordance with the regulatory effects of *IRE*s. These findings were derived at through the identification of conserved motif elements in the UTRs of *IRE*-regulated genes that function as binding sites for miRNAs. Additionally,

several putative iron metabolism genes with no known functions were assigned to functional categories through the assessment of their motif sharing with elements of known functions. In conclusion the identification of elements in the UTRs of *IRE*-regulated genes have allowed for the inference of a co-regulatory network between iron metabolism and transcription/translation regulation, reproduction, amino acid biosynthesis, inflammation and immune response and genes involved in signaling pathways. Such intricate interplay was derived based on the shared elements among *IRE*-regulated genes, which further supports their previously defined functional classifications. Additionally the identified *IRE*-regulated genes were assigned to GO-categories, in order to gain insight into their function. As a result, several putative iron trafficking genes were proposed. Moreover, several transcription and translation regulation genes were identified, with a probable role in regulating genes with downstream effects in maintaining iron homeostasis.

The presence of these genes in other vertebrates may further be assessed to derive at a set of genes that function as regulators of iron metabolism across all species. Since, the identification of *IRE*s in the UTRs of ferritin and *TfR*s, several other *IRE*-regulated mRNAs have been identified in this study, while many more remain to be characterized. Such identification and characterization in a wet-lab setting is time consuming and costly. Hence, the computational approach provides a ground from which experimental validation can be undertaken. The sequenced genome of *Glossina morsitans* have allowed for such analysis, the result of which is summarized as a set of *IRE*-regulated genes, many of which have no known implications in iron metabolism. As a result of this study, it is evident that the iron metabolism network extends beyond transport and storage, with an extensive interplay between genes involved in transcription/translation regulation, energy metabolism, stress and immune response, and signal transduction.

Chapter 3

Validating the role of *IRE*-regulated genes using microarray data

Abstract

Background:

Our findings in Chapter 2 highlight the role of a set of putative *IRE*-regulated genes with no previously known implications in iron metabolism. Evidence of gene expression for these putative *IRE*-regulated genes in response to a blood meal will add further support for a role in iron regulation and leads to the identification of key genes whose timely expression drives the network of events that prevent iron-induced toxicity. The published work of Marinotti *et al.* (2006), in profiling gene expression in response to a blood meal at seven different time points using the Affymetrix chip, provides a good knowledgebase, to assess the expression patterns of 150 putative *IRE* regulated genes. A re-analysis of the microarray data was performed with a fold change > 1.0 and p-value < 0.05.

Methods:

The homologs of the 150 putative *IRE*-regulated genes were identified in Anopheles and those with probe IDs on the Affymetrix gene chip were further analyzed. Several Glossina *IRE*-regulated genes were not present in the reported set of significantly expressed transcripts presented by Marinotti *et al.* (2006), due to their stringent selection criteria (*i.e.* p-value < 0.001). The re-analysis of the data was carried out using the Genespring software and included pre-

processing and normalization of the data, and the t-test analysis. The selection criteria were based on the p-value of less than 0.05 and the fold change of greater than one.

Results:

A total of 103 Glossina putative *IRE*-regulated genes were identified with a homolog in Anopheles having a probe ID on the Anopheles Gene Chip. Of the 103 *IRE*-regulated homologous genes 80 were significantly upregulated, while 23 were significantly down-regulated following the blood meal. Transcription and translation were enriched as functional categories in the differentially expressed genes. The assessment of the expression patterns of these genes led to the identification of several early response and late responses among which, the expression of nine cell envelope and transport genes were evident. Seven of the nine genes are signal peptides.

Conclusion:

Based on the Anopheles gene expression dataset, further support for a role in iron-metabolism was added to the 103 putative *IRE*-regulated genes. The remainder 47 genes may either have no probe ID on the Anopheles Gene Chip or may be considered as Glossina-specific genes involved in maintaining iron homeostasis. Furthermore, it is plausible to speculate that genes identified in this study are early and late responders to iron from the blood meal, hence representing candidates for targeting different stages of insect biology in the context of iron metabolism and assisting in the development of novel vector control strategies.

3.1 Introduction

Trypanosomes and all other pathogens alike, may be considered as scavengers of host's iron resources. However, in fulfilling their need the trypanosomes are faced with tightly regulated "iron-withholding" strategies set in place by the insect vector as an integral part of the innate immune response (Marx, 2002). It has been shown that approximately 40% and more of tsetse flies challenged with trypanosome-infected blood can survive parasite invasion and self-cure (Lehane *et al.*, 2008). However, no detailed knowledge exists about the mechanism by which the insect becomes refractory to trypanosomes. Deriving at the exact differences between the refractory and the infected flies is a major challenge. This allows for speculations about the possible role of iron sequestration as part of an immune response by the insect in preventing the pathogen from proliferating and the subsequent invasion of the salivary gland. If so, identification of genes responsible for

iron binding, transport and storage, and the further profiling of their expression can give important clues about Tsetse's mechanisms of refractoriness and survival. Furthermore, reactive oxygen species (ROS) are known to be used by insects against pathogens. As in other insects the ROS are implicated in tsetse response to infection (Hao *et al.*, 2003; Lehane *et al.*, 2003; Munks *et al.*, 2005; MacLeod *et al.*, 2007). In line with this, an increased prevalence of trypanosomes have been observed when treating tsetse flies with antioxidants (Maudlin and Welburn, 1987; MacLeod *et al.*, 2007). Implications of iron in the production of ROS is well-established, which further supports the possible role of iron metabolism genes in tsetse's vectorial competence. Besides the implications of iron at the host pathogen interface, this trace metal is of essence for the reproductive success and survival of all insect vectors including tsetses. To this end these insects have adapted mechanisms, by which the toxic effects of iron overload can be impeded. This is dependent on the timely expression of critical genes that allow for digestion of blood meal iron, its direction to eggs and ovaries as well as other organs, and the excretion of the excess amount. Thus assessment of the expression profile of genes in response to iron could provide compelling evidence about the nature of iron regulation in hematophagous insects. Tsetse flies reproduce viviparously with an unusually low reproductive capabilities, thus reducing fecundity can have profound impact on vector control. Even more interesting is the fact that both male and female tsetses are hematophagous, which further signifies the role of the iron-metabolism genes as targets for the control of the tsetse population (males and females) (Aksoy *et al.*, 2001).

To date, the perception of iron metabolism in insects has mainly been that a few major proteins including ferritin and transferrin are responsible for the binding and the sequestration of iron (Dunkov *et al.*, 1995; Yoshiga *et al.*, 1997). However, our findings in previous chapters add complexity to the current knowledge by introducing several distinctive iron-binding proteins that have possibly evolved to meet tsetse's need of blood meal with iron as its major constituent. To characterize the mechanism by which, putative *IRE*-regulated genes may induce a response to blood meal iron, the knowledge of gene expression would be of essence. Despite the work of Lehane *et al.* (2003) on the expression analysis of immune response genes in tsetse fly, current literature is devoid of

such knowledge about iron-responsive genes. In the absence of blood-induced microarray data for *Glossina morsitans*, the well-established database for *Anopheles gambiae* represents a good alternative. Though levels of gene expression conservation may vary between orthologs, this analysis helps derive at a very broad picture about the events that may occur following a blood meal, under the assumption that both blood-feeding insects *i*.e. Anopheles and Glossina share similarities in their mechanisms of iron metabolism and regulation. Albeit Drosophila is more closely related to Glossina with the last common ancestor (LCA) of ˜ 47.9 Mya, making it an important insect for the comparative assessment and understanding of tsetse fly. However, the non-blood feeding nature of Drosophila limits our ability to infer blood meal-induced iron-responsive expression for the orthologs of Glossina IRE-regulated genes. As a result, Anopheles (LCA $>$ 200 mya) represents a good alternative. Hence, considering the predictive nature of these findings future research in this area will help shed light on the clarity of our results. Here, we aim at assessing the expression of the predicted *IRE*-regulated genes of Glossina in Anopheles in response to blood meal whereby, a putative map of events occurring from 3H-15D post blood meal will be inferred for *Glossina morsitans*.

3.2 Materials and Methods

3.2.1 Data retrieval

Gene Chip Plasmodium/Anopheles genome array (Affymetrix) including probe sets to approximately 14906 *A. gambiae* genes. The raw intensity data retrieved from Gene Expression Omnibus (GEO) (`http://www.ncbi.nlm.nih.gov/geo/`) was used to evaluate blood meal induced gene expression in adult female mosquitoes. Furthermore, the homologs of putative *IRE*-regulated genes in Anopheles were acquired from previous analysis described in Chapter 2 (Section 2.2.4).

3.2.2 Pre-processing and statistical analysis

Despite existing literature on the analysis of Anopheles genome array (Marinotti *et al.*, 2006), many putative *IRE*-regulated genes of Glossina are not present in the reported set of significantly expressed transcripts following a blood meal, hence to capture changes in the expression of these genes, the reanalysis of the data where the selection criteria is based on the p value of < 0.05 and the fold change of $>=1.0$ was carried out. Though it is plausible to argue that a higher P value may influence the number of false positives, it is important to note that integrating the power of both the P value and the fold change may help reduce this number, while providing a more biologically meaningful assessment of the results. Also, if the predicted IRE-regulated genes are indeed responsive to iron it would be expected that the expression of these genes are altered in iron-induced samples. Considering that such alterations may be very small, using a very stringent threshold for the log2 fold change may negatively affect the outcome and result in some of the data to be eliminated. Hence we have modified our criteria to accommodate for genes with more subtle response to iron.

As part of pre-processing steps data summarization, log2 transformation, and baseline transformation were carried out using the Genespring 5.1 software (Silicon Genetics, CA). Average signal intensities were then normalized using the Loess curve for intensity dependent normalization followed by a per gene median normalization. Quality control of the samples was carried out, using principle component analysis (PCA), which calculates the PCA scores and represents them in a 3D scatter plot. Furthermore, fold change analysis

was carried out by comparing experimental conditions (blood-fed samples) to control conditions (non-blood-fed samples), using parametric approach (t-test) whereby probe sets that satisfy a p value of < 0.05 and a fold change of $>=1.0$ in at least one condition are accepted as up-regulated genes. To organize putative *IRE*-regulated genes into clusters based on the similarity of their expression profile the Hierarchical clustering algorithm was used as implemented in Genespring with the distance matrix set to euclidean and the linkage rule set to median.

3.3 Results and Discussions

3.3.1 Identification of *Anopheles gambiae* iron-regulated transcripts

The availability of blood-induced microarray expression data in *Glossina morsitans* would present a compelling knowledge source to better understand iron metabolism in this blood-feeding insect. However, in the absence of such knowledge the well-established database of microarray expression data for *Anopheles gambiae* represents a good alternative. The integrated knowledge of *IRE*-regulated genes in *Glossina morsitans* and the expression profile of *Anopheles gambiae* genes in response to blood meal allowed for drawing a putative map of iron metabolism events.

136 of 150 putative Glossina *IRE*-regulated genes had a blast match to *Anopheles gambiae*, while 16 genes were Glossina-specific. Of the 136 genes 103 have Anopheles homologs with existing probe IDs on the Anopheles Gene Chip, which were selected for further analysis. The global patterns of greater than one fold up-regulation or down-regulation of genes with P-values lower than 0.05 were established by comparing transcript levels in adult *A. gambiae* at six time points post ingestion of blood meal to the levels in the non-blood-fed mosquitoes. Despite existing literature on the genome-wide analysis of gene expression in adult *Anopheles gambiae* (Marinotti *et al.*, 2006), the results of this study through identifying putative *IRE*-regulated genes have allowed for an in depth assessment of the iron-induced expression of these genes following the blood meal.

3.3.2 Immediate response to blood meal iron

The effect of blood meal iron on mosquitoes can be summarized as a series of physiological and morphological changes that take place from the early hours to several days following blood meal. Detailed analysis of variations in gene expression at 3H post-blood meal revealed a differential expression for 59 out of 103 Anopheles homologs to the putative *IRE*-regulated Glossina genes. Among these, 35 showed an increased expression while 24 were significantly downregulated. The most commonly represented genes at 3H belong

to transcription and translation as well as biosynthesis of nucleotide, amino acids and co-factors. Furthermore, two putative iron transport genes namely the gustatory receptor isoform c and ferritin heavy chain were upregulated at 3H. Additionally, the expression of several cell envelope, central intermediary, and energy metabolism genes are evident. Of the 35 genes significantly upregulated at this time point nine may be considered as early response genes, due to their 3H-specific expression (Table 3.1). This gene set comprises of six enzymes and four putative signal peptides. At 3H, the blood meal is in preparation for digestion, which allows for elucidating an association of genes explicitly expressed at this time point, with signals responsible for the digestion of heme-derived and transferrin-bound iron from the blood meal.

Table 3.1: **Early response genes to iron overload from the blood meal in Anopheles, extrapolated for *Glossina morsitans*.** Four signal peptides are present among these genes which points to their importance as both regulators of iron and putative targets for developing control strategies.

Gene ID	Gene Name	GO	Type
TMP005335	Zinc carboxypeptidase a1	Amino_acid_biosynthesis	Non-secretory
TMP012491	Amino-acylase-1	Translation	Non-secretory
TMP002821	GMC oxidoreductase	Central_intermediary_metabolism	Signal Peptide
TMP002406	Cuticle protein	Cell_envelope	Signal Peptide
TMP004987	Cuticular protein 67B	Cell_envelope	Signal Peptide
TMP005160	Homeobox protein	Central_intermediary_metabolism	Non-secretory
TMP005357	LIX1-like protein	Energy_metabolism	Non-secretory
TMP012412	Sticks and stones isoform a	Amino_acid_biosynthesis	Non-secretory
TMP008329	Gustatory receptor isoformc	Transport_and_binding	Signal Peptide

Among these, the expression of cuticular proteins as cell envelope proteins as well as one transport protein i.e. the gustatory receptor isoform c are evident. Gustatory receptors play crucial roles in detecting food, selecting sites to lay eggs, and recognize mates, through recognition of specific stimuli. The putative up-regulation of this gene in an *IRE*-regulated manner suggests a role for the gustatory receptor in sensing iron in blood meal, hence inducing downstream events. Bahadorani and Hilliker (2009), through investigating the effect of iron and other metals in *D. melanogaster* gustation have proposed the involvement of fly's gustatory system in behavioral responses to high dosage iron in maintaining fly survival and reproductive success.

Thus, these results further add values to our previous findings (Chapter 2) pertaining to *IRE*-regulated genes, by providing a possible reflection of iron metabolism events in Glossina at 3H PBM. Furthermore, the nine early-response genes proposed here represent putative targets for development of control strategies, as functional disruption of these genes may halt the digestion, absorption and later distribution of iron to necessary organs, hence affect insect's viability.

3.3.3 Late response to blood meal iron

At 24H the early response genes are no longer expressed, which is possibly due to the fact that these genes need to be quickly restored to the pre-blood fed levels allowing for an efficient acquiring of the next blood meal. From 24H to 15D post blood meal the expression of several transcripts are evident, which we will refer to as late response genes. These include 35 out of 103 putative *IRE*-regulated genes identified in Glossina that were mapped to Anopheles expression data (Table 3.2). Several GO terms are evident among the late response genes including components of translation, transcription and regulation, energy metabolism, amino acid biosynthesis, cell envelope and central intermediary metabolism. A single transporter gene, *Unc-50* was expressed throughout 24H-15D. *Unc-50* is a conserved integral Golgi membrane protein involved in intracellular trafficking. The blood meal-induced expression of this gene in an *IRE*-regulated manner allows for speculating a role for this gene in iron trafficking in blood feeding insects.

Among the late response genes five are putative signal peptides namely angiotensin-converting enzyme, insulin-like peptide1, integrator4, glucosidase, *and tm2d1.* This provides compelling evidence about the importance of these genes as targets for control strategies. Furthermore, the majority of the late response genes are located in the cytoplasm and the nucleus. Additionally the pattern of expression is similar at 24H, 48H and 72H following blood meal, with the number of expressed genes, increasing to 87, 92 and 89 respectively compared to genes expressed at 3H (Appendix C). This increase in the number of *IRE*-regulated genes may be explained by the fact that with the progression of digestion between 24-72H, the ingested iron needs to be absorbed, transferred to ovaries and eggs, while the excess iron needs to be excreted.

Table 3.2: **Late response genes to iron overload from the blood meal.** These constitute of five signal peptides and a single transporter gene among other responses, expressed from 24H-15D.

Gene ID	Gene Name	GO	Type
TMP005511	atg5	Amino_acid_biosynthesis	Non-secretory
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis	Non-secretory
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism	Non-secretory
TMP013886	trna(uridine-2-o-)-methyl transferase trm7	Central_intermediary_metabolism	Non-secretory
TMP012772	reverse transcriptase	Energy_metabolism	Non-secretory
TMP014030	GJ12585	Replication_and_transcription	Non-secretory
TMP008274	CG15118	Translation	Non-secretory
TMP003060	GK22170	Translation	Non-secretory
TMP006826	ring-box protein2	Translation	Non-secretory
TMP008943	Glucosidase	Cell_envelope	Signal
TMP012387	cytochrome P450	Cell_envelope	Non-secretory
TMP007137	hypothetical conserved protein	Regulatory_functions	Non-secretory
TMP003617	host cell factor	Translation	Non-secretory
TMP002863	angiotensin-converting enzyme	Cell_envelope	Signal
TMP004746	myosin heavy chain type II	Amino_acid_biosynthesis	Non-secretory
TMP006019	insulin-like peptide 1	Cell_envelope	Signal
TMP007537	integrator 4	Cell_envelope	Signal
TMP011421	tm2d1_drome	Energy_metabolism	Signal
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis	Non-secretory
TMP006377	coenzyme Q-binding protein	Amino_acid_biosynthesis	Non-secretory
TMP007653	Abdominal-B protein	Central_intermediary_metabolism	Non-secretory
TMP013948	Chch-domain-containing protein2	Energy_metabolism	Non-secretory
TMP002520	nucleoporin	Purines_and_pyrimidines	Non-secretory
TMP004047	GJ14155	Purines_and_pyrimidines	Non-secretory
TMP002718	shaggy	Regulatory_functions	Non-secretory
TMP004121	zinc finger protein2	Regulatory_functions	Non-secretory
TMP007216	kinesin heavy chain	Regulatory_functions	Non-secretory
TMP012225	phd finger protein	Regulatory_functions	Non-secretory
TMP013958	GYF domain-containing protein	Translation	Non-secretory
TMP004413	Wnk,isoformA	Central_intermediary_metabolism	Non-secretory
TMP002587	alternative splicing factor SRP20/9G8	Translation	Non-secretory
TMP008448	autophagy-related protein	Translation	Non-secretory
TMP013100	sox box protein isoform a	Translation	Non-secretory
TMP004517	importin9	Purines_and_pyrimidines	Non-secretory
TMP010016	tpr repeat-containing protein	Purines_and_pyrimidines	Non-secretory
TMP014133	unc-50	Transport_and_binding	Non-secretory

The number of expressed genes associated with transcription and translation reached its peak at these time points. Furthermore, several previously up-regulated energy metabolism genes such as acetyl cholinesterase, *LIX1*-like protein, and serine-pyruvate mitochondrial were significantly down-regulated at 24H, 48H, and 72H post blood meal (PBM). Additionally the previously expressed cuticular protein belonging to the cell envelope category was down-regulated at these time points.

The findings of this study regarding the expression of early and late response genes post blood meal are to a great extent inline with the results of Ribeiro (2003), who has

presented a catalogue of Anopheles transcripts significantly expressed following a blood meal, however in this study we add further evidence for their roles in iron metabolism. we have further proposed an iron-responsive expression for these genes. Among the putative *IRE*-regulated genes in this study, three have previously been shown by Lehane *et al.* (2008) to express in response to trypanosome-infected blood meal in Glossina. These include the antigen 5-related proteins, ATP synthase, and 60S ribosomal protein. It has been suggested that following trypanosoma infection, the self-cured flies induce an oxidative stress response (Lehane *et al.*, 2008). Thus these may indeed be considered as genes functioning at the host-pathogen interface to fight infection either through iron sequestration or the production of reactive oxygen species (ROS) and imperative targets for future studies. Though several genes were commonly identified between this study and Lehane's work, the latter mainly assesses the expression of genes within the fat body, while the results presented here may be considered a snapshot of events that control iron-metabolism at the genome-wide scale.

Besides the above-mentioned genes, the implications of some of the *IRE*-regulated genes in immune response against trypanosomes have previously been demonstrated (Attardo *et al.*, 2006). Among the fat body genes identified by Attardo *et al.* (2006) in response to trypanosome infection, shaggy protein (protein kinase) was identified in this work, as being *IRE*-regulated, which further points to the implication of this gene in the fight against the pathogen. Additionally, despite the identification of several known iron metabolism genes through analyzing the fat body transcriptome of *Glossina morsitans* (Attardo *et al.*, 2006) namely transferrin, ferritin, and divalent metal transporter (*Nramp1*), only ferritin heavy chain was identified in this study to be *IRE*-regulated. This either suggests mechanisms other than *IRE*-regulation for transferrin and *Nramp1* in Glossina or perhaps possible *IRE*s in their UTRs do not resemble any of the canonical or SELEX-based patterns. Given the absence of transferrin receptors (*TfR*s) in insects to date, it is probable that modifications have occurred in the mechanism of regulation of transferrin. Furthermore, it is possible that other genes identified in this study may have evolved in Glossina, such that they compensate for the absence of *IRE* regulatory elements in the UTRs of transferrin.

3.3.4 Constitutively expressed *IRE*-regulated genes

Furthermore, a number of *IRE*-regulated genes were significantly up- or down-regulated during all time-points. To gain insight into possible roles for these genes in iron metabolism a brief review of their functions will be provided. One of the proteins with increased expression post blood meal is the *Kelch-like* protein belonging to the nucleotide biosynthesis category. *Kelch-like* proteins function as portable substrate-specific adaptors of *E3*-ubiquitin-protein ligase complex. *Kelch repeats* are proposed to be involved in protein-protein interactions for circadian clock related substrate for ubiquitination (Yasuhara *et al.*, 2004). Many biological processes are dependent and controlled by selective ubiquitin-based degradation of key substrate (Pintard *et al.*, 2003). Furthermore, several *kelch-like* proteins have also been shown to be involved in the organization of actin and microtubule dynamic (Robinson and Cooley, 1997). Additionally, as part of the intricate iron regulatory mechanism, the ubiquitination system has been recognized, such as the degradation of iron-regulatory protein2 (*IRP*2) in iron-replete cells (Vashisht *et al.*, 2009). It has been proposed that *IRP*1 under *ISC* biogenesis impairment is sensitized for iron-dependent degradation via ubiquitin-protease pathway (Wang *et al.*, 2007). These evidence further support the significant implication of the ubiquitin-proteasome pathway in iron homeostasis. Besides the *kelch-like* proteins several other proteins were significantly up-regulated at 3H-15D post blood meal, including ubiquitin-conjugating enzyme 9 and *E3*-ubiquitin protein ligase.

Another protein shown to be significantly expressed throughout all the time points is Serine-threonine protein kinase, which is an enzyme involved in the phosphorylation of the OH group of serine or threonine. Receptors of this enzyme play roles in the regulation of cell proliferation, apoptosis, cell differentiation and embryonic development (Cross *et al.*, 2000; Roskoski, 2010). Several lines of evidence have suggested a role for serine/threonine protein kinases in iron metabolism (Corradini *et al.*, 2011; Ganz *et al.*, 2012; Knutson, 2010). Myotonic dystrophy kinase-related *Cdc42*-binding kinaseα (*MRCKa*) is a Serine-threonine kinase that regulates actin/myosin assembly. However, this protein is predicted as an *IRE*-regulated protein, which provide clues to its role in iron metabolism. Cmejla and colleagues (2010) have further provided support for iron-regulated expression of

MRCKa, which has also been implicated in transferrin iron uptake, through the regulation of *TF-TFR* endocytosis/endosome trafficking. Another *IRE*-regulated gene identified in this study and proved to be significantly expressed at all the studied time points is the DNA replication-licensing factor *MCM7*. This protein is critical for the initiation of eukaryotic genome replication. *MCM* proteins play essential roles in DNA replication and cell proliferation. Studies have identified *MCM* proteins as negative regulators of hypoxia inducible factor a protein known to be *IRE*-regulated. Hypoxia results in decreased proliferation through the functioning of *HIF-1*. *MCM7* is known to promote *HIF-1a* ubiquitination and proteosomal degradation. Hypoxia inducible factor 1 (*HIF*-1) controls the expression of genes involved in hypoxia induced adaptive responses (Hubbi *et al.*, 2011). *HIF-1* also mediates cell cycle arrest in response to hypoxia hence the increased *HIF-1* levels are associated with diminished cell proliferation (Biswas *et al.*, 2010). Under iron deprivation *HIF-1a* has been shown to stimulate *TFR* gene transcription (Bianchi *et al.*, 1999).

Suppressor of tumorigenicity protein 7 is another protein being unregulated in response to blood meal at several time points. *TS7* is considered a Tumor Suppressor Gene (*TSG*), involved in the development of certain types of cancers. Considering that iron overload is associated with carcinogenesis in human a support for the possible role of *TS7* in iron metabolism may be drawn. Another gene presenting blood-meal-induced up-regulation is the elongation factor 1α that is involved in the GTP-dependent binding of aminoacyl tRNAs to the ribosome (Maroni, 1993). The elongation phase of eukaryotic protein systems is regulated by phosphorylation, which is a critical step in maintaining a balanced rate of protein synthesis under several conditions such as mitosis, nutrient starvation, stimulation with insulin, growth factors or mitogens (Archambaud *et al.*, 2005; Browne and Proud, 2002). Various lines of evidence support the stimulation-induced phosphorylation-based regulation of eukaryotic elongation factors, while several protein kinases have been identified to be involved in this process. Besides its general role in translation *eEF1a* plays a critical role in nuclear export of proteins (Khacho *et al.*, 2008). Additional functions have been suggested for this gene such as the control of phosphotidylinasitol signaling; actin remodeling, cell motility and oncogenic activities, import of tRNAs into mitochondria

and nuclear export of proteins (Khacho *et al.*, 2008; Panasyuk *et al.*, 2008; Bouzaidi-Tiali *et al.*, 2007). Furthermore, *EF1A* has been shown to bind to cytoskeleton proteins and mediate their interactions (Gross and Kinzy, 2005, 2007). However, no exact evidence exists to date supporting the role of this gene in iron regulation and metabolism. As such the result of this study and the presence of *IRE* in the UTR of Glossina *EF1A* allows for speculations about its putative role in the maintenance of iron homeostasis, while the exact mechanism remains to be elucidated.

Spire is another significantly expressed gene in response to blood meal that is proposed to be involved in iron metabolism. *Spire* is an actin nucleation factor, which is critical for several biological processes and molecular functions. Actin cytoskeleton is essential for processes such as shape determination and intracellular transport. Drosophila *Spire* has been shown to be required for oocyte and embryo development as well; processes known to be iron-dependent (Quinlan *et al.*, 2005). Additional upregulated gene in response to blood meal is the Barren, which is a non-Structural Maintenance of Chromosome (*SMC*) subunit of condensin involved in chromosome condensation (Somma *et al.*, 2003). Eukaryotic chromosome condensation requires condensin that is a 13S multi-protein complex (Kimura *et al.*, 2001). Studies have suggested a role for Barren protein as an activator of DNA topoisomerase II (Lupo *et al.*, 2001). Lavoie and colleagues have shown that Barren plays a critical role during cell cycle when chromosome condensation occurs (Lavoie *et al.*, 2000). The importance of iron in cell cycle and mitotic process is well established. Hence, the speculation for the involvement of Barren protein in iron metabolism is plausible.

Furthermore, Gamma-tubulin A, a member of globular proteins family is another significantly expressed gene in response to blood meal. Gamma-tubulin A is important in the nucleation and polar orientation of microtubulin. Microtubule (MT) cytoskeleton is involved in cell motility, trafficking, signal transduction as well as apoptosis and cell division (Kelleher and Titus, 1998; Gundersen and Cook, 1999; Small *et al.*, 1999). The results of this study suggest that the up-regulation of Gamma-tubulin A expression could be iron-induced, however the mechanism by which this may occur or the downstream effect of this over-expression in maintaining iron-homeostasis is not well-understood.

Eukaryotic translation initiation factor 3 (*eIF3*), a multi-protein complex that binds to the 40S-ribosome, hence allow the maintenance of 40S and 60S ribosomal subunits dissociated was also identified to be blood-induced and *IRE*-regulated. It has been suggested that *eIF3* is involved in the formation of 40S initiation complex. *EIf* is responsible for efficient initiation and regulation of mRNA translation. Regulated mRNA translation is an important aspect of gene expression and contributes to diverse biological processes in various cell types. Studies have shown enhanced levels of *eIF3* in response to hepatic iron load (Liu *et al.*, 2012). Furthermore, in *IRE*-regulated mRNAs the binding of *IRE*-binding proteins (*IRP*s) in the 5´-UTR of *e.g.* ferritin have been shown to prevent the binding of the 43S pre-initiation complex (consisting of 40S-ribosomal subunit *eIF3*, and *eIF2*, GTP and tRNA) with the cap-binding complex *eIF4F* (Muckenthaler *et al.*, 2003; Mazumder *et al.*, 2003). However, it is considered that *IRE*/*IRP* binding does not impede the interaction of initiation factors that bind before the 40S-ribosomal subunit (Wilkie *et al.*, 2003). The presence of *IRE* in the UTR of *eIF3*, allows for the inference of an iron-regulated mechanism for this gene. In line with this finding, Theurl and colleagues have suggested an increased expression of *eIF3* mRNA in response to iron treatment of HepG2 cells (Theurl *et al.*, 2004). This further supports the expression patterns observed in Anopheles following blood meal whereby *eIF3* expression was up-regulated at all experimental time points PBM. This up-regulation in turn results in efficient expression of mRNAs that could possibly be involved in the iron metabolism pathway.

GH18161-like protein is another identified *IRE*-regulated mRNA that was significantly expressed post-blood meal. The exact function of this gene is not fully known, however it contains RNA polymerase *Rpb4* domain, which is characteristic of *Rpb4* proteins. *Rpb4* is an essential subunit of RNA polymerase III (Siaut *et al.*, 2003). Furthermore, *Rpb4* subunit has been shown to contribute to co-transcriptional recruitment of 3'-UTR processing factors (Runner *et al.*, 2008). RNA polymerase II is responsible for efficient synthesis of mRNAs, whose controlled recruitment to promoter region is crucial for transcription regulation (Choder, 2004). Polymerase II is a nuclear complex, as such *Rpb4* and *Rpb7* reside in the nucleus. However, these proteins have been reported to be present in both

nucleus and cytoplasm. Other lines of evidence suggest that *Rpb4* and *7* localization is environmental, as *Rpb4* is formed in large numbers within the cytoplasm in response to high temperature stress in response to iron overload as well where by *Rbp4* expression is up-regulated and therefore, localized such that its function in iron metabolism could be carried out (Khazak *et al.*, 1995; Farago *et al.*, 2003). *Rpb4* has also been shown to be re-located in response to heat shock in *S. pombe* cells. These findings support the role of *Rpb4* is mRNA export (Kimura *et al.*, 2001). Furthermore, expression data analysis suggests the significant expression of arrestin at 3H-15D. Arrestins are a family of proteins involved in signal transduction that is responsible for regulating the function of G-protein-coupled receptors (*GPCRs*) (Wilden *et al.*, 1986). Arrestin-related trafficking adaptors (*ARTs*) act as adaptors for ubiquitin ligases in response to specific endocytic transport signals (Mittal and McMahon, 2009). Plasma membrane proteins including transporters, receptors and channels may be targets of specifics arrestin-related transport adaptors for endocytic down regulation, which is achieved through the recruitment of ubiquitin *E3* ligase *Rsp5*. Endocytosis regulates plasma membrane proteins in response to environmental changes (Becuwe *et al.*, 2012). The results of this study, propose an iron-induced activation of arrestin-related proteins as well as an *IRE*-regulating mechanism for iron-induced activities of these proteins.

Protein disulfide-isomerase (*PdI*), an enzyme localized in the cytoplasmic reticulum in eukaryotes, is another protein up-regulated in response to blood meal throughout the experimental time points. *PdI* recognizes and interacts with unfolded/misfolded protein substrates and in doing so it allows the substrate to reach its native arrangement of disulfides (Wilkinson and Gilbert, 2004). However, the exact mechanism of *PdI* in the endoplasmic reticulum, its coordination and regulation are not well-understood. Despite the absence of extensive literature on the linkage between iron metabolism and *PdI*, Sliskovic and Mutus (2006) showed the involvement of both *PdI* and iron in the induction and/or inhibition of Caspase3. It was observed that iron (Fe^{3+}) inhibits Caspase3 activity, which was shown to be reversible in the presence of *PdI* as the *PdI*-Fe^{3+} complex allowed for the activation of CaspaseB (Sliskovic and Mutus, 2006). The major function of *PdI* is the isomerization and rearrangement of disulfide bonds (Darby *et al.*, 1994). *PdI* has

been reported to have zinc, copper and iron binding activities through its active site thiols (Solovyov and Gilbert, 2004; Sliskovic and Mutus, 2006). The binding of *PdI* to iron is possible due to the resemblance of *PdI* active site to [2Fe-2S]-ferrodoxin iron binding domain (Yeh *et al.*, 2000).

Additionally, several genes were downregulated at 3H-15D in response to the iron-rich blood meal, including the mitochondrial processing peptidase, mitochondrial ATP synthase, meiosis-specific nuclear structural 1, synaptotagmin, Bruno-3, *SAM* motif-containing protein, GJ11024 and Bruton's tyrosin kinase. Mitochondrial-processing peptidase B belongs to the peptidase gene family located within the mitochondrial matrix that catalyzes the cleavage of proteins imported into the mitochondria. Furthermore, mitochondrial ATP synthase is an enzyme involved in energy metabolism within the mitochondria, which is the unit for cellular energy production, iron metabolism, programmed cell death, and the production of reactive oxygen species (Lill and Malenhoff, 2008; Boveris *et al.*, 1972; Jones and Fink, 1982). One of the critical roles of mitochondria is the synthesis of iron-sulfur clusters (*ISCs*) important for several catalytic and structural functions (Lill and Malenhoff, 2008). The iron regulated nature of this gene is plausible as it is under high iron concentration that the iron-sulfur cluster is assembled within the iron regulatory protein 1 (*IRP*1) that prevents its binding to 5'-IRE.

Another down-regulated gene belongs to the reproductive machinery, named meiosis-specific nuclear structural1 (*MNS1*). *MNS1* is a skeletal protein that plays a crucial role in spermatogenesis as well as motile ciliary function (Zhou *et al.*, 2012; Cao *et al.*, 2006). The implications of iron in reproductive health are well established, however the exact role of *MNS1* gene is iron-metabolism remains to be elucidated.

Sterile alpha motif (*SAM*) containing protein is another down-regulated gene in response to blood meal. *SAM* domains have been identified in several proteins involved in a variety of cellular processes including cell signaling. *SAM* domain has been associated with *Eph* family of tyrosine kinase receptors, *AF6,* protein tyrosin phosphatase, *ETS* family of transcription factors, polychromatic proteins, diacylglycerol kinase liprins, the

connector enhancer of *KSR*, serine-threonine kinases, and adaptor proteins (Hirai *et al.*, 1987; Stein *et al.*, 1998; Kyba and Brock, 1998; Sakane *et al.*, 1996; Ponting, 1995; Therrien *et al.*, 1998). The mechanism by which *SAM*-containing proteins maintain iron homeostasis is not well-understood. Additionally synaptotagmin, a Ca^{2+} binding protein identified as a putative *IRE*-regulated gene shows a significant suppression at all experimental time points as well. In response to an initial signal transmission, Ca^{2+} triggers the release of neurotransmitter through possible activation of synaptotagmin (FernĞndez-ChacŮn *et al.*, 2001). There is a well-established link between iron and neurotransmission, whereby an altered functionality of dopaminergic response is observed under iron deprivation (Konofal *et al.*, 2004; Youdim and Green, 1977; Quiroz *et al.*, 2010). Another blood meal-induced suppressed gene is the Bruton's tyrosin kinase (*Btk*) that is a multi-signaling pathway component involved in a diversity of functions. *Btk* has a key role in B-cell proliferation, development, differentiation survival, and apoptosis (Satterthwaite *et al.*, 1998; Islam and Smith, 2000). Bruno-3 is an RNA-binding gene belonging to the BRUNOL family (Kim-Ha *et al.*, 1995) that is involved in the post-transcriptional regulation of gene expression. The main regulatory role of Bruno-3 in the nucleus is the control of alternative splicing. Furthermore, Bruno-3-like gene families have been shown to control translation and mRNA stability in the cytoplasm Barreau *et al.* (2006). Finally GJ11024 is a *MPV17* domain-containing protein, often encoded in *MPV17* gene is another significantly repressed gene. *MPV17* is a component of mitochondrial inner membrane involved in the metabolism of reactive oxygen species (Spinazzola *et al.*, 2006). Additionally, mutations in human *MPV17* gene have been identified as a causative factor in mitochondrial DNA depletion syndrome (Wang *et al.*, 2007). However, no direct link between the function of this gene and iron metabolism is known.

3.4 Conclusion

In conclusion, the mapping of the predicted *IRE*-regulated genes in Glossina to the available blood-induced expression data for Anopheles have provided a comprehensive knowledge of iron metabolism in blood feeding insects and putatively in tsetse flies. Though Drosophila is more closely related to Glossina with the last common ancestor ~ 47.9 Mya

compared to ~274.9 Mya with Anopheles, the blood-feeding nature of mosquitoes, allows for deriving at a more biologically relevant picture of events occurring following the ingestion of iron from blood meal. Furthermore, research is needed to validate these findings and define the relationship between these genes in the context of vectorial competence. Accordingly, several *IRE*-regulated cell-envelope proteins beside ferritin were identified with significant expression at several time points PBM. Furthermore, among the assigned GO terms to *IRE*-regulated genes, transcription, translation and amino acid biosynthesis were identified as over-represented functions that remained significantly expressed at all experimental time points. To this end, genes identified in this study with implications in iron metabolism, whose timely expression or repression allows for maintaining iron homeostasis, may be of interest as targets for future work.

Chapter 4

Evolution of *IRE*-regulated genes in hematophagous insects

Abstract

Background:

Identification and characterization of genes involved in iron regulation from hematophagous insects are useful for the elucidation of evolutionary mechanisms that have allowed for their adaptation to the iron-rich blood meal. Although, several components of iron metabolism have been described in some of the blood-feeding insects, the exact mechanism by which such regulation takes place remains unclear. This study takes a closer look at iron regulation through, identifying the evolutionary patterns associated with Hematophagy. The putative *IRE*-regulated genes of blood feeding and non-blood feeding insect species were compared to derive at signatures specific to blood-feeding.

Methods:

Detection of orthologs for 150 putative *IRE*-regulated genes in several blood feeding (BF) (*A. gambiae*, *A. aegypti*, *P. humanus*, *C. quinquefasciatus*, *I. scapularis*) and non-blood feeding (NBF) insects (*D. melanogaster*, *A. mellifera*, *A. pisum*, *B. mori*, *D. Plexippus*) were carried-out using QuartetS. Multiple sequence alignment was performed using MUSCLE, while MEGA was used to generate ML trees. Several evolutionary hypotheses were assessed using PAML. These include the neutrality tests where one-ratio and two-ratio models were compared with

their alternative models. The significance of the results for each model was assessed using the Likelihood Ratio Tests (LRTs). The Likelihood Ratios were then compared with the $\chi 2$ value with a degree of freedom to assess their significance. Additionally, we investigated selection pressure among sites for the putative *IRE*-regulated genes, whereby several models including M0, M1a, M2a, M3, M7 and M8 were used. To assess the significance of the findings, the LRTs were conducted where by M0, M1a and M7 were compared to the alternative models i.e. M2a, M3 and M8. Furthermore, sites with a Bayesian posterior probability ($\geq$95%) were considered to be under positive selection.

Results:

Both blood feeding and non-blood feeding insect genes with putative roles in iron metabolism are experiencing purifying selection (though the scale at which such pressure is experienced differs). In addition, several variations in the evolution of *IRE*-regulated genes were observed. These can be summarized as genes undergoing accelerated evolution namely, the phosphoinositide3-kinase, ubiquitin-conjugating enzyme9, general transcription factor IIH subunit1, serine-pyruvate mitochondrial, pentatricopeptide Transcription and translation, tRNA(uridine-2-o-)-methyl-transferase-trm7, mediator 100kD, as well as genes with sites under positive selection including, nucleoporin, eukaryotic translation initiation factor 3, bruno-3 transcript, general transcription factor IIH subunit1, ferritin heavy-chain protein, serine-pyruvate mitochondrial, brain chitinase and chia.

Conclusion:

In the context of iron metabolism the results of this study provides new perspective on hematophagy in insects through identifying positive selection as the evolutionary mechanism that drives insect adaptation to blood meal iron.

4.1 Introduction

4.1.1 Problem identification

A key role of evolutionary biology is to understand variation in traits associated with fitness. Iron metabolism is an important component of fitness. This is supported by several lines of evidence pointing to the important implications of iron in survival. If indeed iron metabolism is so closely related to fitness, it seems paradoxical to find an extreme level of polymorphism in iron metabolism genes. Furthermore, it is expected that natural selection will purge deleterious mutations and quickly fix beneficial mutations (Aylmer,

1930; Smith *et al.*, 1974; Charlesworth, 1993). However, it is important to note that iron metabolism is extremely dynamic and affected by several factors. In spite of great interest in elucidating the principles of iron metabolism, many basic questions remain unresolved. Hence, the lack of consensus about the incident, type and strength of selection acting on iron-metabolism genes of blood-feeding insects as well as the identification of 150 *IRE*-regulated genes (Chapter 2) motivates this study and allows for the elucidation of the evolutionary variations within and between blood-feeding and non-blood-feeding insects. Several evolutionary hypotheses will be tested to start disentangling the contribution of different factors to the evolution of *IRE*-regulated genes in blood-feeding insects.

Haematophagy or blood-feeding insects have evolved at least six times during the Jurassic and Cretaceous periods *i.e.* 145-65 mya (Balashov, 1984). The blood-feeding nature of certain insects have been explained by several hypothesis based on which the origin of blood sucking behavior of insects have been linked to their close association with vertebrate hosts. Another hypothesis suggests that an ancestral form of certain insects were parasitic, which have then evolved alongside vertebrates (Lehane, 2005). These and several other hypotheses exist, while evolutionary analysis of genes present in blood feeding insects compared with their non-blood feeding counterparts may lead to the detection of blood-feeding-specific signatures hence, shed light on molecular changes that have allowed for such adaptation to blood meal and their success in nature. Blood feeding is a complex process whereby the insect faces several drawbacks while, feeding on hosts blood, such as hemostasis, and inflammation as part of the host defense. According to transcriptome studies it is now clear that salivary proteins play a profound role in the process of blood feeding as part of more complex interdependent events at work (Alves-Silva *et al.*; Valenzuela *et al.*, 2003; Olivares *et al.*, 2009; Ribeiro, 2003). Another important danger imposed on haematophagous insects is the iron overload occurring following a blood meal. However, it is evident that a well-coordinated network of mechanisms are in place to prevent iron-induced toxicity, while directing the consumed iron toward processes such as egg development (Chapter 3). Hence, understanding such adaptation in the context of blood feeding and iron regulation allows for understanding evolutionary drivers at play.

4.1.2 Phylogenetic analysis methods in molecular evolution

Understanding the evolutionary processes of DNA and proteins in different organisms is the main theme of molecular evolution. Such studies provide insights into the evolution of organisms in which these molecules have evolved, while also shedding light on the biological functions of these molecules (Felsenstein, 1988). Three main processes govern the distribution of alleles in a population namely mutation, genetic drift, and selection. The process by which new alleles are introduces into a population is referred to as mutation, whereby a novel genetic variation may pass along during the process of DNA replication. The role of selection can therefore be explained as a control mechanism on the random drifting of allele frequencies. The tendency of beneficial alleles to increase and deleterious alleles to decrease over time is described by Darwinian selection. Accordingly positive selection is explained as the tendency of alleles with adaptive advantage to increase over time, while purifying selection refers to the tendency of deleterious alleles to decrease in frequency. Positive selection potentially affects the fate of a population while being transient *i.e.* a period of positive selection may be followed by a period of purifying selection, once a novel property has been established in the population.

4.1.3 Applications

Advances in molecular techniques have allowed the study of signatures of adaptive events in the genome and transcriptome of different species. Several attempts have been made to discern changes that are evolutionarily adaptive and have undergone positive selection from those that are purely the result of drifts (Hughes and Nei, 1988).

The evolutionary relationship and the pattern of evolutionary history of a group of sequences are often represented as a phylogenetic tree (Page and Holmes, 2009). To achieve this a multiple sequence alignment is first generated, and a phylogeny is reconstructed. As a result changes in rate and pattern of gene evolution can be estimated. A phylogeny describes how a single common ancestral gene has evolved to give rise to descendent genes. Furthermore, a phylogenetic tree is categorized into two components namely the tree topology and the branch length, with the former describing the pattern of divergence

and the latter the amounts of evolutionary time between divergences. However the decisiveness of such analysis relies heavily on generating accurate alignment of sequences based on which the evolutionary trees are drawn. Several tools have therefore been introduced to meet the need of such analysis including ClustalW for closely related sequences (Thompson *et al.*, 1994), Molecular Evolutionary Genetics Analysis (MEGA) (Tamura *et al.*, 2007), Multi-LAGAN for genome-wide homologous sequences (Brudno *et al.*, 2003) and several others.

The next step in analysis is the reconstruction of a phylogenetic tree, for which several algorithms exist that can be classified into distance-based methods and character-based methods (Salemi and Vandamme, 2003). Neighbor-joining (Saitou and Nei, 1987) is an example of distance-based methods, according to which, the evolutionary distance between each pair of sequences is measured based on their sequence similarity. As a result sequences are clustered according these distance measures.

Character-state methods on the other hand including maximum parsimony method (Kluge and Farris, 1969; Fitch, 1971), maximum likelihood method (Felsenstein, 1981) and Bayesian Inference method (Huelsenbeck *et al.*, 2001), consider each site position in the alignment as a character, and the sequence at each site as the state for this character. Under specific criterion, this approach then identifies an optimal tree (Swofford *et al.*, 1996). The maximum parsimony method aims to find the optimal tree with the smallest number of evolutionary changes that can explain the differences of states in the sequences (Fitch, 1971), while maximum likelihood and Bayesian methods search for optimal trees with highest likelihood or posterior probability values. The next step of the analysis involves the detection of evolutionary constraints acting on the sequences. This is often achieved by modeling the underlying substitution process *i.e.* the rate at which one nucleotide is substituted for another. To distinguish between synonymous and non-synonymous substitutions that are possibly subject to purifying, neutral, or positive selection several codon substitution models have been introduced and are commonly used. As such the ratio of the non-synonymous (dN) to synonymous (dS) substitution rate denoted as ω reflects the type of selective pressure experienced. Therefore $\omega<1$ is associated with

purifying selection, ω>1 is evidence of positive selection, and ω=1 is considered neutral selection (Bielawski and Yang, 2003). Generally dN is expected to be smaller than dS (ω<1) because non-synonymous substitutions are more likely to disrupt function than to be beneficial for the protein.

In understanding the process of molecular evolution statistical methods such as maximum likelihood estimation of the dN/dS ratio (Goldman and Yang, 1994; Muse and Gaut, 1994) as well as the likelihood ratio test (Muse and Gaut, 1994; Nielsen and Yang, 1998) for positively selected genes are of great value.

4.1.3.1 Test of selection in the likelihood context

Several attempts have been made to identify codons under positive selection, many of which have made use of phylogenetic trees and likelihood based models of evolution implemented in PAML v4.0 (Yang, 2007). These models estimate whether a site is under positive selection, based on the way omega varies over sites (M0-M8) or over lineages (branch models and branch site models). Furthermore, to test the significance of results, the fitness of the nested models allowing positive selection should be compared with models that do not. Nested models fit exactly the same evolutionary scenario, except that one of the two allows positive selection and the other does not. To do this, Likelihood ratio tests (LRTs) (Yang and Nielsen, 1998) are conducted whereby the results are significant if only a better fit is observed in the likelihood. The LRT is calculated as twice the difference in the maximum likelihood values between nested models.

Considering the significant physiological importance of blood-feeding in insects, this chapter focuses on the identification of selection forces that have shaped the evolution of hematophagous insects in the context of iron-metabolism. To achieve this, several evolutionary models will be tested. The results of which may help shed light on the mechanism by which these insect have adapted to iron-overload from the blood-meal.

4.2 Materials and Methods

4.2.1 Data retrieval and sequence alignment

The *IRE*-regulated genes (identified in chapter 2) were used to identify corresponding orthologs in several blood feeding insects (*A. gambiae*, *A. aegypti*, *P. humanus*, *C. quinquefasciatus*, *I. scapularis*) as well as non-blood feeding insects (*D. melanogaster*, *A. mellifera*, *A. pisum*, *B. mori*, *D. plexippus*). Of the 150 putative *IRE*-regulated genes in Glossina, 34 have orthologs in at least two blood-feeding and two non-blood-feeding insects and Silk worm (out-group for later analysis). The corresponding cDNA sequences were retrieved from Ensembl (`http://metazoa.ensembl.org/info/data/ftp/index.html`). Furthermore, ORF Finder (`http://www.ncbi.nlm.nih.gov/projects/gorf/`) was used to clarify the longest open reading frame (ORF). The collected cDNA sequences were further translated to amino acid sequences using translatorx_vLocal.pl (Abascal *et al.*, 2010) to obtain an amino acid guided cDNA alignment. Moreover, as part of the parameters, building a multiple alignment using MUSCLE (Edgar, 2004) was specified. Furthermore, MEGA v.5.0 (Kumar *et al.*, 1994) was used to generate the Maximum Likelihood (ML) trees, where the number of bootstrap replication was set to 1000 and the tree inference was performed by specifying Nearest Neighbor Interchange (NNI) as the ML heuristic method to be used. The resulting trees were then used as guide trees for the program PAML v4.0 (Yang, 2007). As a result 34 *IRE*-regulated genes were further assessed using PAML. The Codeml program requires three different input files, one containing the sequences, a tree file, and the control file *i.e.* codeml.ctl containing parameters of the models. To achieve this, sequences were converted to PHYLIP format, while the tree is in newick format. Furthermore, to assess selection on blood feeding compared to non-blood feeding insects the branches pertaining to each group were labeled on the newick formatted tree file.

We tested for selection in the insect phylogeny by comparing the number of non-synonymous substitutions per nonsynonymous sites (dN) to the number of synonymous substitutions per synonymous site (dS) in a maximum likelihood (ML) framework. A dN/dS ratio $>$ 1 suggests that many nonsynonymous mutations offer a fitness advantage and are fixed in populations at a rate greater than synonymous changes, and is evidence of positive

selection. A dN/dS ratio = 1 indicates neutral evolution and a dN/dS ratio < 1, refers to purifying selection.

4.2.2 Investigating selection pressure among lineages

In order to investigate the putative role of selection on the adaptation of *IRE*-regulated genes in blood-feeding insects the neutrality tests were carried out, whereby one-ratio and two-ratios models were compared with their corresponding alternative models (omega was fixed to one, ω=1). Figure 4.1 represents several evolutionary assumptions for the neutrality tests used to detect selection in the *IRE*-regulated genes among blood feeding and non-blood feeding insects. It is notable that the tree layout differs for each of the IRE-regulated genes assessed and that figure 4.1 is a schematic presentation of the tree for a single IRE-regulated gene.

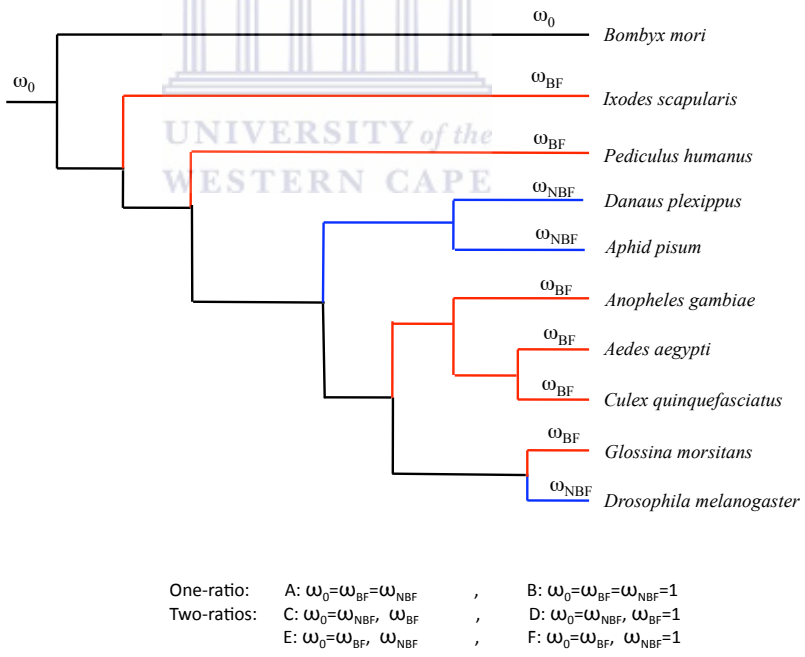


Figure 4.1: Schematic illustration of hypotheses used for examination of the type of selection acting on the two lineages (blood-feeding (BF) and non-blood-feeding (NBF) insects) using tests of neutrality including the one-ratio and two-ratio tests, where assumptions A, C and E are compared with B, D and F respectively.

The one-ratio model estimates a single ω value for the entire tree (BF and NBF insects).

The results were then compared with the alternative hypothesis assuming ω=1 for the entire tree. Under a likelihood framework, the results of models are significant only if they show a better fit in the Likelihood Ratio Tests (LRT)s relative to their null models. The LRT is calculated as twice the difference in maximum likelihood values between nested models. The degree of freedom is calculated as the difference of numbers of parameters between two models (df=1). The significance of the LRT statistics were, evaluated using χ^2 distributions with a degree of freedom.

The two ratio models on the other hand estimate the omega value for a specific lineage in the tree, and compare it to the null hypothesis *i.e.* ω=1 to determine whether the specified lineage is under selection. In order to specify the lineage of interest for the two-ratio analysis, it is required that the tree is labeled, whereby branches on which the neutrality test is carried out are labeled using "#1" character in the input tree file. This was carried out for both blood feeding and non-blood-feeding insect branches separately (Figure 4.2).

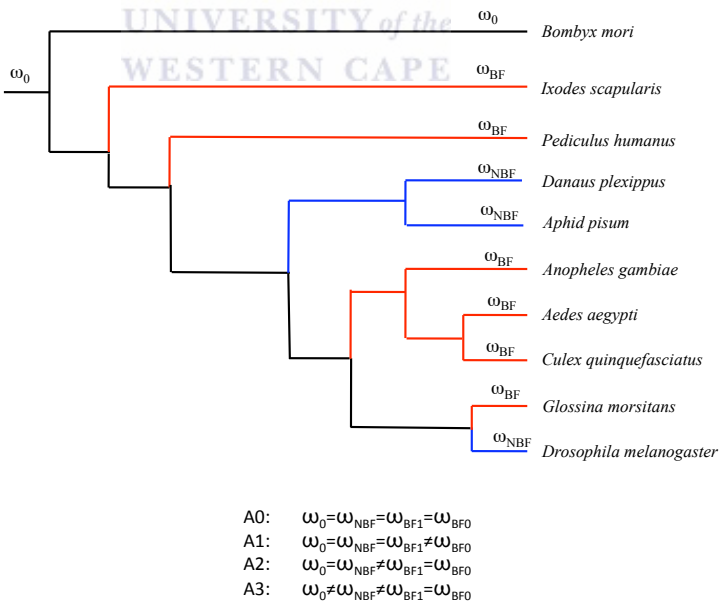


Figure 4.2: Schematic illustration of hypotheses used for examination of the possible variations in selection pressure between the two lineages (blood-feeders and non-blood-feeders) using branch models of adaptive evolution. Accordingly A1 and A2 are compared with A0, while A2 and A3 are also compared.

The following assumptions: A1) the mutation rate of *IRE*-regulated genes in BF-insects have increased relative to NBF-insects, A2) a difference in selection have occurred following the speciation event that gave rise to blood-feeding insect and A3) a long-term change in selection pressure have led to the evolution of blood-feeding insects, which was evaluated using branch models (one-ratio (A0), two-ratios (A1 and A2), and three ratios (A3) models) were used. A series of comparisons were carried out including, A0 *vs.* A1, A0 *vs.* A2 and A2 *vs.* A3 the significance of which were evaluated using $\chi 2$ distributions with a degree of freedom.

4.2.3 Investigating selection pressure among sites

As adaptive evolution is restricted to a small subset of sites, modeling variable selective pressures among sites provide a better understanding compared to modeling variable selective pressure among branches (Yang and Bielawski, 2000). Accordingly various models have been introduced by Nielsen and Yang (1998) as well as Yang and Bielawski (2000) all of, which are implemented in the CODML program. Furthermore, to examine sites that have evolved under positive Darwinian selection, Likelihood ratio tests (LRTs) (Yang and Nielsen, 1998) are conducted whereby models M0 (one-omega ratio), M1a (neutral) and M7 (beta) are compared with more complex models that allow for positive selection namely; M2a (selection), M3 (district) and M8 (beta + ω) as the latter models are commonly employed in detecting adaptive evolution among sites (Nielsen and Yang, 1998; Anisimova *et al.*, 2001). Furthermore, Bayesian posterior probabilities of site classes are calculated for each amino acid site (Nielsen and Yang, 1998) and positive selection is indicated when a freely estimated ω parameter is > 1 and the LRT is significant (Bielawski and Yang, 2003). Therefor sites with a significant Bayesian posterior probability ($\geq 95\%$) would be considered under positive selection.

4.3 Results and Discussions

4.3.1 Selection pressure among lineages

The evolutionary selective pressures acting on *IRE*-regulated genes between BF and non-BF insects were studied, using various branch models implemented in PAML. Table 4.1 shows the parameter estimates and likelihood scores for each model, whereby model A estimates a single ω for the entire tree (ω_0=ω_{BF}=ω_{NBF}), model C estimates ω for BF insects (ω_0=ω_{NBF}, ω_{BF}) and model E estimate ω value for the NBF insects (ω_0=ω_{BF}, ω_{NBF}). Accordingly models B-F fix ω value to 1 for each of the alternative models (null hypotheses). The estimated likelihood scores for each of the branch models are then used in the calculation of likelihood ratios, whereby different models are compared (*i.e.* models A *vs.* B, C *vs.* D, and E *vs.* F are compared, accordingly). These likelihood ratios were compared with the χ^2 value to assess their significance with p-value $< 0.01/0.05$.

The comparison of models A and B, which tests selection in the entire tree of BF and NBF insects suggest that all 34 *IRE*-regulated genes examined are subject to selection. This can be deduced from the significant likelihood ratios ranging from 503.62 to 6742.91 for these genes ($p<0.01$; ${\chi_1}^2=6.63$, $p<0.05$; $\chi2=3.84$). The estimated ω value for model A (one-ratio) is significantly less than one ($0.00 < \omega < 0.09$), which is an evidence of purifying selection associated with functional constraints acting on all *IRE*-regulated genes in insects (blood feeding and non-blood feeding). To better understand the results it is important to determine whether this selection is influenced by the selection in blood-feeding, non-blood-feeding, or both branches (lineages). The comparison of models C and D, which examine selection in the blood-feeding insect lineage, shows selection to be acting on *IRE*-regulated genes with significant likelihood ratios ranging from 227.39 to 2637.60. Further assessment of the ω value points to purifying selection which is evident from $0.00 < \omega < 0.16$. Similar analysis was carried out for the non-blood feeding insect lineage by comparing model E and F, whereby all 34 *IRE*-regulated genes were found to be under purifying selection as the likelihood ratio values ranging from 81.47 to 2367.78 ($p<0.01$; ${\chi_1}^2=6.63$, $p<0.05$; $\chi2=3.84$) and $0.00 < \omega < 0.10$ are significant (Table 4.1).

Table 4.1: Parameter estimates and likelihood scores for the predicted *IRE*-regulated genes under neutrality test of selection using branch-specific models.

Gene	No. BF	No. NBF	A. One-ratio $\omega_0=\omega_{BF}=\omega_{NBF}$			B. One-ratio $\omega_0=\omega_{BF}=\omega_{NBF}=1$		LRT A *vs.* B	C. Two-ratios $\omega_0=\omega_{NBF},\omega_{BF}$			D. Two-ratios $\omega_0=\omega_{NBF},\omega_{BF}=1$		LRT C *vs.* D	E. Two-ratios $\omega_0=\omega_{BF},\omega_{NBF}$				F. Two-ratios $\omega_0=\omega_{BF},\omega_{NBF}=1$		LRT E *vs.* F
			P	*l*	ω_0	P	*l*		P	*l*	ω_0	ω_{BF}	P	*l*		P	*l*	ω_0	ω_{NBF}	P	*l*	
TMP002520	4	2	14	-43642.35	0.09	13	-45642.00	3999.29**	15	-43636.93	0.04	0.10	14	-44955.73	2637.60**	15	-43641.42	0.10	0.07	14	-43894.42	506.00**
TMP002587	5	3	18	-3406.02	0.01	17	-3672.80	533.55**	19	-3402.36	0.16	0.02	18	-3577.31	349.90**	19	-3400.93	0.02	0.16	18	-3441.66	81.47**
TMP003059	4	3	18	-7838.94	0.00	17	-8765.05	1852.22**	19	-7840.66	0.00	0.04	18	-8213.48	745.63**	19	-7843.15	0.03	0.00	18	-8173.96	661.62**
TMP003060	4	4	18	-6921.97	0.00	17	-7461.43	1078.93**	19	-6921.02	0.00	0.10	18	-7105.66	369.27**	19	-6925.16	0.08	0.00	18	-7152.94	455.57**
TMP003203	5	4	20	-17212.62	0.03	19	-19631.53	4837.83**	21	-17202.14	0.01	0.04	20	-18309.25	2214.21**	21	-17205.30	0.04	0.01	20	-18131.13	1851.66**
TMP003297	5	2	16	-16786.83	0.01	15	-19041.00	4508.33**	17	-16776.08	0.00	0.03	16	-18117.25	2682.33**	17	-16782.07	0.03	0.00	16	-17180.10	796.06**
TMP003500	5	3	18	-17516.08	0.00	17	-18450.55	1868.95**	19	-17516.02	0.00	0.00	18	-18022.95	1013.86**	19	-17516.02	0.00	0.00	18	-17767.28	502.51**
TMP003577	4	4	18	-11511.05	0.02	17	-13053.01	3083.93**	19	-11501.60	0.01	0.03	18	-12238.29	1473.39**	19	-11503.33	0.03	0.01	18	-12066.67	1126.69**
TMP003602	5	4	20	-52850.45	0.07	19	-56221.90	6742.91**	21	-52811.28	0.01	0.08	20	-54626.81	3631.05**	21	-52820.86	0.08	0.01	20	-53871.37	2101.03**
TMP004240	5	3	18	-6801.49	0.02	17	-8272.97	2942.95**	19	-6801.37	0.03	0.02	18	-7624.46	1646.17**	19	-6798.33	0.02	0.03	18	-7234.08	871.49**
TMP005511	5	4	20	-6649.62	0.03	19	-7329.83	1360.42**	21	-6641.41	0.00	0.04	20	-7014.66	746.49**	21	-6644.37	0.04	0.00	20	-6873.27	457.81**
TMP005554	5	3	18	-15691.83	0.00	17	-17068.56	2753.48**	19	-15691.82	0.00	0.00	18	-16490.77	1597.89**	19	-15691.82	0.00	0.00	18	-15998.76	613.87**
TMP005595	5	4	20	-32708.12	0.06	19	-34204.22	2992.20**	21	-32693.49	0.01	0.09	20	-33469.26	1551.55**	21	-32697.89	0.08	0.01	20	-33199.76	1003.76**
TMP006377	5	4	20	-6399.86	0.01	19	-6756.73	713.73**	21	-6396.81	0.01	0.02	20	-6560.61	327.60**	21	-6395.55	0.06	0.01	20	-6540.08	289.05**
TMP006617	4	4	18	-6491.00	0.02	17	-7297.69	1613.37**	19	-6489.40	0.02	0.03	18	-6842.91	707.02**	19	-6489.30	0.03	0.02	18	-6764.77	550.95**
TMP006998	5	4	20	-12562.81	0.01	19	-13500.06	1874.50**	21	-12559.89	0.01	0.00	20	-13078.52	1037.26**	21	-12559.73	0.00	0.01	20	-12836.67	553.89**
TMP007507	4	4	18	-19734.66	0.00	17	-22475.43	5481.54**	19	-19730.11	0.00	0.01	18	-20886.94	2313.66**	19	-19731.89	0.01	0.00	18	-20915.77	2367.78**
TMP008000	4	4	18	-8671.81	0.00	17	-9358.45	1373.28**	19	-8671.71	0.00	0.00	18	-8961.05	578.69**	19	-8671.72	0.00	0.00	18	-8984.03	624.62**
TMP008942	5	4	20	-33804.28	0.06	19	-36422.55	5236.54**	21	-33773.59	0.00	0.08	20	-34978.42	2409.66**	21	-33789.27	0.08	0.04	20	-34821.20	2063.86**
TMP009157	5	4	20	-2911.88	0.01	19	-3736.69	1649.61**	21	-2905.98	0.01	0.01	20	-3225.74	639.53**	21	-2904.12	0.01	0.00	20	-3272.21	736.18**
TMP009871	4	3	16	-6522.79	0.04	15	-7484.80	1924.01**	17	-6522.59	0.04	0.04	16	-6947.42	849.67**	17	-6522.41	0.04	0.05	16	-6834.93	625.04**
TMP010016	5	3	18	-15633.21	0.02	17	-16840.51	2414.60**	19	-15629.49	0.01	0.03	18	-16430.96	1602.94**	19	-15629.45	0.03	0.01	18	-15879.04	499.19**
TMP010317	4	4	18	-14943.00	0.04	17	-16248.73	2611.47**	19	-14930.19	0.01	0.06	18	-15495.39	1130.41**	19	-14932.27	0.05	0.01	18	-15399.89	935.24**
TMP010673	4	3	16	-4873.94	0.00	15	-5244.33	740.79**	17	-4874.33	0.00	0.10	16	-4988.03	227.39**	17	-4873.61	0.00	0.00	16	-5021.96	296.70**
TMP010846	4	4	18	-10645.46	0.00	17	-11163.02	1035.13**	19	-10640.86	0.01	0.11	18	-10819.14	356.56**	19	-10645.28	0.01	0.01	18	-10884.53	478.49**
TMP011104	4	3	16	-8988.69	0.02	15	-9755.54	1533.71**	17	-8984.91	0.01	0.00	16	-9449.88	929.95**	17	-8984.93	0.00	0.01	16	-9182.82	395.78**
TMP011448	5	2	16	-12871.11	0.00	15	-13604.11	1466.01**	17	-12877.59	0.01	0.07	16	-13316.01	876.85**	17	-12880.67	0.06	0.01	16	-12975.95	190.57**
TMP011786	4	4	18	-11736.66	0.01	17	-12587.01	1700.69**	19	-11736.54	0.01	0.01	18	-12178.10	883.12**	19	-11736.66	0.01	0.01	18	-12076.34	679.36**
TMP013137	5	4	20	-21677.45	0.05	19	-22690.94	2026.98**	21	-21662.48	0.01	0.09	20	-22233.60	1142.25**	21	-21670.91	0.08	0.02	20	-21924.30	506.78**
TMP013886	4	4	18	-7409.28	0.03	17	-8293.37	1768.18**	19	-7407.24	0.00	0.04	18	-7792.18	769.88**	19	-7407.55	0.04	0.01	18	-7726.14	637.17**
TMP013948	4	4	18	-4076.22	0.05	17	-4328.03	503.62**	19	-4075.80	0.04	0.06	18	-4172.49	193.37**	19	-4075.98	0.06	0.04	18	-4168.82	185.68**
TMP014133	5	4	20	-7361.03	0.02	19	-8096.76	1471.46**	21	-7355.90	0.00	0.03	20	-7775.83	839.86**	21	-7358.37	0.03	0.00	20	-7585.20	453.67**
TMP014187	5	4	20	-28885.08	0.06	19	-30667.62	1782.54**	21	-28869.47	0.01	0.08	20	-29765.81	1792.68**	21	-28873.44	0.07	0.00	20	-29370.27	496.83**

Note: * Significant ($p<0.05$; $\chi^2_1=3.84$), ** Extremely significant ($p<0.01$; $\chi^2_1=6.63$)

Gathered from these results it is evident that *IRE*-regulated genes are experiencing similar types of evolution *i.e.* purifying selection. As such to further investigate possible variations in selection pressure between the two lineages (blood-feeders and non-blood-feeders) several extra evolutionary hypotheses were assessed (Section 5.2.2, Figure 4.2). According to the results model A1 fits the data significantly better than the null hypothesis (A0) for seven *IRE*-regulated genes indicating an increased mutation rate in BF-insects relative to non-blood feeding insects. These include TMP003602 (phosphoinositide3-kinase), TMP009157 (ubiquitin-conjugating enzyme9), TMP010317 (general transcription factor IIH subunit1), TMP011104 (serine-pyruvate mitochondrial), TMP013137 (pentatricopeptide Transcription and translation), TMP013886 (tRNA(uridine-2-o-)-methyl-transferase-trm7) and TMP014187 (mediator 100kD). However this scenario does not apply to the rest of *IRE*-regulated genes (*i.e.* 27), where A1 was rejected, suggesting that most *IRE*-regulated genes under study do not experience an increased rate of mutation in relation to blood-feeding insects.

Furthermore, model A2 fits the majority of *IRE*-regulated genes (more than 70% of genes) significantly better that the null model, where a difference in selection pressure have occurred following the hypothetical speciation event that has given rise to the blood-feeding *IRE*-regulated genes. This is evident from the significant likelihood ratios ranging from 4.08 to 78.33. Furthermore, to evaluate if there has been a long term shift in selection pressure model A3 was compared to A2, whereby, the assumption that the putative *IRE*-regulated genes have experienced a long term shift was rejected except for three genes, namely TMP002520 (nucleoporin), TMP004240 (elongation factor1 alpha) and TMP009157 (ubiquitin-conjugating enzyme9) (Table 4.2).

Thus, the pattern that emerges from this study points to similar selection *i.e.* purifying selection acting on *IRE*-regulated genes in blood-feeding and non-blood-feeding insects. Such purifying selection was expected for genes known to be important for survival and fitness. Despite similar types of selection, the scale at which such selection is experienced is indeed different between BF and NBF insects, as supported by A2 model. Additionally, the increased rate of mutations in a subset of genes, though few, may point to their

functional adaptation in the context of the iron regulation in BF insects.

4.3.2 Investigating selection among sites

Genes affected by natural selection may represent an excess of functionally important molecular changes that are referred to as genomic regions under positive selection, mainly due to the fact that positive selection favors new genetic variations. As such excess of non-synonymous substitutions over synonymous substitutions beyond what would be expected provide strong evidence of positive selection at the protein level.

To further assess possible variations in the evolution of *IRE*-regulated genes between BF and NBF insects several evolutionary site models *i.e.* M0-M8 were used. Table 4.3 illustrates the LRT ratios and sites under positive selection, whereby site positions represented in bold are sites that are significant at 99% ($p< 0.01$) while the rest are significant at 95% ($p< 0.05$). Accordingly, only sites supported by more than one selection model (*i.e.* M3 and M8) were accepted as positively selected sites. Furthermore, M2a did no fit the data, due to insignificant LRTs when compared to the null hypothesis, *i.e.* M1a. Considering the importance of iron regulation in blood feeding insects it is expected that no or very few synonymously substituted sites exist. In supporting this the results indicate the presence of positively selected sites within seven blood-feeding *IRE*-regulated genes namely TMP002520 (nucleoporin), TMP008942 (eukaryotic translation initiation factor 3), TMP009871(bruno-3 transcript), TMP010317 (general transcription factor IIH subunit1), TMP010673 (ferritin heavy-chain protein), TMP011104 (serine-pyruvate mitochondrial) and TMP011448 (brain chitinase and chia). This suggests a possible role for these genes in the adaptive evolution of iron metabolism in blood-feeding insects. Genes responsible for adaptive evolution often resemble high rate of evolutionary change and are under positive selection (Biswas and Akey, 2006; Turner *et al.*, 2005). These genes are commonly involved in the co-evolutionary processes, such as immune response, host-pathogen interactions and reproduction (Hughes and Nei, 1988; Swanson *et al.*, 2003; Woolhouse *et al.*, 2002). Furthermore, to gain insight into the evolutionary process of iron metabolism the possible function of genes with positively selected sites were clarified.

Table 4.2: Parameter estimates and likelihood scores for the predicted *IRE*-regulated genes under branch-specific models.

Gene	A0		A1				A2				A3					A0 vs A1	A0 vs A2	A2 vs A3
	ω_0 =ω_{NBF} =ω_{BF1} =ω_{BF0}	l	ω_0 =ω_{NBF} =ω_{BF1}	ω_{BF0}	l	ω_{BF0}	ω_0 =ω_{NBF}	ω_{BF1} =ω_{BF0}	l	ω_0	ω_{NBF}	ω_{BF1} =ω_{BF0}	ω_{BF1} =ω_{BF0}	l			
TMP002520	0.09	-43642.35	0.09	530.16	-43641.97	0.04	0.10	0.10	-43636.93	0.10	0.06	0.00	0.00	-43633.22	0.76	10.85**	7.41**
TMP002587	0.01	-3406.02	0.01	0.00	-3405.68	0.15	0.01		-3402.36	0.01	0.16	0.00		-3400.88	0.69	7.32**	2.96
TMP003059	0.00	-7838.94	0.00	8.38	-7838.02	0.00	0.03		-7838.66	0.03	0.00	0.00		-7838.46	1.83	0.56	0.39
TMP003060	0.00	-6921.97	0.00	4.18	-6921.06	0.00	0.09		-6921.02	0.09	0.00	0.00		-6920.77	1.81	1.89	0.49
TMP003203	0.03	-17212.62	0.03	19.10	-17211.69	0.01	0.03		-17202.14	0.03	0.01	0.00		-17201.47	1.85	20.94**	1.34
TMP003297	0.01	-16786.83	0.01	5.48	-16786.83	0.00	0.03		-16776.08	0.03	0.00	0.00		-16776.07	0.00	21.51**	0.03
TMP003500	0.00	-17516.08	0.00	0.06	-17516.06	0.00	0.00		-17516.02	0.00	0.00	0.00		-17516.02	0.04	0.11	0.01
TMP003577	0.02	-11511.05	0.02	0.26	-11510.37	0.00	0.03		-11501.60	0.03	0.00	0.00		-11501.38	1.35	18.90**	0.44
TMP003602	0.07	-52850.45	0.07	999.00	-52847.85	0.00	0.08		-52811.28	0.08	0.00	0.00		-52811.19	5.20*	78.33**	0.19
TMP004240	0.02	-6801.49	0.02	0.01	-6800.89	0.02	0.02		-6801.37	0.02	0.03	0.00		-6794.77	1.20	0.24	13.21**
TMP005111	0.03	-6649.62	0.03	999.00	-6648.64	0.00	0.04		-6641.41	0.04	0.00	0.00		-6640.99	1.95	16.42**	0.84
TMP005554	0.00	-15691.83	0.00	202.05	-15691.03	0.00	0.00		-15691.82	0.00	0.00	0.00		-15691.72	1.59	0.01	0.21
TMP005595	0.06	-32708.12	0.06	999.00	-32706.80	0.00	0.08		-32693.49	0.08	0.00	0.00		-32693.25	2.63	29.26**	0.47
TMP006377	0.01	-6399.86	0.01	102.34	-6398.93	0.00	0.02		-6396.81	0.02	0.00	0.00		-6396.81	1.87	6.11**	0.00
TMP006617	0.02	-6491.00	0.02	290.00	-6490.64	0.01	0.03		-6489.40	0.03	0.01	0.02		-6489.22	0.71	3.21	0.36
TMP006998	0.00	-12562.81	0.00	245.90	-12562.62	0.00	0.00		-12559.89	0.00	0.00	0.00		-12559.71	0.38	5.83**	0.37
TMP007507	0.00	-19734.66	0.00	3.68	-19734.59	0.00	0.01		-19730.11	0.01	0.00	0.00		-19729.96	0.14	9.09**	0.31
TMP008000	0.00	-8671.81	0.00	156.46	-8671.00	0.00	0.00		-8671.71	0.00	0.00	0.00		-8671.69	1.62	0.20	0.04
TMP008942	0.06	-33804.28	0.06	999.00	-33802.48	0.00	0.08		-33773.59	0.08	0.00	0.00		-33773.38	3.60	61.39**	0.42
TMP009157	0.00	-2911.88	0.00	0.00	-2906.75	0.00	0.01		-2905.98	0.01	0.00	0.04		-2903.32	10.26**	11.81**	5.31*
TMP009871	0.04	-6522.79	0.04	0.02	-6522.67	0.04	0.03		-6522.59	0.03	0.04	0.03		-6522.39	0.24	0.41	0.40
TMP010016	0.02	-15633.21	0.02	0.00	-15633.03	0.00	0.02		-15629.49	0.02	0.00	0.01		-15629.35	0.38	7.46**	0.27
TMP010317	0.04	-14943.00	0.04	0.00	-14937.77	0.00	0.05		-14930.19	0.05	0.00	0.00		-14929.97	10.46**	25.61**	0.43
TMP010673	0.00	-4873.94	0.00	33.76	-4873.92	0.00	0.10		-4873.33	0.10	0.00	0.00		-4873.20	0.03	1.20	0.27
TMP010846	0.00	-10645.46	0.00	33.42	-10644.62	0.00	0.10		-10640.86	0.10	0.00	0.00		-10640.42	1.67	9.19**	0.88
TMP011104	0.02	-8988.69	0.00	0.00	-8985.28	0.00	0.00		-8984.91	0.00	0.00	0.04		-8984.49	6.81**	7.56**	0.84
TMP011448	0.00	-12871.11	0.00	2.54	-12871.03	0.00	0.06		-12870.59	0.06	0.00	0.00		-12870.49	0.15	1.04	0.20
TMP011786	0.00	-11736.66	0.00	2.31	-11736.60	0.00	0.00		-11736.54	0.00	0.00	0.00		-11735.74	0.13	0.25	1.60
TMP013137	0.05	-21677.45	0.01	0.00	-21668.10	0.00	0.08		-21662.48	0.08	0.00	0.00		-21662.10	18.70**	29.95**	0.74
TMP013886	0.02	-7409.28	0.00	0.03	-7407.24	0.00	0.03		-7407.24	0.03	0.01	0.03		-7406.55	4.08*	4.08*	1.39
TMP013948	0.05	-4076.22	0.05	0.06	-4076.20	0.04	0.06		-4075.80	0.06	0.04	0.03		-4075.80	0.04	0.83	0.00
TMP014133	0.02	-7361.03	0.02	0.00	-7359.92	0.00	0.03		-7355.90	0.03	0.00	0.00		-7355.65	2.22	10.26**	0.50
TMP014187	0.05	-28885.08	0.00	0.00	-28876.93	0.00	0.08		-28869.47	0.08	0.00	0.00		-28869.06	16.31**	31.23**	0.82

Note: * Significant ($p<0.05$; χ_{21}=3.84), ** Extremely significant ($p<0.01$; χ_{21}=6.63)

Genes presenting signatures of positive selection were found to be involved in structural protein function, translation, central intermediary metabolism, cell envelope and energy metabolism. Additionally, several positively selected sites were found to be located within protein domains, which further highlights their importance at the functional level (Appendix D).

It may therefore be elucidated that blood-feeding behavior may have played a role in shaping the evolution and the functional adaptations of these genes. This is evident from the positively selected sites identified specific to blood feeding insects. Such lineage-specific changes might be important for functional specificity and to maintain a new functional characteristic. Furthermore, non-blood feeding *IRE*-regulated genes with sites under selection include TMP003203 (DNA replication licensing factor *mcm7*), TMP003297 (*TRP* channel protein) and TMP007507 (topoisomerase3 beta) (Table 4.4). It is notable that no positive selection was identified on the sites of these genes in their blood-feeding counterparts. The significant difference in the number of positively selected sites between blood-feeders and non-blood feeders could be due to either: 1) a number of positively selected sites were missed when using PAML (Yang, 2007), 2) limited number of available sequences for non-blood feeders (Castric and Vekemans, 2007) and 3) the most likely reason may be the presence of different selection pressures between blood feeding and non-blood feeding insects.

Table 4.3: Parameter estimates and likelihood scores for the predicted *IRE*-regulated genes under site models in blood-feeding insects.

BF Ortholog Gene	l		dN/dS	LRT M0 vs. M3 (df=2)	PS sites	l		LRT M1 vs. M2 (df=2)	l		dN/dS	LRT M7 vs. M8 (df=4)	PS sites	No. of common PS sites between M3 and M8
	M0	M3				M1	M2		M7	M8				
TMP002520	-24279.01	-23882.86	16.08	792.31**	**276**,**390**,**809**,832,**869**,**877**,878,881,898,909,1155,1205,1206,1320,1323,1380,1381,1387,1402,1423,1467,1487,1500,1584,1928,1967,2094	-23998.16	-23998.16	0.00	-23923.65	-23888.61	50.92	70.09**	390,809,869,877,1205,1381	6
TMP002587	-1515.94	-1488.31	7.62	55.27**	-	-1509.20	-1509.20	0.00	-1489.37	-1489.37	5.15	0.00	-	-
TMP002821	-8199.05	-8109.10	57.64	179.91**	-	-8157.79	-8157.79	0.00	-8096.08	-8092.42	1.00	7.32	-	-
TMP003059	-3895.79	-3815.86	0.26	159.85**	-	-3871.62	-3871.62	0.00	-3815.33	-3815.24	1.00	0.18	-	-
TMP003060	-2375.61	-2343.41	10.13	64.40**	-	-2352.87	-2352.87	0.00	-2343.92	-2343.92	91.52	0.01	-	-
TMP003203	-8546.82	-8470.96	68.24	151.72**	**8**,30,**94**,**111**,**119**,**143**,196,**325**,**453**,**597**,674,699	-8459.91	-8459.91	0.00	-8409.32	-8401.40	58.99	15.84**	57,264	-
TMP003297	-8771.69	-8543.34	0.44	456.70**	-	-8594.25	-8594.25	0.00	-8546.92	-8544.85	185.86	4.13	-	-
TMP003500	-7677.52	-7538.23	0.33	278.57**	-	-7611.71	-7611.71	0.00	-7543.60	-7543.58	1.00	0.03	-	-
TMP003577	-6110.32	-5970.14	0.32	280.35**	-	-6034.43	-6034.43	0.00	-5974.09	-5970.97	1.00	6.24	-	-
TMP003602	-27953.52	-27312.18	0.41	1282.67**	-	-27530.17	-27530.17	0.00	-27333.89	-27310.69	1.00	46.40**	-	-
TMP004047	-14476.03	-14156.42	0.68	639.22**	-	-14233.09	-14233.09	0.00	-14167.99	-14157.62	3.97	20.73**	-	-
TMP004240	-4495.06	-4372.85	12.86	244.41**	-	-4387.61	-4387.61	0.00	-4371.58	-4371.58	12.59	0.00	-	-
TMP004289	-4715.85	-4679.55	0.04	72.60**	-	-4711.66	-4711.66	0.00	-4680.50	-4680.24	1.00	0.53	-	-
TMP004552	-23389.57	-22860.32	4.29	1058.51**	**10**,203,**208**,711,929,**1227**,1313,1322,1327,1475,1488,1516,1606,1614,1765,1926,2031,2064,2226	-22993.77	-22993.77	0.00	-22886.21	-22861.52	5.33	49.37**	10,208,1227,1327,1765,2031	6
TMP004553	-1422.77	-1412.10	38.38	21.33**	29,50,128	-1411.71	-1411.71	0.00	-1404.00	-1402.91	20.36	2.18	-	-
TMP005026	-10027.71	-9816.31	0.32	422.81**	-	-9878.53	-9878.53	0.00	-9830.70	-9818.20	1.00	24.99**	-	-
TMP005160	-5323.27	-5246.84	9.19	152.87**	105,126,153,159,174,175,176,**189**,190,193,237,238,353,**626**,627,629,**631**	-5269.45	-5269.45	0.00	-5266.34	-5247.49	8.55	37.70**	105,153,159,174,175,176,**189**,237,238,353,**626**,627,629,631	14
TMP005357	-3452.35	-3442.67	8.39	19.37**	211	-3437.67	-3437.67	0.00	-3375.42	-3373.33	4.86	4.18	199	-
TMP005511	-3295.92	-3254.18	0.07	83.48**	-	-3277.89	-3277.89	0.00	-3254.70	-3254.04	27.87	1.32	-	-
TMP005554	-9200.40	-8945.76	0.12	509.27**	-	-9045.12	-9045.12	0.00	-8949.68	-8948.22	1.00	2.92	-	-
TMP005595	-13249.79	-12978.00	0.34	543.58**	-	-13044.69	-13044.69	0.00	-12990.03	-12980.38	24.31	19.29**	-	-
TMP006315	-1483.46	-1483.46	44.20	0.00	-	-1482.57	-1482.57	0.00	-1477.75	-1477.75	1.00	0.00	-	-
TMP006377	-3359.89	-3312.64	1.08	94.51**	-	-3327.67	-3327.67	0.00	-3317.66	-3315.12	1.00	5.08	-	-
TMP006617	-3071.73	-2981.00	0.26	181.48**	-	-3016.41	-3016.41	0.00	-2982.46	-2981.54	58.34	1.85	-	-
TMP006868	-5004.77	-4875.85	5.60	257.85**	**23**,27,38,41,48,107,110,112,113,**116**,125,127,139,144,149,151,161,178,193,194,**199**,267,270,403,452	-4884.74	-4884.02	1.44	-4883.45	-4882.64	338.73	1.63	-	-
TMP006998	-6216.22	-6111.71	0.06	209.01**	-	-6171.04	-6171.04	0.00	-6117.64	-6115.95	1.00	3.40	-	
TMP007122	-3492.17	-3412.51	1.80	159.31**	105,107,151,**154**,171,184,225,235,236,239,248,262,266,**286**,298,308,334,**358**,360,362,365,**366**	-3424.33	-3424.33	0.00	-3417.04	-3414.30	1.82	5.48	107,154,236,262,286,358,366	7
TMP007507	-9520.51	-9378.64	11.03	283.75**	-	-9475.03	-9475.03	0.00	-9382.35	-9380.79	22.27	3.11	-	-
TMP007537	-14877.36	-14648.87	0.20	456.99**	-	-14791.92	-14791.92	0.00	-14660.22	-14650.72	1.00	19.00**	-	-
TMP008000	-4162.83	-4040.58	0.06	244.49**	-	-4089.51	-4089.51	0.00	-4045.51	-4045.51	20.61	0.00	-	-
TMP008747	-14928.85	-14509.31	0.67	839.07**	-	-14628.56	-14628.56	0.00	-14517.63	-14513.72	23.72	7.83	-	-
TMP008942	-15718.39	-15298.63	13.77	839.51**	410,427,**774**,783,801,**858**,862,868,**886**,894,910,**945**,**989**,998,1020,1026,1074,1083,1097,1122,1145,1146	-15402.74	-15402.61	0.27	-15325.68	-15293.52	14.09	64.32**	410,774,801,858,868,886,945,989,998,1074,1097,1122,1145	13
TMP009157	-1800.20	-1772.03	3.32	56.34**	-	-1797.13	-1797.13	0.00	-1772.68	-1772.68	1.00	0.00	-	-
TMP009871	-3297.00	-3276.06	56.84	41.88**	**68**,**69**,**103**,383	-3276.50	-3275.29	2.42	-3268.83	-3263.14	83.33	11.37*	70,103	1
TMP010016	-7688.81	-7556.30	0.22	265.03**	-	-7639.06	-7639.06	0.00	-7561.91	-7557.76	1.00	8.30	-	-
TMP010317	-6779.26	-6635.46	1.58	287.60**	165,**277**,**290**,**293**,367,368,**444**,472,473,485,487	-6693.32	-6693.32	0.00	-6649.11	-6634.61	2.09	29.01**	165,277,290,293,368,444,473,485,487	9
TMP010673	-2738.72	-2685.51	9.80	106.41**	33,**34**,46,47,133,144,148,152,204,206	-2699.62	-2698.56	2.13	-2695.36	-2686.17	9.40	18.38**	34,46,47,133,148,206	6
TMP010846	-4534.56	-4515.76	0.18	37.61**	-	-4526.66	-4526.66	0.00	-4517.55	-4517.29	1.00	0.52	-	-
TMP011104	-4296.00	-4231.66	11.43	128.68**	142,146	-4261.38	-4261.38	0.00	-4236.41	-4231.68	11.88	9.46	142	1
TMP011448	-6816.14	-6677.76	45.10	276.75**	63,648,**663**,**672**,675,794	-6703.03	-6703.02	0.01	-6686.44	-6676.98	58.99	18.91**	663,672,794	3
TMP011617	-5069.85	-4971.86	0.29	195.98**	-	-4987.32	-4987.32	0.00	-4981.15	-4972.34	1.00	17.62**	-	-
TMP011786	-5959.64	-5840.81	1.71	237.65**	404	-5886.11	-5886.11	0.00	-5843.59	-5840.28	2.28	6.61	-	-
TMP013137	-7739.54	-7623.35	0.03	232.40**	-	-7678.55	-7678.55	0.00	-7625.41	-7625.38	82.35	0.07	-	-
TMP013886	-3408.61	-3351.85	0.06	113.53**	-	-3378.90	-3378.90	0.00	-3349.13	-3346.71	15.64	4.85	-	-
TMP013948	-1391.75	-1363.26	1.08	56.98**	-	-1374.49	-1374.49	0.00	-1364.95	-1364.95	32.47	0.00	-	-
TMP014133	-2918.83	-2871.16	18.33	95.34**	-	-2894.35	-2894.35	0.00	-2870.95	-2870.95	21.71	0.00	-	-
TMP014187	-15530.49	-15310.73	0.12	439.53**	-	-15430.17	-15430.17	0.00	-15320.91	-15316.72	1.00	8.40	-	-

Note: * Significant ($p<0.05$; $\chi_{22}=5.99$ or $\chi_{24}=9.49$), ** Extremely significant ($p<0.01$; $\chi_{22}=9.21$ or $\chi_{24}=13.28$)

Table 4.4: Parameter estimates of likelihood scores for the predicted *IRE*-regulated genes under site models in Non-blood-feeding insects.

NBF ortholog Gene	l		dN/dS	LRT M0 vs. M3 (df=2)	PS sites	l		LRT M1 vs. M2 (df=2)	l		dN/dS	LRT M7 vs. M8 (df=4)	PS sites	No. of common PS Sites between M3 and M8
	M0	M3				M1	M2		M7	M8				
TMP002520	-14322.45	-14224.50	0.02	195.90**	-	-14225.29	-14225.29	0.00	-14225.55	-14225.55	38.36	0.00	-	-
TMP002587	-1687.72	-1650.57	2.03	74.29**	-	-1674.53	-1674.53	0.00	-1653.92	-1653.92	2.80	0.00	-	-
TMP002821	-6014.62	-5963.74	0.11	101.77**	-	-5987.58	-5987.58	0.00	-5967.12	-5967.12	20.81	0.00	-	-
TMP003059	-3527.81	-3447.06	72.36	161.49**	-	-3487.99	-3487.92	0.13	-3446.88	-3446.86	93.25	0.03	-	-
TMP003060	-3089.03	-3014.24	0.45	149.58**	-	-3047.52	-3047.52	0.00	-3015.89	-3013.74	1.00	4.29	-	-
TMP003203	-7736.75	-7602.31	2.68	268.89**	311,582	-7667.86	-7667.86	0.00	-7613.64	-7603.97	60.75	19.35**	311	1
TMP003297	-5258.80	-5189.95	1.14	137.69**	71,75,111,114,115,118,122,123,127,202,206,211,227,236,295,308,346,355,366,434,437,441,444,471,481,485,488,489,528,540,544,545,548,580,587,591,599,676,693,696,698,702,766,776,778,795,816,827,831,838	-5189.95	-5189.95	0.00	-5192.83	-5190.31	21.45	5.05	-	-
TMP003500	-6354.25	-6319.35	8.58	69.81**	16,19,70,111,131,152,215,220,296,412,536,601	-6319.46	-6319.35	0.21	-6310.55	-6307.24	14.61	6.63	-	-
TMP003577	-6240.23	-6163.57	0.03	153.32**	-	-6193.54	-6193.54	0.00	-6164.52	-6164.54	25.69	-0.03	-	
TMP003602	-21057.98	-20713.70	3.31	688.57**	771,113	-20848.27	-20848.27	0.00	-20729.85	-20717.01	60.92	25.68**	-	0
TMP004047	-11730.32	-11594.63	1.04	271.37**	-	-11611.52	-11611.52	0.00	-11596.21	-11596.17	134.01	0.08	-	-
TMP004240	-3181.30	-3119.71	0.76	123.18**	-	-3120.04	-3120.05	0.00	-3120.18	-3120.18	3.59	0.00	-	-
TMP004289	-3371.62	-3317.49	4.11	108.26**	-	-3321.77	-3321.77	0.00	-3317.95	-3317.93	403.52	0.04	-	-
TMP004552	-20135.00	-19874.54	14.98	520.94**	-	-19917.56	-19917.56	0.00	-19872.78	-19873.78	254.16	-2.01	-	-
TMP004553	-832.67	-824.42	15.61	16.50**	39,41,49,54,59,77,79,88,93	-824.75	-824.42	0.67	-824.89	-824.30	34.04	1.18	-	-
TMP005026	-5697.44	-5637.03	2.26	120.82**	-	-5642.61	-5642.61	0.00	-5638.91	-5638.91	179.53	0.01	-	-
TMP005160	-3106.68	-3074.62	188.00	64.12**	-	-3074.97	-3074.97	0.00	-3074.89	-3074.89	9.96	0.00	-	-
TMP005357	-1939.62	-1895.29	0.05	88.66**	-	-1895.46	-1895.46	0.00	-1895.73	-1895.73	581.00	-0.01	-	-
TMP005511	-2745.29	-2699.90	41.79	90.77**	-	-2738.57	-2738.57	0.00	-2700.82	-2700.82	56.90	0.00	-	-
TMP005554	-6990.01	-6864.09	1.40	251.83**	199	-6908.05	-6908.05	0.00	-6869.61	-6866.20	46.34	6.81	-	-
TMP005595	-13609.61	-13329.23	0.02	560.77**	-	-13454.42	-13454.42	0.00	-13332.14	-13332.14	1.00	0.00	-	-
TMP006315	-1003.98	-991.92	20.80	24.11**	-	-992.97	-992.98	0.00	-992.39	-992.38	208.96	0.03	-	-
TMP006377	-2303.39	-2279.55	10.79	47.68**	-	-2290.69	-2290.69	0.00	-2277.91	-2277.91	1.00	0.00	-	-
TMP006617	-3066.37	-3021.98	0.01	88.78**	-	-3053.45	-3053.45	0.00	-3021.26	-3017.89	1.00	6.74	-	-
TMP006868	-1715.61	-1691.93	0.80	47.36**	-	-1691.93	-1691.93	0.00	-1693.01	-1692.26	2.21	1.48	-	-
TMP006998	-5779.10	-5734.56	0.22	89.07**	-	-5740.04	-5740.04	0.00	-5738.92	-5735.22	3.53	7.41	-	-
TMP007122	-3109.66	-3094.54	0.92	30.24**	-	-3094.81	-3094.81	0.00	-3094.62	-3094.59	9.97	0.07	-	-
TMP007507	-9549.90	-9269.81	34.44	560.18**	32,141,144,248,275,469,763,781	-9349.67	-9349.67	0.00	-9277.46	-9266.84	34.73	21.25**	781	1
TMP007537	-6735.98	-6699.57	2.23	72.83**	498,741	-6699.57	-6699.57	0.00	-6701.36	-6701.36	83.76	-0.01	-	-
TMP008000	-4004.64	-3912.21	0.18	184.86**	-	-3944.79	-3944.79	0.00	-3913.97	-3913.54	1.00	0.87	-	-
TMP008747	-8768.92	-8652.36	2.42	233.12**	5,14,16,18,19,33,42,43,44,51,59,60,62,68,73,77,121,126,132,144,147,150,151,165,171,180,190,214,233,234,241,246,249,252,254,256,257,261,269,271,278,280,282,284,285,286,294,296,300,301,304,305,307,310,312,316,346,349,351,355,356,357,362,363,364,366,367,368,395,396,410,411,412,437,439,443,447,448,449,451,452,471,476,482,483,491,497,498,499,506,512,513,516,526,527,542,543,549,551,552,557,559,597,625,628,631,651,653,654,655,660,675,678,691,692,696,697,701,708,710,712,713,727,728,729,731,739,741,744,745,746,764,832,834,925,935,936,937,977,979,1007,1037,1103,1241,1243,1245,1251,1259,1278,1286,1299,1300,1303,1304,1307,1308,1310,1311	-8652.36	-8652.36	0.00	-8656.07	-8652.60	10.66	6.96	-	-
TMP008942	-15339.34	-14841.06	1.96	996.56**	429,770,778,928,929,948,958,993,1055,1059,1169,1193,1195,1197,1199,1437	-14966.12	-14966.12	0.00	-14849.26	-14836.54	2.44	25.44**	-	-
TMP009157	-1274.03	-1258.38	32.75	31.30**	-	-1271.18	-1271.17	0.00	-1258.39	-1258.40	76.89	-0.03	-	-
TMP009871	-3031.96	-3010.79	6.67	42.33**	-	-3026.14	-3026.14	0.00	-3012.98	-3012.98	37.62	0.00	-	-
TMP010016	-8779.00	-8659.56	0.05	238.88**	-	-8688.87	-8688.87	0.00	-8660.36	-8660.29	1.00	0.15	-	-
TMP010317	-6726.10	-6604.49	11.04	243.22**	272,357,436,442	-6669.11	-6669.11	0.00	-6616.70	-6606.82	9.63	19.75**	-	-
TMP010673	-1965.41	-1919.06	2.90	92.70**	130	-1923.83	-1923.83	0.00	-1920.74	-1919.57	26.58	2.33	-	-
TMP010846	-4773.51	-4723.49	0.11	100.05**	-	-4735.65	-4735.65	0.00	-4728.63	-4725.44	5.86	6.36	-	-
TMP011104	-3742.52	-3718.17	0.86	48.70**	-	-3729.08	-3729.08	0.00	-3719.01	-3719.01	77.37	0.00	-	-
TMP011448	-3534.45	-3516.56	61.23	35.79**	-	-3516.56	-3516.56	0.00	-3515.93	-3515.74	7.36	0.37	-	-
TMP011617	-2912.46	-2907.36	25.88	10.21**	374,407,422,432,438	-2907.75	-2907.36	0.78	-2907.41	-2907.41	80.22	0.00	-	-
TMP011786	-5386.51	-5276.43	38.10	220.17**	171	-5300.55	-5300.55	0.00	-5279.57	-5274.58	36.37	9.97**	-	-
TMP013137	-9446.63	-9348.47	1.15	196.33**	-	-9393.65	-9393.65	0.00	-9354.10	-9349.43	4.83	9.34	-	-
TMP013886	-3231.95	-3161.21	15.45	141.47**	106,299	-3192.87	-3192.87	0.00	-3164.11	-3160.57	1.16	7.10	-	-
TMP013948	-1704.75	-1649.64	0.90	110.21**	-	-1676.32	-1676.32	0.00	-1650.93	-1649.01	1.00	3.84	-	-
TMP014133	-3251.21	-3185.28	2.98	131.86**	-	-3221.56	-3221.56	0.00	-3191.23	-3189.16	3.41	4.16	-	-
TMP014187	-11145.16	-11015.70	0.06	258.92**	-	-11107.16	-11107.16	0.00	-11017.85	-11017.84	27.85	0.02	-	-

Note: * Significant ($p<0.05$; $\chi 22=5.99$ or $\chi 24=9.49$), ** Extremely significant ($p<0.01$; $\chi 22=9.21$ or $\chi 24=13.28$)

4.4 Conclusion

Despite being successful in nature, blood-feeding insects are constantly faced with the toxic effects of iron overload. Adaptation to such behavior is mainly to maintain fly's reproductive success, as iron is a key contributing factor. To address evolutionary changes that have allowed blood-feeding insects to compensate for iron toxicity, this chapter have identified the type of selection acting on *IRE*-regulated genes. As a result, the majority of these genes were experiencing functional constraints (*i.e.* subject to purifying selection). However, a subset of genes were identified to have experienced an increased rate of evolution, while several of these were recognized to contain sites under positive selection. Considering that positive selection allows for functional variation it maybe elucidated that functions of these genes have evolved such that they meet the blood-feeding and iron regulation needs of these insects.

Chapter 5

Concluding Remarks

Lack of conservation of the key players of human iron metabolism in insects indicates that iron regulation is conveyed by different mechanisms. As a co-factor in the kreb cycle and oxidative phosphorylation, iron is critical in the production of ATP for tsetse's energy requirements. Furthermore, the implications of iron in processes such as the immune response, maintenance of circadian rhythms as well as developmental and aging-related processes is well established (Mandilaras *et al.*, 2013).

Several lines of evidence exist in understanding the fate of blood meal iron in mosquitoes, however no such extensive work has been done in tsetse fly. Furthermore, research in understanding iron metabolism in insects has mostly focused on a few genes including ferritin (Pham and Chavez, 2005; Pham *et al.*, 2005; Dunkov *et al.*, 1995, 2002), transferrin (Harizanova *et al.*, 2005; Yoshiga *et al.*, 1997), iron regulatory protein 1 (Zhang *et al.*, 2002), and the divalent metal transporter1 (*DMT1*) (MartŠnez-Barnetche *et al.*, 2007). However, no regulatory network of iron metabolism have been given to date, which provides a unique perspective for the role of iron in insects biology and further understanding of how blood feeding insects utilize iron and direct it towards egg developments while preventing iron-induced toxicity. Such a broad knowledge base provides a ground for future research in the area of vector control.

Since the last review addressing iron metabolism in insects (Nichol and Winzerling, 2002), several insect vectors have been sequenced that accommodate more insect-specific investigations. In line with this, some of questions addressed in this study are summarized as follows: a) If *IRE*s controls the expression of iron responsive genes at the post-transcriptional level, what are the *IRE*-regulated genes in the genome of *Glossina morsitans*, b) What are the functional categories to which the *IRE*-regulated genes belong, c) What is the mechanism of iron transport and trafficking employed by tsetse flies, d) From the host-pathogen point of view what are the possible genes responsible for iron sequestration and iron-induced immune response, e) What are the principle genes responsible for the reproductive success of insects?, f) Is *IRE* the only element conserved in the UTRs of iron metabolism genes, or are there other factors at play that co-exert their effects on gene expression under iron deprivation and iron overload?, g) Having derived at a set of *IRE*-regulated genes, how does the timely expression of these genes contribute to the events occurring from 3h-15D following a blood meal? and finally h) What evolutionary processes have allowed for the adaptation of blood feeding insects to iron-rich blood meal. Addressing these questions have led to the identification of critical protein targets in the iron metabolism pathway as new candidates for the development of preventative measures protecting against the trypanosome infection. These can be summarized as 150 putative *IRE*-regulated genes in *Glossina morsitans*, that mostly function in transcription and translation regulation, and metabolism. Many of the identified genes have no known implications in iron metabolism and may be considered novel. Furthermore, besides ferritin heavy chain and *MRCK*-alpha, that are known to be under *IRE*-regulation, the rest of the proposed genes are novel with respect to their *IRE*-mediated mechanism of regulation at the post-transcriptional level.

We further demonstrate that iron trafficking and sequestration in Glossina is mainly achieved through the function of several genes out of which ferritin heavy chain, beta-carotene dioxygenase, serine-threonine protein kinase can be named. Beside these, many novel genes with no previously defined roles in iron metabolism are introduced, that could have evolved such that they meet the blood-feeding needs of hematophagous insects. Also the participation of a number of these genes in secretory pathways, further provides prob-

able targets for innovative control strategies. As such a new insight into iron metabolism is provided for Glossina and possibly other blood feeding insects.

Furthermore, our results attest to the tight co-regulation that exists between iron metabolism and reproduction processes in tsetse fly, with *UNR*-bs and *SXL*-bs being the over-represented elements in the UTRs of putative *IRE*-regulated genes. In light of recent findings where iron has been shown to regulate the activity of the miRNA pathway (Li *et al.*, 2012), our results, further support the link between iron regulation and miRNA activity, through the identification of miRNA binding motifs in the UTRs of several *IRE*-regulated genes in Glossina, including the *GY-box*, *Brd-box* and *K-box* motifs. Moreover, a detailed analysis of miRNAs in Glossina is underway, which provides the ground for further investigating the putative *IRE*-regulated genes as miRNAs targets.

The findings of this work, pertaining to the expression of *IRE*-regulated genes following the blood meal, indicate an overrepresented expression of transcription and translation, and amino acid biosynthesis genes. Moreover, we introduce *Unc-50* as an important transporter whose timely expression throughout 24H-15D PBM remarks its importance to mechanisms associated with iron trafficking. Likewise, several other *IRE*-regulated transport and cell envelope proteins were identified to up-regulate in response to blood meal, which allows for speculating an iron-induced expression of these genes.

Finally, our findings have provided an overview of the evolutionary forces that have allowed for the adaptation of blood feeding insects to iron-rich blood meal, whereby several variations were observed in the evolution of *IRE*-regulated genes, including genes experiencing an increased rate of evolution and those with sites under positive selection. These finding have allowed for speculating a functional adaptation of these genes in meeting the blood-feeding behavior of insects, hence their imperative role in orchestrating a network of events leading to iron metabolism and regulation.

The knowledge provided here allows for a better understanding of key players in the

acquisition, transport and storage of iron in *Glossina morsitans*.

Bibliography

Abascal, F., Zardoya, R., and Telford, M. J. (2010). Translatorx: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*, 38(Web Server issue):W7–13.

Abaza, I., Coll, O., Patalano, S., and Gebauer, F. (2006). Drosophila unr is required for translational repression of male-specific lethal 2 mrna during regulation of x-chromosome dosage compensation. *Genes Dev*, 20(3):380–389.

Abd-Alla, A. M., Cousserans, F., Parker, A. G., Jehle, J. A., Parker, N. J., Vlak, J. M., Robinson, A. S., and Bergoin, M. (2008). Genome analysis of a glossina pallidipes salivary gland hypertrophy virus reveals a novel, large, double-stranded circular dna virus. *Journal of virology*, 82(9):4595–4611.

Addess, K. J., Basilion, J. P., Klausner, R. D., Rouault, T. A., and Pardi, A. (1997). Structure and dynamics of the iron responsive element rna: implications for binding of the rna by iron regulatory binding proteins. *J Mol Biol*, 274(1):72–83.

Aisen, P., Enns, C., and Wessling-Resnick, M. (2001). Chemistry and biology of eukaryotic iron metabolism. *Int J Biochem Cell Biol*, 33(10):940–959.

Aksoy, S., Berriman, M., Hall, N., Hattori, M., Hide, W., and Lehane, M. J. (2005). A case for a glossina genome project. *Trends in parasitology*, 21(3):107–111.

Aksoy, S., Maudlin, I., Dale, C., Robinson, A. S., and O'Neill, S. L. (2001). Prospects for control of african trypanosomiasis by tsetse vector manipulation. *Trends Parasitol*, 17(1):29–35.

Allocco, D. J., Kohane, I. S., and Butte, A. J. (2004). Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics*, 5:18.

Alves-Silva, J., Ribeiro, J. M. C., and Van Den Abbeele.

Anderson, C. P., Shen, M., Eisenstein, R. S., and Leibold, E. A. (2012). Mammalian iron metabolism and its control by iron regulatory proteins. *Biochim Biophys Acta*, 1823(9):1468–1483.

Anderson, L. J., Wonke, B., Prescott, E., Holden, S., Malcolm Walker, J., and Pennell, D. J. (2002). Comparison of effects of oral deferiprone and subcutaneous desferrioxamine on myocardial iron concentrations and ventricular function in beta-thalassaemia. *The Lancet*, 360(9332):516–520.

Anisimova, M., Bielawski, J. P., and Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*, 18(8):1585–1592.

Apelqvist, Å., Li, H., Sommer, L., Beatus, P., Anderson, D. J., Honjo, T., de Angelis, M. H., Lendahl, U., and Edlund, H. (1999). Notch signalling controls pancreatic cell differentiation. *Nature*, 400(6747):877–881.

Arca, B., Lombardo, F., Valenzuela, J. G., Francischetti, I. M., Marinotti, O., Coluzzi, M., and Ribeiro, J. M. (2005). An updated catalogue of salivary gland transcripts in the adult female mosquito, anopheles gambiae. *Journal of experimental biology*, 208(20):3971–3986.

Archambaud, C., Gouin, E., Pizarro-Cerda, J., Cossart, P., and Dussurget, O. (2005). Translation elongation factor ef-tu is a target for stp, a serine-threonine phosphatase involved in virulence of listeria monocytogenes. *Mol Microbiol*, 56(2):383–396.

Artavanis-Tsakonas, S., Rand, M. D., and Lake, R. J. (1999). Notch signaling: cell fate control and signal integration in development. *Science*, 284(5415):770–776.

Attardo, G., Strickler-Dinglasan, P., Perkin, S., Caler, E., Bonaldo, M., Soares, M., El-Sayeed, N., and Aksoy, S. (2006). Analysis of fat body transcriptome from the adult tsetse fly, glossina morsitans morsitans. *Insect molecular biology*, 15(4):411–424.

Attardo, G. M., Ribeiro, J. M., Wu, Y., Berriman, M., and Aksoy, S. (2010). Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the tsetse fly (glossina morsitans morsitans). *BMC Genomics*, 11:160.

Aylmer, F. R. (1930). The genetical theory of natural selection.

Azar, D. and Nel, A. (2012). Evolution of hematophagy in non-biting midges (diptera: Chironomidae). *Terrestrial Arthropod Reviews*, 5(1):15–34.

Bahadorani, S. and Hilliker, A. J. (2009). Biological and behavioral effects of heavy metals in drosophila melanogaster adults and larvae. *Journal of insect behavior*, 22(5):399–411.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). Meme suite: tools for motif discovery and searching. *Nucleic acids research*, 37(suppl 2):W202–W208.

Bailey, T. L., Bodén, M., Whitington, T., and Machanick, P. (2010). The value of position-specific priors in motif discovery using meme. *BMC Bioinformatics*, 11:179.

Balashov, Y. S. (1984). Interaction between blood-sucking arthropods and their hosts, and its influence on vector potential. *Annual review of entomology*, 29(1):137–156.

Bandyopadhyay, J., Song, H.-O., Park, B.-J., Singaravelu, G., Sun, J. L., Ahnn, J., and Cho, J. H. (2009). Functional assessment of nramp-like metal transporters and manganese in caenorhabditis elegans. *Biochem Biophys Res Commun*, 390(1):136–141.

Barreau, C., Paillard, L., MŐreau, A., and Osborne, H. B. (2006). Mammalian celf/bruno-like rna-binding proteins: molecular characteristics and biological functions. *Biochimie*, 88(5):515–525.

Barton, H. A., Eisenstein, R. S., Bomford, A., and Munro, H. N. (1990). Determinants of the interaction between the iron-responsive element-binding protein and its binding site in rat l-ferritin mrna. *J Biol Chem*, 265(12):7000–7008.

Bashaw, G. J. and Baker, B. S. (1997). The regulation of the drosophila msl-2 gene reveals a function for sex-lethal in translational control. *Cell*, 89(5):789–798.

Beck, G., Ellis, T. W., Habicht, G. S., Schluter, S. F., and Marchalonis, J. J. (2002). Evolution of the acute phase response: iron release by echinoderm (asterias forbesi) coelomocytes, and cloning of an echinoderm ferritin molecule. *Dev Comp Immunol*, 26(1):11–26.

Beckmann, K., Grskovic, M., Gebauer, F., and Hentze, M. W. (2005). A dual inhibitory mechanism restricts msl-2 mrna translation for dosage compensation in drosophila. *Cell*, 122(4):529–540.

Becuwe, M., Vieira, N., Lara, D., Gomes-Rezende, J., Soares-Cunha, C., Casal, M., Haguenauer-Tsapis, R., Vincent, O., Paiva, S., and LÕon, S. (2012). A molecular switch on an arrestin-like protein relays glucose signaling to transporter endocytosis. *J Cell Biol*, 196(2):247–259.

Beerntsen, B. T., James, A. A., and Christensen, B. M. (2000). Genetics of mosquito vector competence. *Microbiol Mol Biol Rev*, 64(1):115–137.

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: Signalp 3.0. *J Mol Biol*, 340(4):783–795.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*, 125(1-2):279–284.

Bettany, A. J., Eisenstein, R. S., and Munro, H. N. (1992). Mutagenesis of the iron-regulatory element further defines a role for rna secondary structure in the regulation of ferritin and transferrin receptor expression. *J Biol Chem*, 267(23):16531–16537.

Bianchi, L., Tacchini, L., and Cairo, G. (1999). Hif-1-mediated activation of transferrin receptor gene transcription by iron chelation. *Nucleic Acids Res*, 27(21):4223–4227.

Bielawski, J. P. and Yang, Z. (2003). Maximum likelihood methods for detecting adaptive evolution after gene duplication. In *Genome Evolution*, pages 201–212. Springer.

Biswas, S. and Akey, J. M. (2006). Genomic insights into positive selection. *Trends Genet*, 22(8):437–446.

Biswas, S., Troy, H., Leek, R., Chung, Y.-L., Li, J.-L., Raval, R. R., Turley, H., Gatter, K., Pezzella, F., Griffiths, J. R., Stubbs, M., and Harris, A. L. (2010). Effects of hif-

1alpha and hif2alpha on growth and metabolism of clear-cell renal cell carcinoma 786-0 xenografts. *J Oncol*, 2010:757908.

Blobel, G. and Dobberstein, B. (1975). Transfer of proteins across membranes. i. presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol*, 67(3):835–851.

Boussadia, O., Niepmann, M., CrŐancier, L., Prats, A.-C., Dautry, F., and Jacquemin-Sablon, H. (2003). Unr is required in vivo for efficient initiation of translation from the internal ribosome entry sites of both rhinovirus and poliovirus. *J Virol*, 77(6):3353–3359.

Bouzaidi-Tiali, N., Aeby, E., CharriŔre, F., Pusnik, M., and Schneider, A. (2007). Elongation factor 1a mediates the specificity of mitochondrial trna import in t. brucei. *EMBO J*, 26(20):4302–4312.

Boveris, A., Oshino, N., and Chance, B. (1972). The cellular production of hydrogen peroxide. *Biochem J*, 128(3):617–630.

Brown, E. C. and Jackson, R. J. (2004). All five cold-shock domains of unr (upstream of n-ras) are required for stimulation of human rhinovirus rna translation. *J Gen Virol*, 85(Pt 8):2279–2287.

Browne, G. J. and Proud, C. G. (2002). Regulation of peptide-chain elongation in mammalian cells. *Eur J Biochem*, 269(22):5360–5368.

Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., , N. I. S. C. C. S. P., Green, E. D., Sidow, A., and Batzoglou, S. (2003). Lagan and multi-lagan: efficient tools for large-scale multiple alignment of genomic dna. *Genome Res*, 13(4):721–731.

Brun, R., Blum, J., Chappuis, F., and Burri, C. (2010). Human african trypanosomiasis. *Lancet*, 375(9709):148–159.

Bullen, J., Griffiths, E., Rogers, H., and Ward, G. (2000). Sepsis: the critical role of iron. *Microbes Infect*, 2(4):409–415.

Butt, J., Kim, H. Y., Basilion, J. P., Cohen, S., Iwai, K., Philpott, C. C., Altschul, S., Klausner, R. D., and Rouault, T. A. (1996). Differences in the rna binding sites of

iron regulatory proteins and potential target diversity. *Proc Natl Acad Sci U S A*, 93(9):4345–4349.

Caljon, G., Van Den Abbeele, J., Stijlemans, B., Coosemans, M., De Baetselier, P., and Magez, S. (2006). Tsetse fly saliva accelerates the onset of trypanosoma brucei infection in a mouse model associated with a reduced host inflammatory response. *Infect Immun*, 74(11):6324–6330.

Calvo, E., Pham, V. M., Lombardo, F., ArcÃă, B., and Ribeiro, J. M. C. (2006). The sialotranscriptome of adult male anopheles gambiae mosquitoes. *Insect Biochem Mol Biol*, 36(7):570–575.

Calvo, S. E., Pagliarini, D. J., and Mootha, V. K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A*, 106(18):7507–7512.

Campillos, M., Cases, I., Hentze, M. W., and Sanchez, M. (2010). Sires: searching for iron-responsive elements. *Nucleic Acids Res*, 38(Web Server issue):W360–W367.

Cao, W., Haig-Ladewig, L., Gerton, G. L., and Moss, S. B. (2006). Adenylate kinases 1 and 2 are part of the accessory structures in the mouse sperm flagellum. *Biol Reprod*, 75(4):492–500.

Castric, V. and Vekemans, X. (2007). Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene srk in brassicaceae? *BMC Evol Biol*, 7:132.

Chang, T.-C., Yamashita, A., Chen, C.-Y. A., Yamashita, Y., Zhu, W., Durdan, S., Kahvejian, A., Sonenberg, N., and Shyu, A.-B. (2004). Unr, a new partner of poly (a)-binding protein, plays a key role in translationally coupled mrna turnover mediated by the c-fos major coding-region determinant. *Genes & development*, 18(16):2010–2023.

Charlesworth, A., Cox, L. L., and MacNicol, A. M. (2004). Cytoplasmic polyadenylation element (cpe)- and cpe-binding protein (cpeb)-independent mechanisms regulate early class maternal mrna translational activation in xenopus oocytes. *J Biol Chem*, 279(17):17650–17659.

Charlesworth, B. (1993). Natural selection on multivariate traits in age-structured populations. *Proc Biol Sci*, 251(1330):47–52.

Choder, M. (2004). Rpb4 and rpb7: subunits of rna polymerase ii and beyond. *Trends Biochem Sci*, 29(12):674–681.

Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P. T., Collins, F. H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N. T., Hoffmann, J. A., Kanzok, S. M., Letunic, I., Levashina, E. A., Loukeris, T. G., Lycett, G., Meister, S., Michel, K., Moita, L. F., M§ller, H.-M., Osta, M. A., Paskewitz, S. M., Reichhart, J.-M., Rzhetsky, A., Troxler, L., Vernick, K. D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., and Kafatos, F. C. (2002). Immunity-related genes and gene families in anopheles gambiae. *Science*, 298(5591):159–165.

Condo, I., Malisan, F., Guccini, I., Serio, D., Rufini, A., and Testi, R. (2010). Molecular control of the cytosolic aconitase/irp1 switch by extramitochondrial frataxin. *Hum Mol Genet*, 19(7):1221–1229.

Corradini, E., Rozier, M., Meynard, D., Odhiambo, A., Lin, H. Y., Feng, Q., Migas, M. C., Britton, R. S., Babitt, J. L., and Fleming, R. E. (2011). Iron regulation of hepcidin despite attenuated smad1,5,8 signaling in mice without transferrin receptor 2 or hfe. *Gastroenterology*, 141(5):1907–1914.

Cross, T. G., Scheel-Toellner, D., Henriquez, N. V., Deacon, E., Salmon, M., and Lord, J. M. (2000). Serine/threonine protein kinases and apoptosis. *Exp Cell Res*, 256(1):34–41.

Dandekar, T., Stripecke, R., Gray, N. K., Goossen, B., Constable, A., Johansson, H. E., and Hentze, M. W. (1991). Identification of a novel iron-responsive element in murine and human erythroid delta-aminolevulinic acid synthase mrna. *EMBO J*, 10(7):1903–1909.

Darby, N. J., Freedman, R. B., and Creighton, T. E. (1994). Dissecting the mechanism of protein disulfide isomerase: catalysis of disulfide bond formation in a model peptide. *Biochemistry*, 33(25):7937–7947.

Day, J. F. (2005). Host-seeking strategies of mosquito disease vectors. *J Am Mosq Control Assoc*, 21(4 Suppl):17–22.

de Moor, C. H. and Richter, J. D. (1999). Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin b1 mrna. *EMBO J*, 18(8):2294–2303.

Dong, Y., Taylor, H. E., and Dimopoulos, G. (2006). Agdscam, a hypervariable immunoglobulin domain-containing receptor of the anopheles gambiae innate immune system. *PLoS Biol*, 4(7):e229.

Dormoy-Raclet, V., Markovits, J., Jacquemin-Sablon, A., and Jacquemin-Sablon, H. (2005). Regulation of unr expression by 5'- and 3'-untranslated regions of its mrna through modulation of stability and ires mediated translation. *RNA Biol*, 2(3):e27–e35.

Dottorini, T., Persampieri, T., Palladino, P., Spaccapelo, R., and Crisanti, A. (2012). Silencing of the hsf gene, the transcriptional regulator of a. gambiae male accessory glands, inhibits the formation of the mating plug in mated females and disrupts their monogamous behaviour. *Pathog Glob Health*, 106(7):405–412.

Drennan, M. B., Stijlemans, B., Van Den Abbeele, J., Quesniaux, V. J., Barkhuizen, M., Brombacher, F., De Baetselier, P., Ryffel, B., and Magez, S. (2005). The induction of a type 1 immune response following a trypanosoma brucei infection is myd88 dependent. *The Journal of Immunology*, 175(4):2501–2509.

Dunkov, B. and Georgieva, T. (2006). Insect iron binding proteins: insights from the genomes. *Insect biochemistry and molecular biology*, 36(4):300–309.

Dunkov, B. C., Georgieva, T., Yoshiga, T., Hall, M., and Law, J. H. (2002). Aedes aegypti ferritin heavy chain homologue: feeding of iron or blood influences message levels, lengths and subunit abundance. *J Insect Sci*, 2:7.

Dunkov, B. C., Zhang, D., Choumarov, K., Winzerling, J. J., and Law, J. H. (1995). Isolation and characterization of mosquito ferritin and cloning of a cdna that encodes one subunit. *Arch Insect Biochem Physiol*, 29(3):293–307.

Dupuy, J., Volbeda, A., Carpentier, P., Darnault, C., Moulis, J.-M., and Fontecilla-Camps, J. C. (2006). Crystal structure of human iron regulatory protein 1 as cytosolic aconitase. *Structure*, 14(1):129–139.

Edgar, R. C. (2004). Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–1797.

Eisenstein, R. S. (2000). Iron regulatory proteins and the molecular control of mammalian iron metabolism. *Annu Rev Nutr*, 20:627–662.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using targetp, signalp and related tools. *Nat Protoc*, 2(4):953–971.

Farago, M., Nahari, T., Hammel, C., Cole, C. N., and Choder, M. (2003). Rpb4p, a subunit of rna polymerase ii, mediates mrna export during stress. *Mol Biol Cell*, 14(7):2744–2755.

Felsenstein, J. (1981). Evolutionary trees from dna sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6):368–376.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual review of genetics*, 22(1):521–565.

FernĞndez-ChacŮn, R., KŽnigstorfer, A., Gerber, S. H., GarcŠa, J., Matos, M. F., Stevens, C. F., Brose, N., Rizo, J., Rosenmund, C., and S§dhof, T. C. (2001). Synaptotagmin i functions as a calcium regulator of release probability. *Nature*, 410(6824):41–49.

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Biology*, 20(4):406–416.

Fleming, M. D., Trenor, 3rd, C., Su, M. A., Foernzler, D., Beier, D. R., Dietrich, W. F., and Andrews, N. C. (1997). Microcytic anaemia mice have a mutation in nramp2, a candidate iron transporter gene. *Nat Genet*, 16(4):383–386.

Folwell, J. L., Barton, C. H., and Shepherd, D. (2006). Immunolocalisation of the d. melanogaster nramp homologue malvolio to gut and malpighian tubules provides evidence that malvolio and nramp2 are orthologous. *J Exp Biol*, 209(Pt 10):1988–1995.

Forbes, J. R. and Gros, P. (2001). Divalent-metal transport by nramp proteins at the interface of host-pathogen interactions. *Trends Microbiol*, 9(8):397–403.

Ganz, T., Goff, J., Klasing, K., Nemeth, E., and Roth, T. (2012). Iod in rhinos–immunity group report: report from the immunity, genetics and toxicology working group of the international workshop on iron overload disorder in browsing rhinoceros (february 2011). *J Zoo Wildl Med*, 43(3 Suppl):S117–S119.

Gebauer, F., Corona, D. F., Preiss, T., Becker, P. B., and Hentze, M. W. (1999). Translational control of dosage compensation in drosophila by sex-lethal: cooperative silencing via the 5 and 3 utrs of msl-2 mrna is independent of the poly (a) tail. *The EMBO journal*, 18(21):6146–6154.

Gebauer, F., Grskovic, M., and Hentze, M. W. (2003). Drosophila sex-lethal inhibits the stable association of the 40s ribosomal subunit with msl-2 mrna. *Molecular cell*, 11(5):1397–1404.

Geiser, D. L. and Winzerling, J. J. (2012). Insect transferrins: multifunctional proteins. *Biochim Biophys Acta*, 1820(3):437–451.

Geiser, D. L., Zhang, D., and Winzerling, J. J. (2006). Secreted ferritin: mosquito defense against iron overload? *Insect Biochem Mol Biol*, 36(3):177–187.

Gelbart, M. E. and Kuroda, M. I. (2009). Drosophila dosage compensation: a complex voyage to the x chromosome. *Development*, 136(9):1399–1410.

Glanfield, A., McManus, D. P., Anderson, G. J., and Jones, M. K. (2007). Pumping iron: a potential target for novel therapeutics against schistosomes. *Trends Parasitol*, 23(12):583–588.

Goldman, N. and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding dna sequences. *Molecular biology and evolution*, 11(5):725–736.

Gooding, R. H. and Krafsur, E. S. (2005). Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. *Annu Rev Entomol*, 50:101–123.

Graca-Souza, A. V., Maya-Monteiro, C., Paiva-Silva, G. O., Braz, G. R., Paes, M. C., Sorgine, M. H., Oliveira, M. F., and Oliveira, P. L. (2006). Adaptations against

heme toxicity in blood-feeding arthropods. *Insect biochemistry and molecular biology*, 36(4):322–335.

Graindorge, A., Militti, C., and Gebauer, F. (2011). Posttranscriptional control of x-chromosome dosage compensation. *Wiley Interdiscip Rev RNA*, 2(4):534–545.

Gray, N. K. and Wickens, M. (1998). Control of translation initiation in animals. *Annu Rev Cell Dev Biol*, 14:399–458.

Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V. A., Horner, D. S., Pavesi, G., Picardi, E., *et al.* (2010). Utrdb and utrsite (release 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mrnas. *Nucleic acids research*, 38(suppl 1):D75–D80.

Gross, S. R. and Kinzy, T. G. (2005). Translation elongation factor 1a is essential for regulation of the actin cytoskeleton and cell morphology. *Nat Struct Mol Biol*, 12(9):772–778.

Gross, S. R. and Kinzy, T. G. (2007). Improper organization of the actin cytoskeleton affects protein synthesis at initiation. *Mol Cell Biol*, 27(5):1974–1989.

Gruber, A. R., Bernhart, S. H., Hofacker, I. L., and Washietl, S. (2008). Strategies for measuring evolutionary conservation of rna secondary structures. *BMC Bioinformatics*, 9:122.

Gundersen, G. G. and Cook, T. A. (1999). Microtubules and signal transduction. *Curr Opin Cell Biol*, 11(1):81–94.

Gunshin, H., Mackenzie, B., Berger, U. V., Gunshin, Y., Romero, M. F., Boron, W. F., Nussberger, S., Gollan, J. L., and Hediger, M. A. (1997). Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature*, 388(6641):482–488.

Guz, N., Attardo, G. M., Wu, Y., and Aksoy, S. (2007). Molecular aspects of transferrin expression in the tsetse fly (glossina morsitans morsitans). *J Insect Physiol*, 53(7):715–723.

Hamady, M., Widmann, J., Copley, S. D., Knight, R., *et al.* (2008). Motifcluster: an interactive online tool for clustering and visualizing sequences using shared motifs. *Genome biology*, 9(8):R128.

Hao, Z., Kasumba, I., and Aksoy, S. (2003). Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (diptera: Glossinidiae). *Insect Biochem Mol Biol*, 33(11):1155–1164.

Hao, Z., Kasumba, I., Lehane, M. J., Gibson, W. C., Kwon, J., and Aksoy, S. (2001). Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proceedings of the National Academy of Sciences*, 98(22):12648–12653.

Harizanova, N., Georgieva, T., Dunkov, B. C., Yoshiga, T., and Law, J. H. (2005). Aedes aegypti transferrin. gene structure, expression pattern, and regulation. *Insect Mol Biol*, 14(1):79–88.

Hathout, Y. (2007). Approaches to the study of the cell secretome. *Expert Rev Proteomics*, 4(2):239–248.

Henderson, B. R., Menotti, E., Bonnard, C., and K§hn, L. C. (1994). Optimal sequence and structure of iron-responsive elements. selection of rna stem-loops with high affinity for iron regulatory factor. *J Biol Chem*, 269(26):17481–17489.

Hentze, M. W., Caughman, S. W., Rouault, T. A., Barriocanal, J. G., Dancis, A., Harford, J. B., and Klausner, R. D. (1987). Identification of the iron-responsive element for the translational regulation of human ferritin mrna. *Science*, 238(4833):1570–1573.

Hentze, M. W. and kuhn, L. C. (1996). Molecular control of vertebrate iron metabolism: mrna-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc Natl Acad Sci U S A*, 93(16):8175–8182.

Hentze, M. W., Muckenthaler, M. U., and Andrews, N. C. (2004). Balancing acts: molecular control of mammalian iron metabolism. *Cell*, 117(3):285–297.

Hirai, H., Maru, Y., Hagiwara, K., Nishida, J., and Takaku, F. (1987). A novel putative tyrosine kinase receptor encoded by the eph gene. *Science*, 238(4834):1717–1720.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., and Nakai, K. (2007). Wolf psort: protein localization predictor. *Nucleic Acids Res*, 35(Web Server issue):W585–W587.

Hower, V., Mendes, P., Torti, F. M., Laubenbacher, R., Akman, S., Shulaev, V., and Torti, S. V. (2009). A general map of iron metabolism and tissue-specific subnetworks. *Molecular bioSystems*, 5(5):422–443.

Hubbi, M. E., Luo, W., Baek, J. H., and Semenza, G. L. (2011). Mcm proteins are negative regulators of hypoxia-inducible factor 1. *Molecular cell*, 42(5):700–712.

Huelsenbeck, J. P., Ronquist, F., *et al.* (2001). Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755.

Hughes, A. L. and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class i loci reveals overdominant selection.

Hunt, S. L., Hsuan, J. J., Totty, N., and Jackson, R. J. (1999). unr, a cellular cytoplasmic rna-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus rna. *Genes & development*, 13(4):437–448.

Iacopetta, B. J. and Morgan, E. H. (1983). Transferrin endocytosis and iron uptake during erythroid cell development. *Biomed Biochim Acta*, 42(11-12):S182–S186.

Iengar, P. and Joshi, N. V. (2009). Identification of putative regulatory motifs in the upstream regions of co-expressed functional groups of genes in plasmodium falciparum. *BMC Genomics*, 10:18.

Islam, T. C. and Smith, C. I. (2000). The cellular phenotype conditions btk for cell survival or apoptosis signaling. *Immunol Rev*, 178:49–63.

Jareborg, N., Birney, E., and Durbin, R. (1999). Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res*, 9(9):815–824.

Jensen, J. (2004). Gene regulatory factors in pancreatic development. *Dev Dyn*, 229(1):176–200.

Jensen, L. J., Gupta, R., Staerfeldt, H.-H., and Brunak, S. (2003). Prediction of human protein function according to gene ontology categories. *Bioinformatics*, 19(5):635–642.

Jeong, S. Y. and David, S. (2003). Glycosylphosphatidylinositol-anchored ceruloplasmin is required for iron efflux from cells in the central nervous system. *J Biol Chem*, 278(29):27144–27148.

Johnson, J. P., Bar-Eli, M., Jansen, B., and Markhof, E. (1997). Melanoma progression-associated glycoprotein muc18/mcam mediates homotypic cell adhesion through interaction with a heterophilic ligand. *Int J Cancer*, 73(5):769–774.

Jones, E. W. and Fink, G. R. (1982). Regulation of amino acid and nucleotide biosynthesis in yeast. *Cold Spring Harbor Monograph Archive*, 11:181–299.

Ke, Y., Sierzputowska-Gracz, H., Gdaniec, Z., and Theil, E. C. (2000). Internal loop/bulge and hairpin loop of the iron-responsive element of ferritin mrna contribute to maximal iron regulatory protein 2 binding and translational regulation in the iso-iron-responsive element/iso-iron regulatory protein family. *Biochemistry*, 39(20):6235–6242.

Ke, Y., Wu, J., Leibold, E. A., Walden, W. E., and Theil, E. C. (1998). Loops and bulge/loops in iron-responsive element isoforms influence iron regulatory protein binding. fine-tuning of mrna regulation? *J Biol Chem*, 273(37):23637–23640.

Kell, D. B. (2009). Iron behaving badly: inappropriate iron chelation as a major contributor to the aetiology of vascular and other progressive inflammatory and degenerative diseases. *BMC medical genomics*, 2(1):2.

Kelleher, J. F. and Titus, M. A. (1998). Intracellular motility: how can we all work together? *Curr Biol*, 8(11):R394–R397.

Khacho, M., Mekhail, K., Pilon-Larose, K., Pause, A., CŹtŐ, J., and Lee, S. (2008). eef1a is a novel component of the mammalian nuclear protein export machinery. *Mol Biol Cell*, 19(12):5296–5308.

Khazak, V., Sadhale, P. P., Woychik, N. A., Brent, R., and Golemis, E. A. (1995). Human rna polymerase ii subunit hsrpb7 functions in yeast and influences stress survival and cell morphology. *Mol Biol Cell*, 6(7):759–775.

Kikinis, Z., Eisenstein, R. S., Bettany, A. J., and Munro, H. N. (1995). Role of rna secondary structure of the iron-responsive element in translational regulation of ferritin synthesis. *Nucleic Acids Res*, 23(20):4190–4195.

Kim-Ha, J., Kerr, K., and Macdonald, P. M. (1995). Translational regulation of oskar mrna by bruno, an ovarian rna-binding protein, is essential. *Cell*, 81(3):403–412.

Kimura, K., Cuvier, O., and Hirano, T. (2001). Chromosome condensation by a human condensin complex inxenopus egg extracts. *Journal of Biological Chemistry*, 276(8):5417–5420.

Klee, E. W. and Ellis, L. B. M. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics*, 6:256.

Kluge, A. G. and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Biology*, 18(1):1–32.

Knutson, M. D. (2010). Iron-sensing proteins that regulate hepcidin and enteric iron absorption. *Annu Rev Nutr*, 30:149–171.

Kohler, S. A., Henderson, B. R., and Kühn, L. C. (1995). Succinate dehydrogenase b mrna of drosophila melanogaster has a functional iron-responsive element in its 5'-untranslated region. *J Biol Chem*, 270(51):30781–30786.

Konofal, E., Lecendreux, M., Arnulf, I., and Mouren, M.-C. (2004). Iron deficiency in children with attention-deficit/hyperactivity disorder. *Arch Pediatr Adolesc Med*, 158(12):1113–1115.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*, 39:309–338.

Kumar, S., Christophides, G. K., Cantera, R., Charles, B., Han, Y. S., Meister, S., Dimopoulos, G., Kafatos, F. C., and Barillas-Mury, C. (2003). The role of reactive oxygen species on plasmodium melanotic encapsulation in anopheles gambiae. *Proc Natl Acad Sci U S A*, 100(24):14139–14144.

Kumar, S., Tamura, K., and Nei, M. (1994). Mega: Molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci*, 10(2):189–191.

Kuriyama-Matsumura, K., Sato, H., Suzuki, M., and Bannai, S. (2001). Effects of hyperoxia and iron on iron regulatory protein-1 activity and the ferritin synthesis in mouse peritoneal macrophages. *Biochim Biophys Acta*, 1544(1-2):370–377.

Kyba, M. and Brock, H. W. (1998). The sam domain of polyhomeotic, rae28, and scm mediates specific interactions through conserved residues. *Dev Genet*, 22(1):74–84.

Lai, E. C., Burks, C., and Posakony, J. W. (1998). The k box, a conserved 3 utr sequence motif, negatively regulates accumulation of enhancer of split complex transcripts. *Development*, 125(20):4077–4088.

Lai, E. C., Tam, B., and Rubin, G. M. (2005). Pervasive regulation of drosophila notch target genes by gy-box-, brd-box-, and k-box-class microrna. *Genes Dev*, 19(9):1067–1080.

Lammert, E., Brown, J., and Melton, D. A. (2000). Notch gene expression during pancreatic organogenesis. *Mechanisms of development*, 94(1):199–203.

Lavoie, B. D., Tuffo, K. M., Oh, S., Koshland, D., and Holm, C. (2000). Mitotic chromosome condensation requires brn1p, the yeast homologue of barren. *Mol Biol Cell*, 11(4):1293–1304.

Laxman, S. and Beavo, J. A. (2007). Cyclic nucleotide signaling mechanisms in trypanosomes: possible targets for therapeutic agents. *Mol Interv*, 7(4):203–215.

Lehane, M., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J., Soares, M., Bonaldo, M., Lehane, S., Hall, N., *et al.* (2003). Adult midgut expressed sequence tags from the tsetse fly glossina morsitans morsitans and expression analysis of putative immune response genes. *Genome Biol*, 4(10):R63.

Lehane, M. J. (2005). *The biology of blood-sucking in insects*. Cambridge University Press.

Lehane, M. J., Gibson, W., and Lehane, S. M. (2008). Differential expression of fat body genes in glossina morsitans morsitans following infection with trypanosoma brucei brucei. *Int J Parasitol*, 38(1):93–101.

Leibold, E. A., Laudano, A., and Yu, Y. (1990). Structural requirements of iron-responsive elements for binding of the protein involved in both transferrin receptor and ferritin mrna post-transcriptional regulation. *Nucleic Acids Res*, 18(7):1819–1824.

Leipuviene, R. and Theil, E. C. (2007). The family of iron responsive rna structures regulated by changes in cellular iron and oxygen. *Cell Mol Life Sci*, 64(22):2945–2955.

Lenaz, G., Bovina, C., D'Aurelio, M., Fato, R., Formiggini, G., Genova, M. L., Giuliano, G., Merlo Pich, M., Paolucci, U., Parenti Castelli, G., and Ventura, B. (2002). Role of mitochondria in oxidative stress and aging. *Ann N Y Acad Sci*, 959:199–213.

Li, Y., Lin, L., Li, Z., Ye, X., Xiong, K., Aryal, B., Xu, Z., Paroo, Z., Liu, Q., He, C., and et al. (2012). Iron homeostasis regulates the activity of the microrna pathway through poly(c)-binding protein 2. *Cell Metabolism*, 15(6):895–904.

Lill, R. and Malenhoff, U. (2008). Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annu Rev Biochem*, 77:669–700.

Lim, J. E., Jin, O., Bennett, C., Morgan, K., Wang, F., Trenor, C. C., Fleming, M. D., and Andrews, N. C. (2005). A mutation in sec15l1 causes anemia in hemoglobin deficit (hbd) mice. *Nature genetics*, 37(11):1270–1273.

Liu, H., Le Trinh, T., Dong, H., Keith, R., Nelson, D., and Liu, C. (2012). Iron regulator hepcidin exhibits antiviral activity against hepatitis c virus. *PloS one*, 7(10):e46631.

Liu, X. and Theil, E. C. (2005). Ferritin as an iron concentrator and chelator target. *Ann N Y Acad Sci*, 1054:136–140.

Locke, M. and Nichol, H. (1992). Iron economy in insects: transport, metabolism, and storage. *Annual review of entomology*, 37(1):195–215.

Luitjens, C., Gallegos, M., Kraemer, B., Kimble, J., and Wickens, M. (2000). Cpeb proteins control two key steps in spermatogenesis in c. elegans. *Genes Dev*, 14(20):2596–2609.

Lupo, R., Breiling, A., Bianchi, M. E., and Orlando, V. (2001). Drosophila chromosome condensation proteins topoisomerase ii and barren colocalize with polycomb and maintain fab-7 pre silencing. *Mol Cell*, 7(1):127–136.

MacLeod, A., Tait, A., and Turner, C. M. R. (2001). The population genetics of trypanosoma brucei and the origin of human infectivity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1411):1035–1044.

MacLeod, E., Maudlin, I., Darby, A., Welburn, S., *et al.* (2007). Antioxidants promote establishment of trypanosome infections in tsetse. *Parasitology*, 134(6):827.

Mandilaras, K., Pathmanathan, T., and Missirlis, F. (2013). Iron absorption in drosophila melanogaster. *Nutrients*, 5(5):1622–1647.

Marinotti, O., Calvo, E., Nguyen, Q. K., Dissanayake, S., Ribeiro, J. M. C., and James, A. A. (2006). Genome-wide analysis of gene expression in adult anopheles gambiae. *Insect Mol Biol*, 15(1):1–12.

Maroni, G. (1993). *An Atlas of Drosophila Genes: Sequences and Molecular Features: Sequences and Molecular Features*. Oxford University Press, USA.

MartŠnez-Barnetche, J., GarcŠa Solache, M., Neri Lecona, A., Tello LŮpez, A. T., del Carmen RodrŠguez, M., Gamba, G., VĞzquez, N., RodrŠguez, M. H., and Lanz-Mendoza, H. (2007). Cloning and functional characterization of the anopheles albimanus dmt1/nramp homolog: implications in iron metabolism in mosquitoes. *Insect Biochem Mol Biol*, 37(6):532–539.

Marx, J. J. M. (2002). Iron and infection: competition between host and microbes for a precious element. *Best Pract Res Clin Haematol*, 15(2):411–426.

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

Maudlin, I. and Welburn, S. C. (1987). Lectin mediated establishment of midgut infections of trypanosoma congolense and trypanosoma brucei in glossina morsitans. *Trop Med Parasitol*, 38(3):167–170.

Mazumder, B., Sampath, P., and Fox, P. L. (2005). Regulation of macrophage ceruloplasmin gene expression: one paradigm of 3'-utr-mediated translational control. *Mol Cells*, 20(2):167–172.

Mazumder, B., Sampath, P., and Fox, P. L. (2006). Translational control of ceruloplasmin gene expression: beyond the ire. *Biol Res*, 39(1):59–66.

Mazumder, B., Sampath, P., Seshadri, V., Maitra, R. K., DiCorleto, P. E., and Fox, P. L. (2003). Regulated release of l13a from the 60s ribosomal subunit as a mechanism of transcript-specific translational control. *Cell*, 115(2):187–198.

McKie, A. T. and Barlow, D. J. (2004). The slc40 basolateral iron transporter family (ireg1/ferroportin/mtp1). *Pflugers Arch*, 447(5):801–806.

McKie, A. T., Marciani, P., Rolfs, A., Brennan, K., Wehr, K., Barrow, D., Miret, S., Bomford, A., Peters, T. J., Farzaneh, F., Hediger, M. A., Hentze, M. W., and Simpson, R. J. (2000). A novel duodenal iron-regulated transporter, ireg1, implicated in the basolateral transfer of iron to the circulation. *Mol Cell*, 5(2):299–309.

Medenbach, J., Seiler, M., and Hentze, M. W. (2011). Translational control via protein-regulated upstream open reading frames. *Cell*, 145(6):902–913.

Meehan, H. A. and Connell, G. J. (2001). The hairpin loop but not the bulged c of the iron responsive element is essential for high affinity binding to iron regulatory protein-1. *Journal of Biological Chemistry*, 276(18):14791–14796.

Mignone, F., Gissi, C., Liuni, S., Pesole, G., *et al.* (2002). Untranslated regions of mrnas. *Genome Biol*, 3(3):1–10.

Min, X. J. (2010). Evaluation of computational methods for secreted protein prediction in different eukaryotes. *Journal of Proteomics & Bioinformatics*, 3.

Missirlis, F., Kosmidis, S., Brody, T., Mavrakis, M., Holmberg, S., Odenwald, W. F., Skoulakis, E. M. C., and Rouault, T. A. (2007). Homeostatic mechanisms for iron storage revealed by genetic manipulations and live imaging of drosophila ferritin. *Genetics*, 177(1):89–100.

Mittal, R. and McMahon, H. T. (2009). Arrestins as adaptors for ubiquitination in endocytosis and sorting. *EMBO Rep*, 10(1):41–43.

Morris, D. R. and Geballe, A. P. (2000). Upstream open reading frames as regulators of mrna translation. *Molecular and Cellular Biology*, 20(23):8635–8642.

Muckenthaler, M., Gunkel, N., Frishman, D., Cyrklaff, A., Tomancak, P., and Hentze, M. W. (1998). Iron-regulatory protein-1 (irp-1) is highly conserved in two invertebrate

species–characterization of irp-1 homologues in drosophila melanogaster and caenorhabditis elegans. *Eur J Biochem*, 254(2):230–237.

Muckenthaler, M., Richter, A., Gunkel, N., Riedel, D., Polycarpou-Schwarz, M., Hentze, S., Falkenhahn, M., Stremmel, W., Ansorge, W., and Hentze, M. W. (2003). Relationships and distinctions in iron-regulatory networks responding to interrelated signals. *Blood*, 101(9):3690–3698.

Muckenthaler, M. U., Galy, B., and Hentze, M. W. (2008). Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (ire/irp) regulatory network. *Annu Rev Nutr*, 28:197–213.

Munks, R. J. L., Sant'Anna, M. R. V., Grail, W., Gibson, W., Igglesden, T., Yoshiyama, M., Lehane, S. M., and Lehane, M. J. (2005). Antioxidant gene expression in the blood-feeding fly glossina morsitans morsitans. *Insect Mol Biol*, 14(5):483–491.

Muñoz, M., Villar, I., and García-Erce, J. A. (2009). An update on iron physiology. *World journal of gastroenterology: WJG*, 15(37):4617.

Muse, S. V. and Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, 11(5):715–724.

Nair, R. and Rost, B. (2005). Mimicking cellular sorting improves prediction of subcellular localization. *Journal of molecular biology*, 348(1):85–100.

Napier, I., Ponka, P., and Richardson, D. R. (2005). Iron trafficking in the mitochondrion: novel pathways revealed by disease. *Blood*, 105(5):1867–1874.

Nichol, H. and Winzerling, J. (2002). Structured rna upstream of insect cap distal iron responsive elements enhances iron regulatory protein-mediated control of translation. *Insect Biochem Mol Biol*, 32(12):1699–1710.

Nielsen, R. and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the hiv-1 envelope gene. *Genetics*, 148(3):929–936.

Ohler, U., Yekta, S., Lim, L. P., Bartel, D. P., and Burge, C. B. (2004). Patterns of flanking sequence conservation and a characteristic upstream motif for microrna gene identification. *RNA*, 10(9):1309–1322.

Olivares, A., Méndez, J. P., Cárdenas, M., Oviedo, N., Palomino, M. A., Santos, I., Perera-Marín, G., Gutiérrez-Sagal, R., and Ulloa-Aguirre, A. (2009). Pituitary-testicular axis function, biological to immunological ratio and charge isoform distribution of pituitary lh in male rats with experimental diabetes. *Gen Comp Endocrinol*, 161(3):304–312.

Ong, S. T., Ho, J. Z. S., Ho, B., and Ding, J. L. (2006). Iron-withholding strategy in innate immunity. *Immunobiology*, 211(4):295–314.

Oswald, F., TŁuber, B., Dobner, T., Bourteele, S., Kostezka, U., Adler, G., Liptay, S., and Schmid, R. M. (2001). p300 acts as a transcriptional coactivator for mammalian notch-1. *Mol Cell Biol*, 21(22):7761–7774.

Page, R. D. and Holmes, E. C. (2009). *Molecular evolution: a phylogenetic approach*. Wiley-Blackwell.

Panasyuk, G., Nemazanyy, I., Filonenko, V., Negrutskii, B., and El'skaya, A. V. (2008). A2 isoform of mammalian translation factor eef1a displays increased tyrosine phosphorylation and ability to interact with different signalling molecules. *Int J Biochem Cell Biol*, 40(1):63–71.

Pantopoulos, K. (2004). Iron metabolism and the ire/irp regulatory system: an update. *Ann N Y Acad Sci*, 1012:1–13.

Patel, G. P., Ma, S., and Bag, J. (2005). The autoregulatory translational control element of poly(a)-binding protein mrna forms a heteromeric ribonucleoprotein complex. *Nucleic Acids Res*, 33(22):7074–7089.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, 8(10):785–786.

Petrak, J. and Vyoral, D. (2005). Hephaestin–a ferroxidase of cellular iron export. *Int J Biochem Cell Biol*, 37(6):1173–1178.

Pham, D. Q.-D. and Chavez, C. A. (2005). The ferritin light-chain homologue promoter in aedes aegypti. *Insect Mol Biol*, 14(3):263–270.

Pham, D. Q.-D., Douglass, P. L., Chavez, C. A., and Shaffer, J. J. (2005). Regulation of the ferritin heavy-chain homologue gene in the yellow fever mosquito, aedes aegypti. *Insect Mol Biol*, 14(3):223–236.

Pham, D. Q. D. and Winzerling, J. J. (2010). Insect ferritins: Typical or atypical? *Biochim Biophys Acta*, 1800(8):824–833.

Piccinelli, P. and Samuelsson, T. (2007). Evolution of the iron-responsive element. *RNA*, 13(7):952–966.

Pickering, B. M. and Willis, A. E. (2005). The implications of structured 5' untranslated regions on translation and disease. *Semin Cell Dev Biol*, 16(1):39–47.

Pierleoni, A., Martelli, P. L., Fariselli, P., and Casadio, R. (2006). Bacello: a balanced subcellular localization predictor. *Bioinformatics*, 22(14):e408–e416.

Pintard, L., Willis, J. H., Willems, A., Johnson, J.-L. F., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers, M., Bowerman, B., and Peter, M. (2003). The btb protein mel-26 is a substrate-specific adaptor of the cul-3 ubiquitin-ligase. *Nature*, 425(6955):311–316.

Ponka, P. (1997). Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. *Blood*, 89(1):1–25.

Ponting, C. P. (1995). Sam: a novel motif in yeast sterile and drosophila polyhomeotic proteins. *Protein Sci*, 4(9):1928–1930.

Provance, Jr, D. W., Gourley, C. R., Silan, C. M., Cameron, L. C., Shokat, K. M., Goldenring, J. R., Shah, K., Gillespie, P. G., and Mercer, J. A. (2004). Chemical-genetic inhibition of a sensitized mutant myosin vb demonstrates a role in peripheral-pericentriolar membrane traffic. *Proc Natl Acad Sci U S A*, 101(7):1868–1873.

Quinlan, M. E., Heuser, J. E., Kerkhoff, E., and Mullins, R. D. (2005). Drosophila spire is an actin nucleation factor. *Nature*, 433(7024):382–388.

Quiroz, C., Pearson, V., Gulyani, S., Allen, R., Earley, C., and FerrÕ, S. (2010). Up-regulation of striatal adenosine a(2a) receptors with iron deficiency in rats: effects on locomotion and cortico-striatal neurotransmission. *Exp Neurol*, 224(1):292–298.

Ribeiro, J. M. C. (2003). A catalogue of anopheles gambiae transcripts significantly more or less expressed following a blood meal. *Insect Biochem Mol Biol*, 33(9):865–882.

Rice, P., Longden, I., and Bleasby, A. (2000). Emboss: the european molecular biology open software suite. *Trends Genet*, 16(6):276–277.

Richardson, D. R. and Milnes, K. (1997). The potential of iron chelators of the pyridoxal isonicotinoyl hydrazone class as effective antiproliferative agents ii: the mechanism of action of ligands derived from salicylaldehyde benzoyl hydrazone and 2-hydroxy-1-naphthylaldehyde benzoyl hydrazone. *Blood*, 89(8):3025–3038.

Riedl, A. and Jacobs-Lorena, M. (1996). Determinants of drosophila fushi tarazu mrna instability. *Mol Cell Biol*, 16(6):3047–3053.

Robinson, D. N. and Cooley, L. (1997). Drosophila kelch is an oligomeric ring canal actin organizer. *J Cell Biol*, 138(4):799–810.

Rodrigues, V., Cheah, P. Y., Ray, K., and Chia, W. (1995). malvolio, the drosophila homologue of mouse nramp-1 (bcg), is expressed in macrophages and in the nervous system and is required for normal taste behaviour. *EMBO J*, 14(13):3007–3020.

Roskoski, Jr, R. (2010). Raf protein-serine/threonine kinases: structure and regulation. *Biochem Biophys Res Commun*, 399(3):313–317.

Roth, A. C. J., Gonnet, G. H., and Dessimoz, C. (2008). Algorithm of oma for large-scale orthology inference. *BMC Bioinformatics*, 9:518.

Rouault, T. A. and Klausner, R. D. (1996). Iron-sulfur clusters as biosensors of oxidants and iron. *Trends Biochem Sci*, 21(5):174–177.

Rougon, G. and Hobert, O. (2003). New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. *Annu Rev Neurosci*, 26:207–238.

Runner, V. M., Podolny, V., and Buratowski, S. (2008). The rpb4 subunit of rna polymerase ii contributes to cotranscriptional recruitment of 3' processing factors. *Mol Cell Biol*, 28(6):1883–1891.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406–425.

Sakane, F., Imai, S., Kai, M., Wada, I., and Kanoh, H. (1996). Molecular cloning of a novel diacylglycerol kinase isozyme with a pleckstrin homology domain and a c-terminal tail similar to those of the eph family of protein-tyrosine kinases. *J Biol Chem*, 271(14):8394–8401.

Salemi, M. and Vandamme, A.-M. (2003). *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*. Cambridge University Press.

Sampath, P., Mazumder, B., Seshadri, V., and Fox, P. L. (2003). Transcript-selective translational silencing by gamma interferon is directed by a novel structural element in the ceruloplasmin mrna 3' untranslated region. *Mol Cell Biol*, 23(5):1509–1519.

Satterthwaite, A. B., Li, Z., and Witte, O. N. (1998). Btk function in b cell development and response. *Semin Immunol*, 10(4):309–316.

Schepens, B., Tinton, S. A., Bruynooghe, Y., Parthoens, E., Haegman, M., Beyaert, R., and Cornelis, S. (2007). A role for hnrnp c1/c2 and unr in internal initiation of translation during mitosis. *EMBO J*, 26(1):158–169.

Schweisguth, F. (2004). Regulation of notch signaling activity. *Current Biology*, 14(3):R129–R138.

Segal, E., Friedman, N., Kaminski, N., Regev, A., and Koller, D. (2005). From signatures to models: understanding cancer using microarrays. *Nat Genet*, 37 Suppl:S38–S45.

Sheftel, A. D., Zhang, A.-S., Brown, C., Shirihai, O. S., and Ponka, P. (2007). Direct interorganellar transfer of iron from endosome to mitochondrion. *Blood*, 110(1):125–132.

Siaut, M., Zaros, C., Levivier, E., Ferri, M.-L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. (2003). An rpb4/rpb7-like complex in

yeast rna polymerase iii contains the orthologue of mammalian cgrp-rcp. *Mol Cell Biol*, 23(1):195–205.

Sierzputowska-Gracz, H. and Theil, E. C. (1995). 15n nmr and cd studies of the ire (iron regulatory element in ferritin mrna with single base substitution in the hairpin loop. *Nucleic Acids Symp Ser*, (33):203–206.

Singson, A., Mercer, K. B., and L'Hernault, S. W. (1998). The c. elegans spe-9 gene encodes a sperm transmembrane protein that contains egf-like repeats and is required for fertilization. *Cell*, 93(1):71–79.

Sliskovic, I. and Mutus, B. (2006). Reversible inhibition of caspase-3 activity by iron (iii): potential role in physiological control of apoptosis. *FEBS letters*, 580(9):2233–2237.

Small, J. V., Kaverina, I., Krylyshkina, O., and Rottner, K. (1999). Cytoskeleton cross-talk during cell motility. *FEBS Lett*, 452(1-2):96–99.

Smith, J. M., Haigh, J., *et al.* (1974). The hitch-hiking effect of a favourable gene. *Genet Res*, 23(1):23–35.

Solovyov, A. and Gilbert, H. F. (2004). Zinc-dependent dimerization of the folding catalyst, protein disulfide isomerase. *Protein Sci*, 13(7):1902–1907.

Somma, M. P., Fasulo, B., Siriaco, G., and Cenci, G. (2003). Chromosome condensation defects in barren rna-interfered drosophila cells. *Genetics*, 165(3):1607–1611.

Soroka, V., Kasper, C., and Poulsen, F. M. (2008). Structural biology of ncam. *Neurochem Res*.

Spinazzola, A., Viscomi, C., Fernandez-Vizarra, E., Carrara, F., D'Adamo, P., Calvo, S., Marsano, R. M., Donnini, C., Weiher, H., Strisciuglio, P., Parini, R., Sarzi, E., Chan, A., DiMauro, S., RŽtig, A., Gasparini, P., Ferrero, I., Mootha, V. K., Tiranti, V., and Zeviani, M. (2006). Mpv17 encodes an inner mitochondrial membrane protein and is mutated in infantile hepatic mitochondrial dna depletion. *Nat Genet*, 38(5):570–575.

Stein, E., Huynh-Do, U., Lane, A. A., Cerretti, D. P., and Daniel, T. O. (1998). Nck recruitment to eph receptor, ephb1/elk, couples ligand activation to c-jun kinase. *J Biol Chem*, 273(3):1303–1308.

Stevens, S. G., Gardner, P. P., and Brown, C. (2011). Two covariance models for iron-responsive elements. *RNA Biol*, 8(5):792–801.

Strickler-Dinglasan, P. M., Guz, N., Attardo, G., and Aksoy, S. (2006). Molecular characterization of iron binding proteins from glossina morsitans morsitans (diptera: Glossinidae). *Insect Biochem Mol Biol*, 36(12):921–933.

Swanson, W. J., Nielsen, R., and Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol*, 20(1):18–20.

Swofford, C. D., Kasckow, J. W., Scheller-Gilkey, G., and Inderbitzin, L. B. (1996). Substance use: a powerful predictor of relapse in schizophrenia. *Schizophr Res*, 20(1-2):145–151.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). Mega4: Molecular evolutionary genetics analysis (mega) software version 4.0. *Mol Biol Evol*, 24(8):1596–1599.

Teichmann, S. A. and Chothia, C. (2000). Immunoglobulin superfamily proteins in caenorhabditis elegans. *J Mol Biol*, 296(5):1367–1383.

Tetley, L., Turner, C. M., Barry, J. D., Crowe, J. S., and Vickerman, K. (1987). Onset of expression of the variant surface glycoproteins of trypanosoma brucei in the tsetse fly studied using immunoelectron microscopy. *J Cell Sci*, 87 (Pt 2):363–372.

Theil, E. C. (1990). The ferritin family of iron storage proteins. *Adv Enzymol Relat Areas Mol Biol*, 63:421–449.

Theil, E. C. (1998). The iron responsive element (ire) family of mrna regulators. regulation of iron transport and uptake compared in animals, plants, and microorganisms. *Met Ions Biol Syst*, 35:403–434.

Therrien, M., Wong, A. M., and Rubin, G. M. (1998). Cnk, a raf-binding multidomain protein required for ras signaling. *Cell*, 95(3):343–353.

Theurl, I., Zoller, H., Obrist, P., Datz, C., Bachmann, F., Elliott, R. M., and Weiss, G. (2004). Iron regulates hepatitis c virus translation via stimulation of expression of translation initiation factor 3. *J Infect Dis*, 190(4):819–825.

Thomas, C., Mackey, M., Diaz, A., and Cox, D. (2009). Hydroxyl radical is produced via the fenton reaction in submitochondrial particles under oxidative stress: implications for diseases associated with iron accumulation. *Redox Report*, 14(3):102–108.

Thomas-Chollier, M., Sand, O., Turatsinze, J.-V., Janky, R., Defrance, M., Vervisch, E., Brohée, S., and van Helden, J. (2008). Rsat: regulatory sequence analysis tools. *Nucleic Acids Res*, 36(Web Server issue):W119–W127.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680.

Tinton, S. A., Schepens, B., Bruynooghe, Y., Beyaert, R., and Cornelis, S. (2005). Regulation of the cell-cycle-dependent internal ribosome entry site of the pitslre protein kinase: roles of unr (upstream of n-ras) protein and phosphorylated translation initiation factor eif-2alpha. *Biochem J*, 385(Pt 1):155–163.

Touati, D. (2000). Iron and oxidative stress in bacteria. *Arch Biochem Biophys*, 373(1):1–6.

Turner, A., Beales, J., Faure, S., Dunford, R. P., and Laurie, D. A. (2005). The pseudo-response regulator ppd-h1 provides adaptation to photoperiod in barley. *Science*, 310(5750):1031–1034.

Valenzuela, J. G., Francischetti, I. M. B., Pham, V. M., Garfield, M. K., and Ribeiro, J. M. C. (2003). Exploring the salivary gland transcriptome and proteome of the anopheles stephensi mosquito. *Insect Biochem Mol Biol*, 33(7):717–732.

Valenzuela, J. G., Pham, V. M., Garfield, M. K., Francischetti, I. M. B., and Ribeiro, J. M. C. (2002). Toward a description of the sialome of the adult female mosquito aedes aegypti. *Insect Biochem Mol Biol*, 32(9):1101–1122.

Vashisht, A. A., Zumbrennen, K. B., Huang, X., Powers, D. N., Durazo, A., Sun, D., Bhaskaran, N., Persson, A., Uhlen, M., Sangfelt, O., Spruck, C., Leibold, E. A., and Wohlschlegel, J. A. (2009). Control of iron homeostasis by an iron-regulated ubiquitin ligase. *Science*, 326(5953):718–721.

Vickerman, K., Tetley, L., Hendry, K. A., and Turner, C. M. R. (1988). Biology of african trypanosomes in the tsetse fly. *Biology of the Cell*, 64(2):109–119.

Vlasova-St Louis, I. and Bohjanen, P. R. (2011). Coordinate regulation of mrna decay networks by gu-rich elements and celf1. *Curr Opin Genet Dev*, 21(4):444–451.

Vogel, C., Teichmann, S. A., and Chothia, C. (2003). The immunoglobulin superfamily in drosophila melanogaster and caenorhabditis elegans and the evolution of complexity. *Development*, 130(25):6317–6328.

Volz, K. (2008). The functional duality of iron regulatory protein 1. *Curr Opin Struct Biol*, 18(1):106–111.

Walden, W. E., Selezneva, A. I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J. C., Theil, E. C., and Volz, K. (2006). Structure of dual function iron regulatory protein 1 complexed with ferritin ire-rna. *Science*, 314(5807):1903–1908.

Wang, X., Trotman, L. C., Koppie, T., Alimonti, A., Chen, Z., Gao, Z., Wang, J., Erdjument-Bromage, H., Tempst, P., Cordon-Cardo, C., *et al.* (2007). Nedd4-1 is a proto-oncogenic ubiquitin ligase for pten. *Cell*, 128(1):129–139.

Weinberg, E. D. and Miklossy, J. (2008). Iron withholding: a defense against disease. *Journal of Alzheimer's Disease*, 13(4):451–463.

Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H., and Breaker, R. R. (2010). Comparative genomics reveals 104 candidate structured rnas from bacteria, archaea, and their metagenomes. *Genome Biol*, 11(3):R31.

Wilden, U., Hall, S. W., and K§hn, H. (1986). Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48-kda protein of rod outer segments. *Proc Natl Acad Sci U S A*, 83(5):1174–1178.

Wilkie, G. S., Dickson, K. S., and Gray, N. K. (2003). Regulation of mrna translation by 5'- and 3'-utr-binding factors. *Trends Biochem Sci*, 28(4):182–188.

Wilkinson, B. and Gilbert, H. F. (2004). Protein disulfide isomerase. *Biochim Biophys Acta*, 1699(1-2):35–44.

Williamson, L. M. (2000). Leucocyte depletion of the blood supply - how will patients benefit? *Br J Haematol*, 110(2):256–272.

Winzerling, J. J. and Pham, D. Q.-D. (2006). Iron metabolism in insect disease vectors: mining the anopheles gambiae translated protein database. *Insect Biochem Mol Biol*, 36(4):310–321.

Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B., and Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet*, 32(4):569–577.

Xue, H., Lu, B., and Lai, M. (2008). The cancer secretome: a reservoir of biomarkers. *J Transl Med*, 6:52.

Yamanaka, K., Minato, N., and Iwai, K. (1999). Stabilization of iron regulatory protein 2, irp2, by aluminum. *FEBS Lett*, 462(1-2):216–220.

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8):1586–1591.

Yang, Z. and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15(12):496–503.

Yang, Z. and Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of molecular evolution*, 46(4):409–418.

Yasuhara, M., Mitsui, S., Hirano, H., Takanabe, R., Tokioka, Y., Ihara, N., Komatsu, A., Seki, M., Shinozaki, K., and Kiyosue, T. (2004). Identification of ask and clock-associated proteins as molecular partners of lkp2 (lov kelch protein 2) in arabidopsis. *J Exp Bot*, 55(405):2015–2027.

Yeh, A. P., Chatelet, C., Soltis, S. M., Kuhn, P., Meyer, J., and Rees, D. C. (2000). Structure of a thioredoxin-like [2fe-2s] ferredoxin from aquifex aeolicus. *Journal of Molecular Biology*, 300(3):587–595.

Yoon, K., Ko, D., Doderer, M., Livi, C. B., and Penalva, L. O. F. (2008). Over-represented sequences located on 3' utrs are potentially involved in regulatory functions. *RNA Biol*, 5(4):255–262.

Yoshiga, T., Hernandez, V. P., Fallon, A. M., and Law, J. H. (1997). Mosquito transferrin, an acute-phase protein that is up-regulated upon infection. *Proc Natl Acad Sci U S A*, 94(23):12337–12342.

Youdim, M. and Green, A. (1977). Biogenic monoamine metabolism and functional activity in iron-deficient rats: Behavioural correlates. In *Ciba Foundation Symposium 51-Iron Metabolism*, pages 201–225. Wiley Online Library.

Yu, C., Zavaljevski, N., Desai, V., and Reifman, J. (2011). Quartets: a fast and accurate algorithm for large-scale orthology detection. *Nucleic Acids Res*, 39(13):e88.

Zeng, G., Cai, S., Liu, Y., and Wu, G.-J. (2012). Metcam/muc18 augments migration, invasion, and tumorigenicity of human breast cancer sk-br-3 cells. *Gene*, 492(1):229–238.

Zeng, G.-F., Cai, S.-X., and Wu, G.-J. (2011). Up-regulation of metcam/muc18 promotes motility, invasion, and tumorigenesis of human breast cancer cells. *BMC Cancer*, 11:113.

Zhang, A.-S., Sheftel, A. D., and Ponka, P. (2006). The anemia of ŞhaemoglobinŤ-deficit(hbd/hbd) mice is caused by a defect in transferrin cycling. *Experimental hematology*, 34(5):593–598.

Zhang, D., Dimopoulos, G., Wolf, A., Minana, B., Kafatos, F., and Winzerling, J. (2002). Cloning and molecular characterization of two mosquito iron regulatory proteins. *Insect biochemistry and molecular biology*, 32(5):579–589.

Zhang, M. and Leong, H. W. (2012). Bbh-ls: an algorithm for computing positional homologs using sequence and gene context similarity. *BMC Syst Biol*, 6 Suppl 1:S22.

Zhang, Z., Wu, Y., Wang, Z., Dunning, F. M., Rehfuss, J., Ramanan, D., Chapman, E. R., and Jackson, M. B. (2011). Release mode of large and small dense-core vesicles specified by different synaptotagmin isoforms in pc12 cells. *Mol Biol Cell*, 22(13):2324–2336.

Zhao, W., Liu, M., and Kirkwood, K. L. (2008). p38alpha stabilizes interleukin-6 mrna via multiple au-rich elements. *J Biol Chem*, 283(4):1778–1785.

Zhou, G., Kohlhepp, P., Geiser, D., Frasquillo, M. D. C., Vazquez-Moreno, L., and Winzerling, J. J. (2007). Fate of blood meal iron in mosquitoes. *J Insect Physiol*, 53(11):1169–1178.

Zhou, J., Yang, F., Leu, N. A., and Wang, P. J. (2012). Mns1 is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet*, 8(3):e1002516.

Appendix A

Glossina Genomics Analysis

The annotation of the tsetse genome have led to the identification of 12220 genes, which were further evaluated and improved through the process of manual curation. As such, a database of curated *Glossina* genes have allowed for further downstream analysis. To gain insight into the function of tsetse genes, ProtFun was used. As a result, 67.93% of genes were assigned into GO categories and 32.02% could not be allocated into any categories. Of the 67.93% proteins 1.74% could not be assigned any biological processes and were considered as uncharacterized genes (Figure A.1a). The functional classifications include genes involved in translation, cell envelope, central intermediary metabolism, amino acid biosynthesis, energy metabolism, transport and binding, regulatory functions, purines and pyramidines, replication and transcription, biosynthesis of cofactors and those involved in fatty acid metabolism (Figure A.1b). Among the assigned GO terms translation is the enriched function, which is mainly due to an increased metabolic need to digest the blood meal and direct it to egg development. These results further support the findings of Ribeiro (2003) who suggested an up-regulated level of transcription/translation control-associated proteins following blood meal. Additionally the classification of enzyme categories using ProtFun have led to the identification of hydrolases, isomerases, ligases, lyases, and oxidoreductases. These enzymes constitute 66.33% of all annotated genes in *Glossina morsitans*. Among the identified enzymatic classifications ligase and lyase are overrepresented.

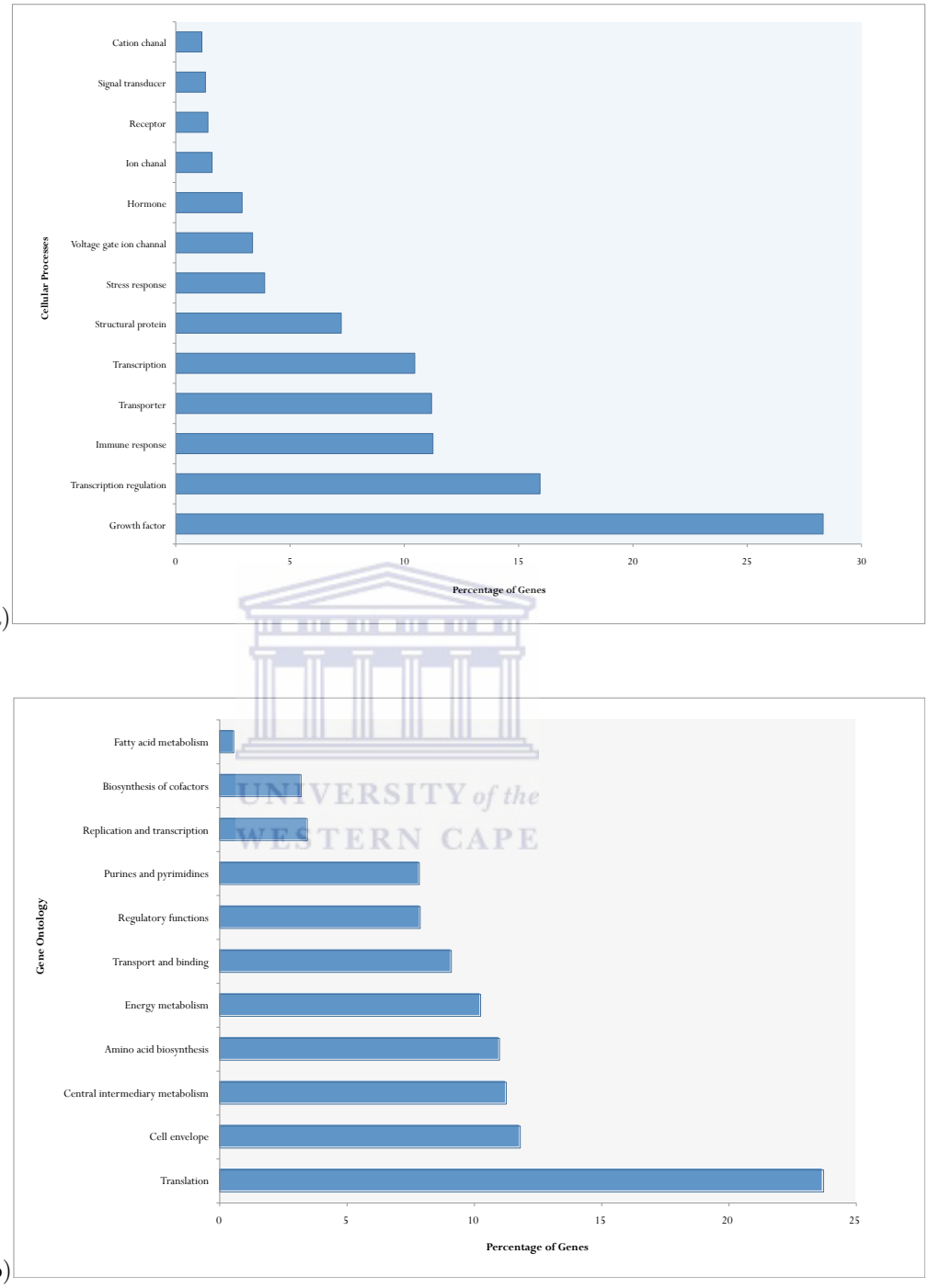


Figure A.1: GO distribution and cellular processes of genes in *Glossina morsitans*. a) shows the assigned cellular processes, and b) represents Go categories. It is evident that translation is the enriched GO and growth factor the enriched cellular process, in which *IRE*-regulated genes are involved.

Despite existing literature on the tissue-specific signal peptides such as the salivary gland

and the fat body (Lehane *et al.*, 2003; Attardo *et al.*, 2010; Alves-Silva *et al.* 2010), no such information at the genome-wide scale exist for *Glossina morsitans*. To this end the results of this study pertaining to the identification of signal peptides, indicative of secretion in Glossina using the genomic information provides novel ground for future studies. As a result the prediction of secreted proteins as part of the ProtFun analysis suggests that 18.02% are secreted proteins among which 12.20% are signal peptides and 5.82% signal anchors. The majority of the identified signal peptides belong to the cell envelope and the transport and binding categories (Figure A.2). Furthermore, a significant fraction of these genes are found to be located in the plasma membrane and the extracellular matrix, which may be considered characteristic features of secreted proteins.

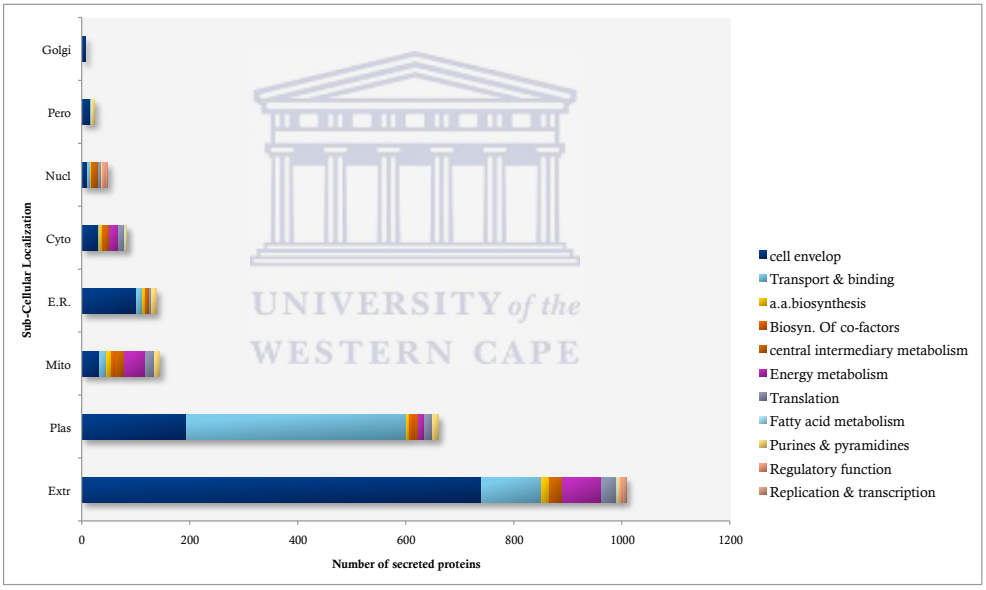


Figure A.2: Putative secreted proteins in the genome of *Glossina morsitans*. As expected the majority of the secreted proteins are in the extracellular matrix and the plasma membrane, while the rest are randomly distributed at other locations. Also most of the putative secreted proteins function as cell envelope and in fatty acid metabolism. This is evident from the dark and light blue bars presented in this figure.

Secreted proteins are of importance due to their accessibility to various drug delivery mechanisms, which is mainly because these proteins are often secreted outside the cell membrane and are therefore presented in the extracellular space. Signal peptides are short peptides present at the N-terminus of the majority of newly synthesized proteins that are involved in the secretary pathway (Blobel and Dobberstein, 1975). Therefore,

further elucidation of genes presented in this work could lead to the development of new interventions for the control of trypanosomiasis.

The localization analysis of Glossina genes resulted in the characterization of sub-cellular localization of 12196 genes, while 24 remained uncharacterized. Most genes are localized within the cytoplasm (31%), nucleus (23%), extracellular (14.16%), and plasma membrane (14.14%).

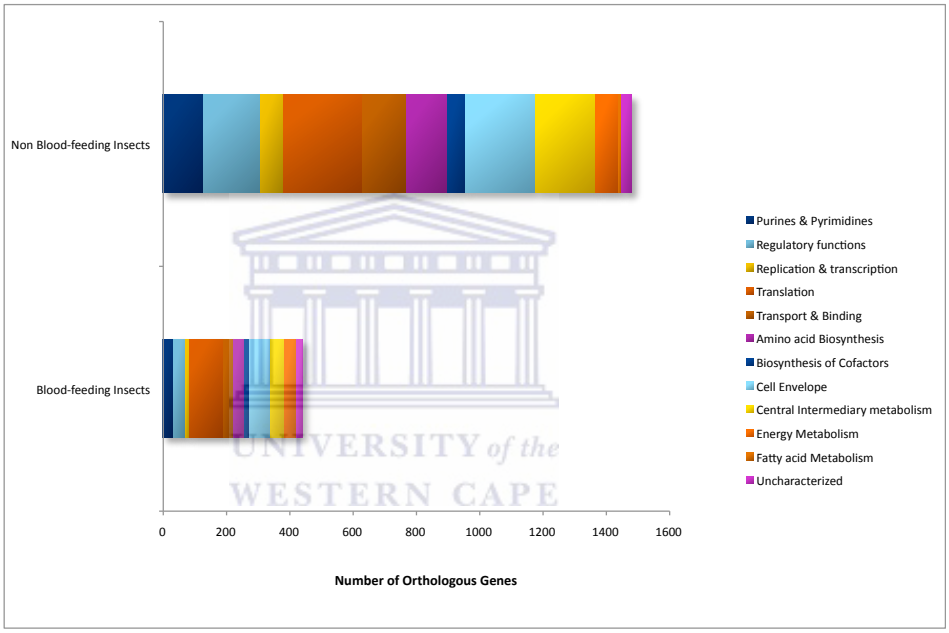


Figure A.3: Glossina genes with blood feeding insect-specific and non blood-feeding insect-specific orthologs. The colours presented in the bar graph point to the cellular processes in which these genes are involved. The majority of blood feeding-specific genes are indicated in translation and cell envelope presented in orange and light blue.

Furthermore, the proteome of *Glossina morsitans* was compared and clustered with several sequenced insect proteome namely *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus*, *Ixodes scapularis*, *Pediculus humanus*, *Drosophila melanogaster*, *Apis mellifera*, *Bombyx mori*, *Danaus Plexippus*, and *Acrythosiphon pisum*, with the objective to define orthology relationships and better understand gene function. Furthermore, such analysis can provide insight into evolutionary processes that are conserved among dipteran species. Orthology and parology are linked to important evolutionary events such as speciation

and duplication respectively. Thus the identification of such relationships are of great significance. Here we identified 7220 gene with orthologs in blood-feeding and non-blood feeding insects, while 5000 are without orthologs and be considered as Glossina-specific. Furthermore, of the 7220 orthologous genes, 412 are only common to blood feeding insects and 1460 corresponding to non-blood feeders (Figure A.3). Furthermore, the conserved genes can be associated with "house-keeping" functions.

Hematophagy has appeared several times during the course of arthropods evolution. It is however notable that, of more than a million insect species only a few thousands (˜14000) have been able to develop blood feeding capabilities (Graca-Souza *et al.*, 2006). This directly points us to the extent of physiological constraints that exist for the evolution of hematophagy. The result of such behavioral changes have made these insects vulnerable to the toxic effects of excessive amounts of iron following digestion. However, considering the success of these insects in nature it is almost certain that a well-coordinated mechanism have allowed them to adapt to high iron-conditions, hence preventing its toxic effects. Furthermore, from the host-pathogen perspective, the parasite spends critical steps of its life cycle within the insect-vector. This implies that hematophagous insect's adaptation to high heme concentrations are consistent with their adaptations to parasitic conditions. To further explore this Kumar *et al.* (2003) showed that a refractory strain of *A. gambiae* to Plasmodium experiences a chronic state of oxidative stress which, was characterized by increased levels of H_2O_2 and superoxide in the hemolymph. In line with this 37 immune response and 16 stress response genes specific to blood-feeding insects with no ortholog in the non-blood feeding counterparts were identified in this study (Figure A.4). In the context of blood feeding these genes have indeed evolved such that to confer refractoriness towards the invading pathogen.

Besides the immunity-related genes other blood feeding-specific genes identified here represent an informative knowledge source to study areas where our insight is sparse and where research is needed to better understand hematophagy and mechanism of its adaptations in insects.

As such the knowledge of species-specific genes can reveal evolutionary processes and

adaptations that have allowed for protection against the deleterious effects of iron, while providing knowledge about key regulators at the vector-pathogen interface.

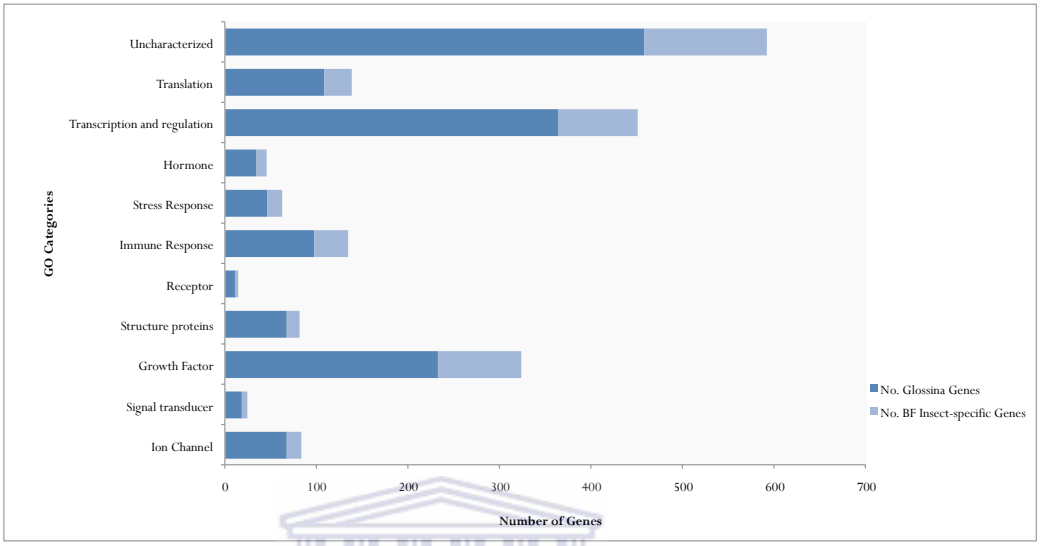


Figure A.4: Blood-feeding-specific genes of *Glossina morsitans* and their GO classifications. As evident a significant proportion of genes have no assigned Go-term. also transcription and regulation, growth factor, immune response and translation have the highest number of BF-specific genes respectively.

Appendix B

IRE-stem loop structures

Table B.1: Details of High-ranked *IRE*-regulated genes in *Glossina morsitans*.

Gene ID	UTR	Gene Name	Start	IRE-pattern
TMP003021	5	kelch-like protein diablo	317	AGAAUUU C AGUUU AAGUUA AGACAAAGUGAA
TMP005554	5	proton-associated sugar transporter a-like	363	AAUGUAA C GAGAA AAGUUC UUCUGCUUAGAUU
TMP005765	5	heterochromatin-associated protein	449	GGCGUUA C UGGAA CCGUGU UAUUCGUUGGCGC
TMP006019	5	insulin-like peptide 1	34	CAAUCUU C AGAGU CAAUGC AUUUUAAGGAUU
TMP006162	5	shaking B	905	UACACAA C GAGUU GGGAGA AAUUGUUUGAAA
TMP006242	5	nmda receptor glutamate-binding chain	234	UAAAGAC C GCAAU GAGUGA GUUGCAAGUCAC
TMP006517	5	GI16026	350	CCCUAUA C GUUCA AAGUUA AGAAUUAUUUUC
TMP006591	5	saccharopine dehydrogenase domain-containing protein	449	GAAACUU C UUGCU GAGUGA UGCAAAAGUUUU
TMP006617	5	20s proteasome regulatory subunit beta type psmb7 psmb10 pup1	893	AAGCUAU C UGUGA AAGUUA UCAUUAAAAAGG
TMP006853	5	visual system homeobox 1	968	GUCCAUA C GUGAU AAGUUA AUUACAAUAAAU
TMP007030	5	ganglioside-induced differentiation-associated protein 1	350	CGAUUCA C UUUUU UAGAAC AAUAAAUAAAAGC
TMP007441	5	mitochondrial processing peptidase beta subunit	213	GCAUGCA C AUAUA CAGGGA CAUAUUGGAUGU
TMP007858	5	Semaphorin	181	UAGAUGA C UUGCU UAGCAG UGCAAUAACUUU
TMP008000	5	nucleosome remodeling factor - isoform a	137	AGAGAAA C UAGCA CAGUGA UGCCAUUUUAAC
TMP008274	5	CG15118	964	AUAUAGA C GAACU CAGAGA AGCUUUACAAUU
TMP008747	5	stoned isoform e	940	AAUUAAA C UGCUU GAGAGU GAGCUUAUAAAU
TMP009134	5	150 Kda dynein-associated polypeptide	629	GCGUAUG C CAGUA AAGUUC GACUGUUUAAAA
TMP009157	5	ubiquitin-conjugating enzyme9	666	AAGGGUA C UCCAU GGGAGG GUGGUUUAUACA
TMP009552	5	synaptotagmin	765	ACUUUAA C AACUG CAGUGU CGAUUUAUAACU
TMP010544	5	GG21264	660	AUAUCUU C CCUUU AAGUUG AUGAGGACGUACU
TMP012387	5	cytochrome P450	69	UUGUCGU C AGGCU UAGGAU AGUUUAUGUUUU
TMP012391	5	cytochrome B561	517	AAAGAAA C GCCGG CAGAGA CUCGGUUCCGCUU
TMP012414	5	Sticks and stones, isoform A	853	ACACUCG C GUUUA AAGUUU CGAACCAAAUGG
TMP013100	5	sox box protein isoform a	788	AUAAUCA C UUUUC AAGUUU GAAAUAUUGAUGA
TMP013254	5	btk family kinase at isoform a	73	CGUUACU C CGUGA AAGUUA UCAUUACAAUCU
TMP013886	5	trna(uridine-2-o-)-methyl transferase trm7	118	CCGAGUG C AGCCG CAGGGC CGGUGUUAGCUU
TMP006651	5	phenylalanyl-tRNA synthase subunit beta	1	AUCCUCU A CAAGA GUGUUU GUAAUUAUUU
TMP008259	5	hypothetical protein	682	AAAUAAA G AUGAU CAGUGA AUCAUUGCUAGC
TMP002718	3	shaggy	855	UUUGAUG C UUCGA UAGAAC CCGAACUUCCAC
TMP002821	3	Glucose-methanol-choline oxidoreductase	393	CGGGGAC C AAUGC CAGAGA GCAUUAGCAACU
TMP003059	3	serine/threonine-protein kinase	816	CGUUUUA C AUCUU GGGAGU GAGCAUUUAAAA
TMP003060	3	GK22170	629	CGUUUUA C AUCUU GGGAGU GAGCAUUUAAAA
TMP004201	3	acetyl cholinesterse	410	CUUUUCA C UAAAU CAGAGC GUUUACAAUUAU
TMP004289	3	f-box protein	323	GGUUACG C AGAUG UAGAAU CAUCGUUUGUAU
TMP004292	3	GH20979	89	UUAAUAG C GUUAA UAGUAU UUAGCAAGUUUU
TMP004552	3	DNA polymerase epsilon	1	CGUUAAA C AUAUU GGGAGA AAUGUAGUUAAU
TMP004553	3	ATP synthase subunit cf6	625	CGUUAAA C AUAUU GGGAGA AAUGUAGUUAAU
TMP004987	3	cuticular protein 67B	683	AUGAAAU C ACGAC CAGUGG GUUGAACAUACU
TMP005243	3	40S ribosomal protein	536	CCUCUUA C UUUUU UAGAAU AGAAAAUGGGAA
TMP005901	3	neuronal pentraxin-1-like	725	ACCAUGA C UAUGG CCGUGA CCAUGACCAUGG
TMP007137	3	hypothetical conserved protein	212	CACCAAU C AGGCG GGGAGA UGCGUACAAUGC
TMP007308	3	folylpolyglutamate synthase, mitochondrial like	546	ACAAAUG C AUGGU CAGGGA GCCAUUUAAUAA
TMP007388	3	hmg coenzyme a isoform a	101	GCAUGCA C AUAUA CAGGGA CAUAUUGGAUGU
TMP007537	3	integrator 4	306	UCCCGAU C CCAAA UAGAAU UUGUGGGAUUUGG
TMP007621	3	26s proteasome regulatory complex subunit psmd5	587	CAGUGCU C GAUGU AAGUUU AUAUCAAGAAUC
TMP008943	3	Glucosidase	341	GAACAGU C GAUUG CAGAGC CAAUUAUGAUUC
TMP009934	3	rhomboid	750	AGUUGUG C CGUCA CCGUGU UGAUGAGAUUGA
TMP010016	3	tpr repeat-containing protein	489	GAAAGUA C AGAAA AAGUUG UUUCUUCGUGCC
TMP010549	3	aldehyde dehydrogenase	159	CAGUGCC C UCAAC AAGUUC GUUGACAGAAUC
TMP010707	3	GH18161	155	UUGAUAA C AUGCA CAGAGU UGUAUUUGUUUG
TMP010821	3	Abdomnal A	302	UUGAAAU C GGUGC CUGUGC GCACAUACAAAC
TMP010846	3	n-methyl-d-aspartate receptor glutamate-binding subunit	615	AAGCGUU C AAGGC CCGUGC GUCUUCCUCAGA
TMP011617	3	E3 ubiquitin-protein ligase	401	UAAUGAG C UGGUG AAGUUG CAUCAACGAUCA
TMP011786	3	disulfide isomerase	882	UUGAAAU C AGUGG CAGAGA UACAUUAAAAUUU
TMP012225	3	phd finger protein	406	GCUCCCA C UGUAA CAGUGC CUACAUAUGAAU
TMP012412	3	sticks and isoform a	425	UGACCAU C CGCGU CUGUGU ACGCCACUGCUG
TMP013137	3	pentatricopeptide	158	CAGACAU C CAUUG CAGUGG CGAAGAAAUAAG
TMP013269	3	UBX domain-containing protein 7	899	CACAUUU C UUCGC CCGUGC GCGGAUCAAUGA
TMP013384	3	GF18386	304	CAUCCAU C GUCCU GAGAGU GGGAAACAGUGA
TMP013948	3	coiled-coil-helix-coiled-coil-helix domain-containing protein2	708	GCUGCCA C CAGCA UAGUAU UGCUGAUUAGCU
TMP014030	3	GJ12585	266	AUUUAAA C UUACA GAGUCA UGUGCAUUUACAA
TMP014204	3	60s ribosomal protein l23	818	CGUCGGG C CUGAU UAGAAU AUUAACCUUGGC
TMP008942	3	eukaryotic translation initiation factor	1	ACAGUCG A UUGCA GAGCCA AUUAUGAUUC

Table B.2: Details of Medium-ranked *IRE*-regulated genes in *Glossina morsitans*.

Gene ID	UTR	Gene Name	Start	IRE-pattern
TMP002406	5	Cuticle protein	83	GUGUUGG C UAGCC GGGAGU GGCUAUGUUGCG
TMP002546	5	hypothetical protein	803	AUGAUGU C UUUGA AAGUUU UCGAAAAGUAUG
TMP002587	5	alternative splicing factor SRP20/9G8	661	GUAGAAU C UGUAA AAGUUA UUGUAAAUAUAA
TMP002921	5	hypothetical protein	915	CCAUCAU C UAGUG CAGAGU UUACUGAUGAUG
TMP003297	5	TRP channel protein	765	AUCAAAA C GAAAG UAGAAC CAUUUCUCUAUA
TMP003577	5	st7-like protein	315	GCCAAUU C CCAUA AAGUUU UGUGGAAAGUUG
TMP003628	5	hypothetical protein	512	GUGAUAA C AGUAC CAGUGU GUGUAUUAAGUA
TMP004029	5	reverse transcriptase	925	GGGACUG C GGACG CAGAGC CGUUCUAACAAA
TMP004047	5	GJ14155	789	GCAAAAA C GAAAU GAGUGA AUUUUUCAAAUC
TMP004240	5	elongation factor1 alpha	252	UGGUGUG C AAUAA AAGUUC UUAUAUUUCAC
TMP004517	5	importin9	373	AAGUCGA C AAAGA UAGGAU UCUUUUGGUAGA
TMP004746	5	myosin heavy chain type II	731	UAAUGAU C GGAAG CAGAGC UUAUUCAUUGGCG
TMP004763	5	spire	165	AAGUUUU C UAAGG AAGUUG CUUUAAACAUGAU
TMP004803	5	GJ11047	598	ACGAUGG C GGUAA UAGAAU UAUACCCAAAAAG
TMP005026	5	beta-carotene dioxygenase	667	GGUUUUG C GGCAU AAGUUC AUGGUUCUAUUC
TMP005160	5	homeobox protein	575	CAAGGCA C CCGUA CAAUGC UAUGUGUAUAGCA
TMP005167	5	arp 23 complex subunit arpc5	693	AGAUUUU C UAUUU AAGUUU GACGUAAAACCCA
TMP005219	5	Unknown	119	CAGAUUG C UUUCA CAAUGC UUGAAAUCAUUAA
TMP005259	5	gly-rich protein	695	UUCAGAA C AAUGA AAGUUU UCAUUUUGAUGU
TMP005357	5	LIX1-like protein	110	GUGGCAA C AGCAG CAGUGU UUGUUUAUUUCG
TMP006315	5	acylphosphatase	79	AGGAAGU C UCUUU UAGUAC AAGGAACAAAUA
TMP006377	5	coenzyme Q-binding protein	729	CGAUCUU C UACUU AAGUUU GACGUAAAACUCA
TMP007122	5	pdgf-and vegf-related factor isoform e	555	CUUACAU C UAAGA AAGUUC UUUUUAAUCUUAU
TMP007350	5	defective proboscisextension response	83	GCAGUAG C AUGUU AAGUUC AAACAUUAUAAAU
TMP007537	5	integrator 4	498	AUAAUCA C CUCCU UAGAAU AGGCGUAACUUU
TMP007548	5	retinoid-and fatty acid-binding glycoprotein	372	UGCAGUU C UUCGU CAGUGG GUGUAAACGCAA
TMP007550	5	GH23986	197	UUUUGUU C AAGGA AAGUUU UCUUUAUAAGUC
TMP007653	5	Abdominal-B protein	716	AUCGCUA C GACGC CAGUGA GUGGUUUAAGUC
TMP007908	5	Unknown	609	CUGAUUA C CUUCU CAGUGG GGAACGUCGAUUA
TMP008329	5	gustatory receptor isoformc	929	AACUUUU C GAAUA AAGUUC UGUUCGUUUAUU
TMP008384	5	LIM domain protein	116	UUUUUUA C UGAGA AAGUUU UUUUAUUAAAAU
TMP008448	5	autophagy-related protein	403	ACACGGU C AAGUU CUGUGU GACUUGCUGUAC
TMP008854	5	receptor tyrosine kinase	120	UCGGAGA C UUUAU AAGUUA AUUAAAUCAAGCC
TMP009102	5	GG15559	552	AUUCACA C UCUUU CAGGGA AAGGGUCGGAAA
TMP009128	5	hypothetical conserved protein	324	CUAUAUU C GCUGC CAGUGU GCGGAUAUGUUAA
TMP009332	5	GF15465	955	UAGCGUU C ACAAU CAGAGA AUUGUGUUUCAA
TMP009821	5	GJ22290	626	UGGAUAU C AAACA GAGCCG UGUUAUAGGUUUA
TMP010651	5	trna delta-isopentenyl pyrophosphate transferase	48	CGGAUUU C AAAUG GAGCCA CCAUUUACUAACA
TMP010673	5	ferritin heavy chain-like protein	642	UCGCCUU C UAUGC CAGUGC GUAUAAAGGUUU
TMP011448	5	brain chitinase and chia	795	CUAAGCU C CUCGA AAGUUC UGCGAGGGAUUUU
TMP012368	5	GJ11024	872	UUACAAA C CGUUA CAGGGU UAUACGUGGGAUG
TMP013701	5	zinc finger protein 106	626	GCCUCGA C UUGCU CAGUGG AGACGAUGCAAAA
TMP014187	5	mediator 100kd	210	UACAUAU C UUUAC UAGUAU GUAUAAAAUAUA
TMP007593	5	barren	757	AGUGCUC G AUGUA AGUUUA UAUCAAGAAUC
TMP007741	5	meiosis-specific nuclear structura l	485	AAUUUUU G GCUUA CAGUGA UGAGCGUGAAGC
TMP002103	3	cg11159-likeprotein	294	UAAUUUU C AAUUU AAGUUA AAAUUAUUUUUA
TMP002357	3	dynein lightchain	384	AUCCUGU C GAACG CAGUGU UGUUUAUUUCGG
TMP002520	3	nucleoporin	562	ACUUGGU C UUUAA AAGUUC UGUAGAAAUUUA
TMP002863	3	angiotensin-converting enzyme	376	UAAAACU C UUUGA AAGUUG UUCAGAAAAAUA
TMP003203	3	DNA replication licensing factor mcm7	492	AAUUCCG C CUACG AAGUUA CAGUGGUUCAAA
TMP003336	3	Ig-like domains containing protein	380	UUCCGGU C AGGGA CAGGGC UCUACUACGACU
TMP003369	3	Serine/threonine-protein kinase	52	CAUAAUG C UAUGU GAGUGA ACAUAGUUCUGU
TMP003500	3	beta-hexosaminidase fdl	477	UAAUUAG C GAUUU GGGAGC AGAUCCGGAAAA
TMP003602	3	phosphoinositide3-kinase	581	ACGAUUG C GUAUC GGGAGU GCAUACUUCAAU
TMP003617	3	host cell factor	318	AAUACCU C AUGAU CAGUGA AUAUAUAUCGCA
TMP004121	3	zinc finger protein2	463	UUAGAAA C AUUUU UAGAAU AGGAUUUCAAAG
TMP004413	3	Wnk,isoformA	108	UACGUAA C CGGAA CAGAGA UUAUUGUUCUGCG
TMP004581	3	hypothetical protein	147	CCAUUAU C AUGUA UAGAAU UAUAUGUAAGCA
TMP005335	3	zinc carboxypeptidase a1	132	CGUGUCU C CACUU CAGAGA AAGCUGAUGGACC
TMP005511	3	atg5	971	GCCACUA C AUAUU UAGUAU AUAUAUUUCA
TMP005595	3	ATP-dependent RNA helicase dhx 36-like	354	AUAAUUU C ACCUG UAGAAU UAAGUACAAAUA
TMP005713	3	tetraspanin42	804	UGGCAAG C GGAUG GAGAGG CAUCGUUUAACA
TMP005827	3	unknown	916	AAAUAUG C CUGCG CAGUGG UGGCAGUGUUUUA
TMP006363	3	fork head box transcription factor	272	ACAACAU C UGUAU CUGUGU AUCAUAAUAAUAU
TMP006826	3	ring-box protein2	222	UUACCAA C AUUUA AAGUUA AGAAUUUAAUAA
TMP006868	3	5-hydroxy tryptamine receptor1	675	AUAAGAC C GGCUA AAGUUC UAGCUGCUGUGC
TMP006998	3	saccharopine dehydrogenase domain containing protein	663	GUAAUCA C UGAAA CUUAGC UUUUAUUUUCGC
TMP007216	3	kinesin heavy chain	609	UAAAAAA C AUUUC CAGAGA GAGAUUUUUAAC
TMP007247	3	beaten path IIa	878	AUGGUAC C GAUGG CAGUGG CUGUCGAGAAUA
TMP007308	3	folylpolyglutamate synthase, mitochondrial like	546	ACAAAUG C AUGGU CAGGGA GCCAUUUAAUAAG
TMP007507	3	topoisomerase 3beta	536	ACAGUCA C GAUAA AAGUUU UAUGUCUACUUUU
TMP008441	3	gamma-tubulin ring complex	944	CCCUCUG C UCAGU CAGAGA AUUAGAUAGAGCA
TMP009871	3	bruno-3 transcript	476	AAUGGUG C CGAUG GAGCCG CAAUUGUUCUAAU
TMP010134	3	clock work isoformb	550	CCUACUA C UUUUG CUUAGC UGAUAAUUUAGGU
TMP010317	3	general transcription factor IIH subunit1	840	AUGCAGA C AUUUU CAGGGA AAAACUUUGCAG
TMP010987	3	sam-motif-containing protein	807	UACUACU C GUUUG CAGUGU CAGCGCAUGCUGU
TMP011104	3	serine-pyruvate mitochondrial	711	UUUCUCU C UUUUU CAGUGU AACGAAAAUUUU
TMP011371	3	laccase-2	263	AGGAACU C GAUAU UAGUAC AUAUUAAAGGAA
TMP011421	3	tm2d1_drome	880	AGGGCUU C UCUGU UAGAAU ACUAGAGACUAGC
TMP011993	3	transposase	739	UAAGGAG C UCGGC GAGUCG GCUAGGCAUUGAU
TMP011994	3	transposase	37	CAUUGUG C AAUUU UAGUAU AAUUUCUCUCUC
TMP012389	3	hypothetical protein	216	GGGUUGC C UUGUU CAGUGA AGCAGGGUCCUG
TMP012491	3	amino-acylase-1	699	AGUAAUU C AGAAA AAGUUU UUUUUAUUUAUU
TMP012534	3	ataxin-2 binding protein	24	CAAGCGG C CGCUG UAGCAG UAGCGGCCGCAG
TMP012770	3	vacuolar H	109	CUCCCUG C UGUUC GAGUGU GAAUACUAGUCA
TMP012772	3	reverse transcriptase	743	AAAUCCU C UGUUU UAGUAU AAAUUGAUUAUA
TMP013603	3	glucose-fructose oxidoreductase domain-containing protein1	474	AGAUUUU C UAUUU AAGUUU GACGUAAACCCAA
TMP013958	3	PERQ amino acid-rich with GYF domain-containing protein	692	AAAUAGG C AAAUU UAGAAC AAUUCUUUUUUC
TMP014133	3	unc-50	95	GUCGGCU C UGUAA UAGCAG UUACAAGUUCCG

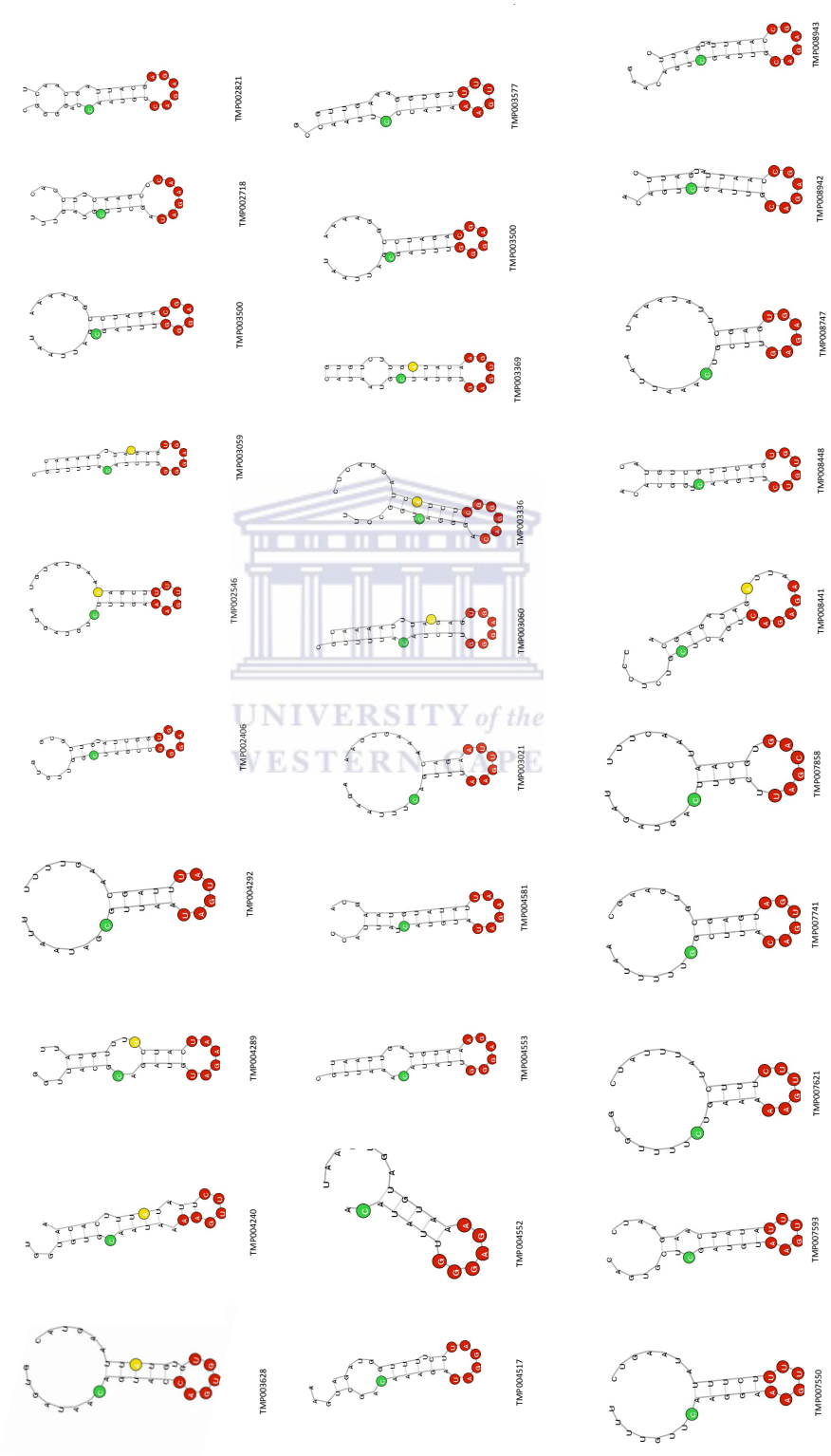
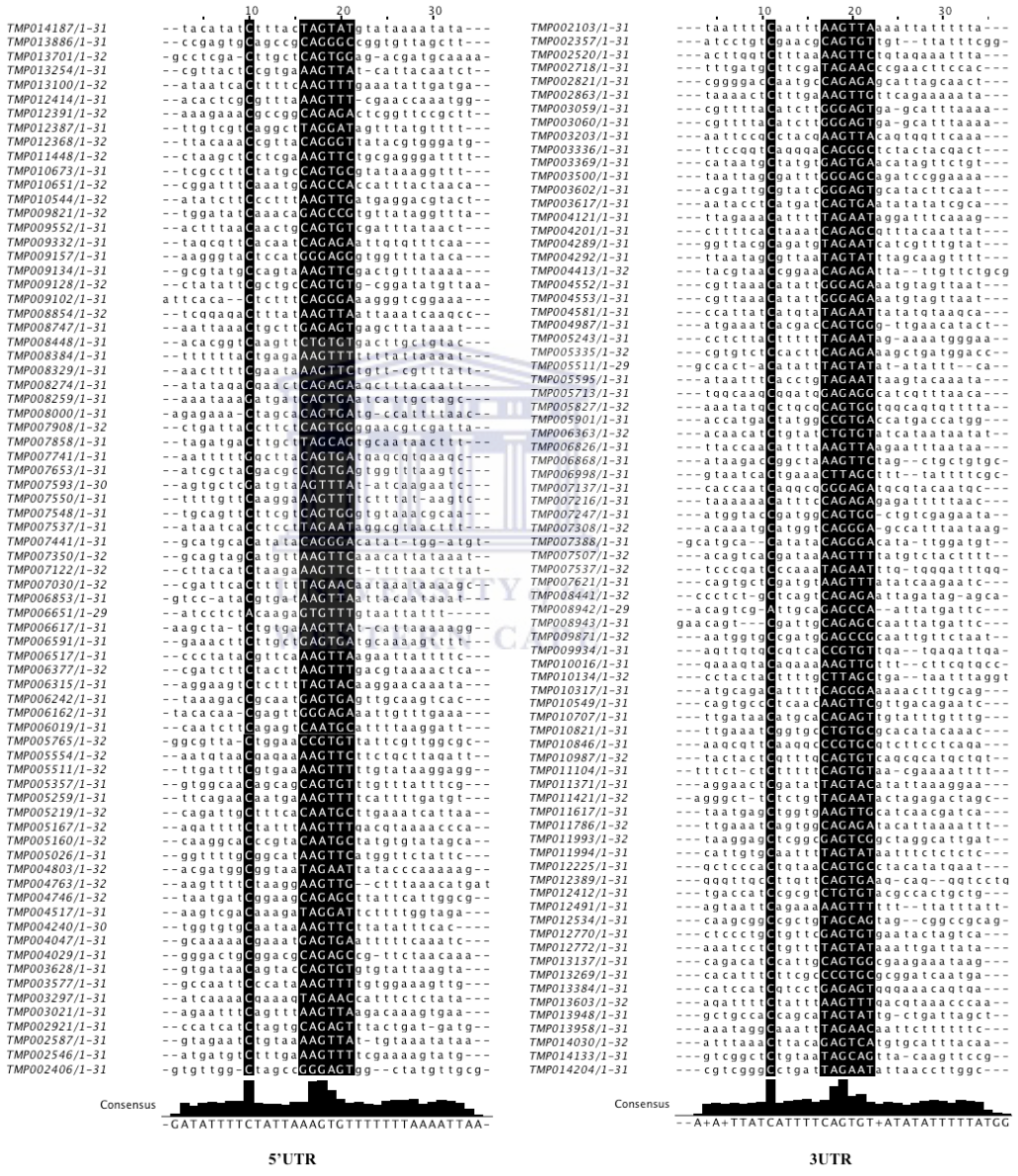


Figure B.1: Secondary stem-loop structures of putative *IRE* structures in *Glossina morsitans*.

Table B.3: Alignment of putative *IRE* sequences highlighting the c-bulge and the loop sequences.



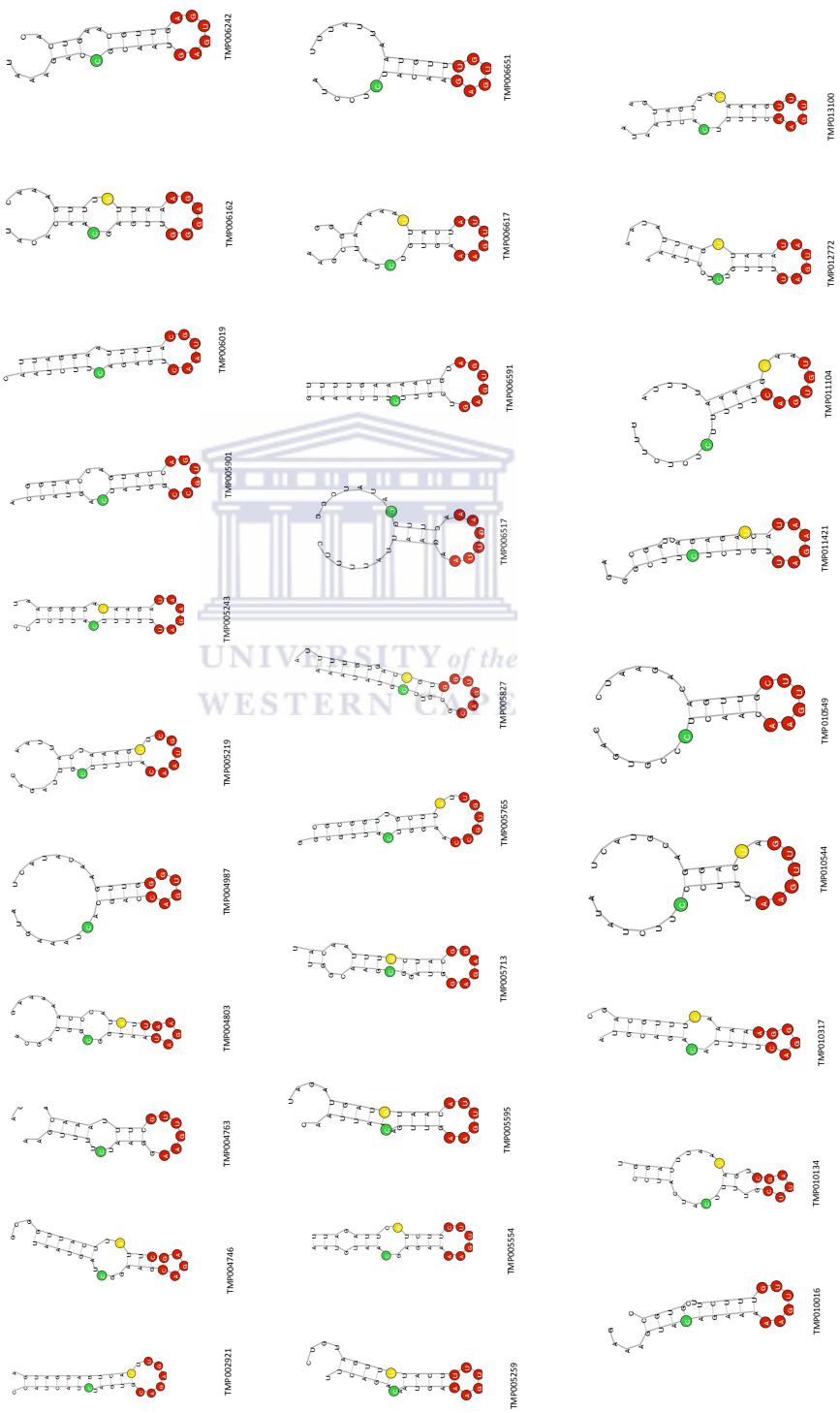


Figure 6.1.1: Continued...

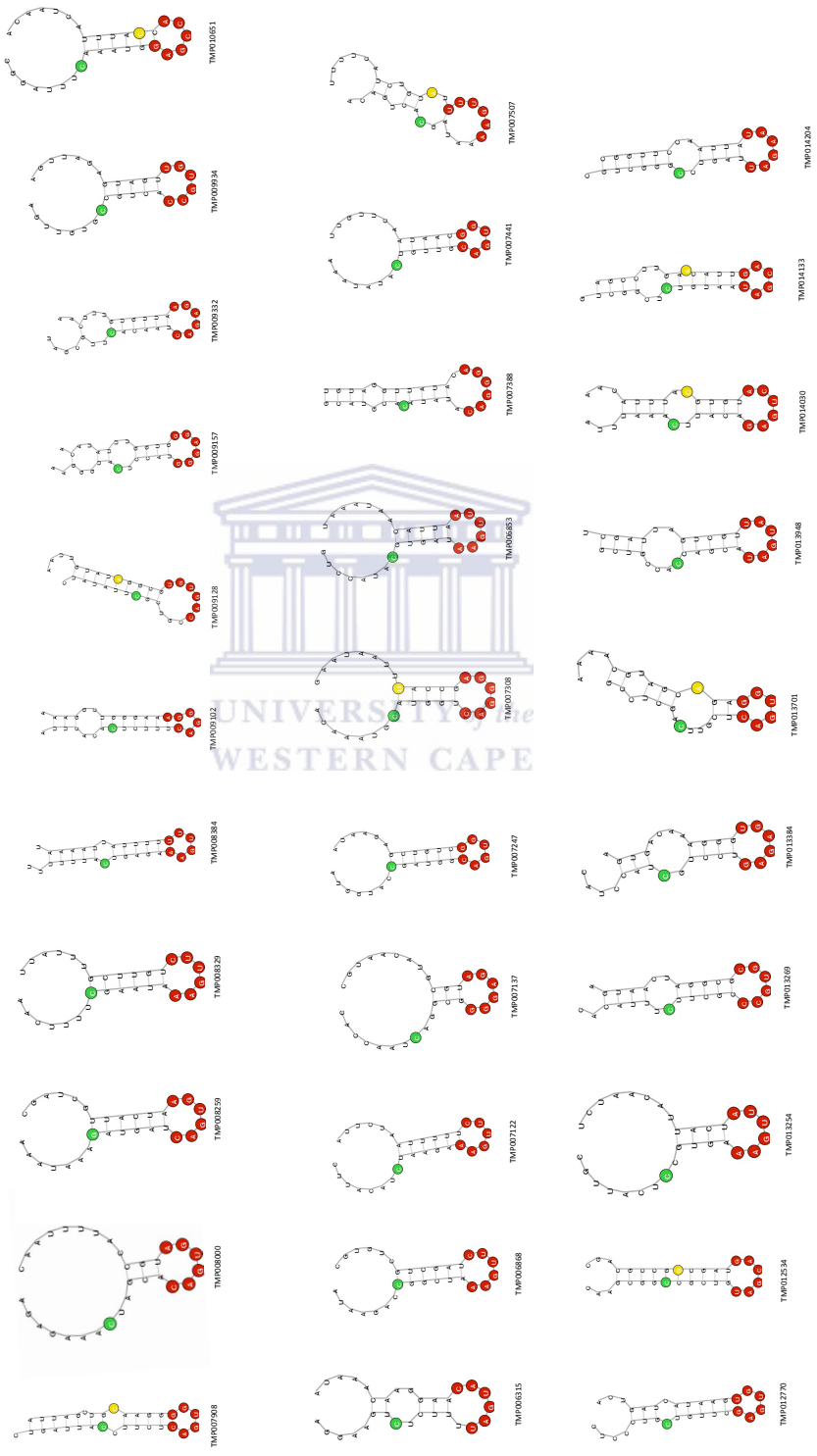


Figure 6.1.1: continued...

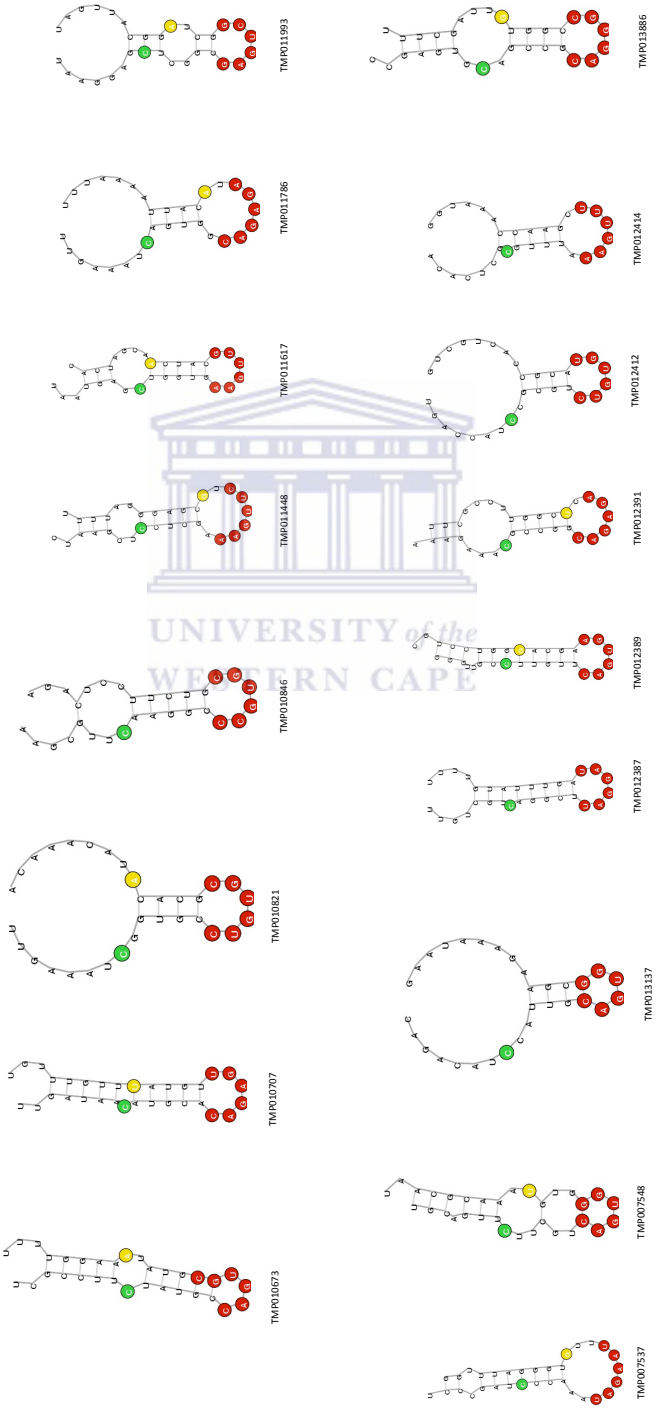
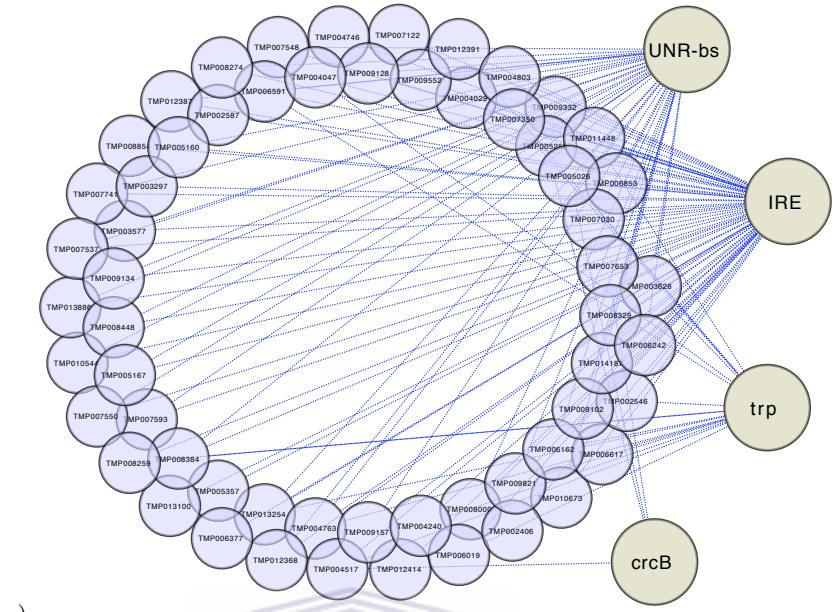
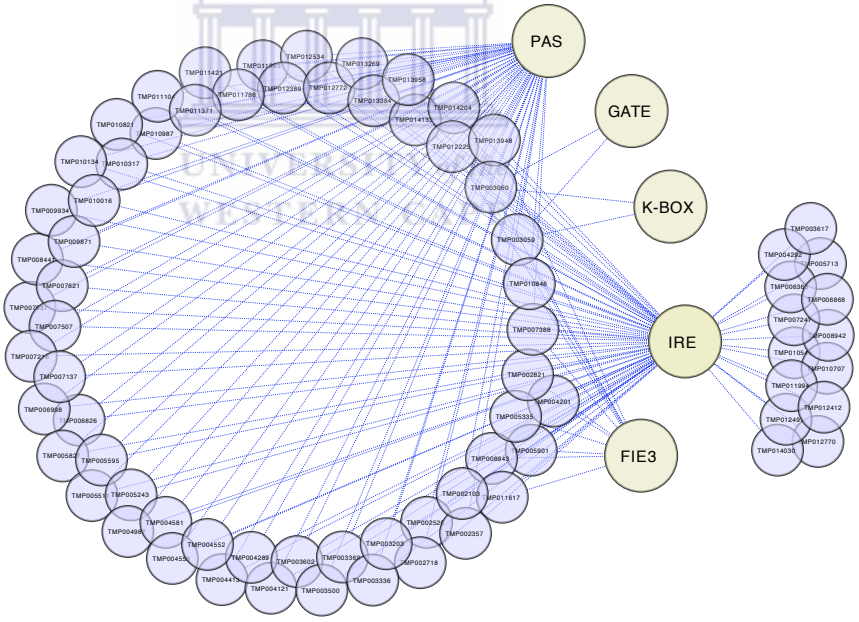


Figure 6.1.1: continued...



a)



b)

Figure B.2: A circular map of the putative *IRE*-regulated genes, sharing one or more motifs. a) represent the identified 5'-UTR elements and b) shows the 3'-UTR elements shared. The majority of the *IRE*-regulated genes share *PAS* and *UNR*-bs motifs as well. Other motifs are presented in a fewer number of *IRE*-regulated genes. Furthermore, no other motifs besides *IRE* was found in the UTRs of 16 genes.

Table B.4: Gene ID conversions for IRE-regulated genes in Glossina.

TMP002103	GMOY000103	TMP007507	GMOY005401
TMP002357	GMOY000357	TMP007537	GMOY005431
TMP002406	GMOY000406	TMP007548	GMOY005442
TMP002520	GMOY000519	TMP007550	GMOY005444
TMP002546	GMOY000545	TMP007593	GMOY005485
TMP002587	GMOY000584	TMP007621	GMOY005513
TMP002821	GMOY000812	TMP007653	GMOY005545
TMP002863	GMOY000853	TMP007741	GMOY005633
TMP002921	GMOY000910	TMP007858	GMOY005749
TMP003021	GMOY001007	TMP007908	GMOY005797
TMP003059	GMOY001045	TMP008000	GMOY005887
TMP003060	GMOY001046	TMP008259	GMOY006141
TMP003203	GMOY001186	TMP008274	GMOY006156
TMP003297	GMOY001279	TMP008329	GMOY006209
TMP003336	GMOY001317	TMP008384	GMOY006264
TMP003369	GMOY001347	TMP008441	GMOY006321
TMP003500	GMOY001475	TMP008448	GMOY006327
TMP003577	GMOY001551	TMP008747	GMOY006619
TMP003602	GMOY001576	TMP008854	GMOY006724
TMP003617	GMOY001591	TMP008942	GMOY006808
TMP003628	GMOY001601	TMP008943	GMOY006809
TMP004029	GMOY001995	TMP009102	GMOY006965
TMP004047	GMOY002012	TMP009128	GMOY006990
TMP004121	GMOY002082	TMP009134	GMOY006996
TMP004201	GMOY002161	TMP009157	GMOY007017
TMP004240	GMOY002199	TMP009332	GMOY007187
TMP004289	GMOY002244	TMP009552	GMOY007406
TMP004292	GMOY002247	TMP009821	GMOY007668
TMP004413	GMOY002368	TMP009871	GMOY007718
TMP004517	GMOY002470	TMP009934	GMOY007780
TMP004552	GMOY002504	TMP010016	GMOY007858
TMP004553	GMOY002505	TMP010134	GMOY007975
TMP004581	GMOY002533	TMP010317	GMOY008151
TMP004746	GMOY002696	TMP010544	GMOY008376
TMP004763	GMOY002713	TMP010549	GMOY008381
TMP004803	GMOY002753	TMP010651	GMOY008480
TMP004987	GMOY002936	TMP010673	GMOY008502
TMP005026	GMOY002975	TMP010707	GMOY008535
TMP005160	GMOY003108	TMP010821	GMOY008646
TMP005167	GMOY003115	TMP010846	GMOY008670
TMP005219	GMOY003166	TMP010987	GMOY008806
TMP005243	GMOY003190	TMP011104	GMOY008920
TMP005259	GMOY003206	TMP011371	GMOY009183
TMP005335	GMOY003280	TMP011421	GMOY009232
TMP005357	GMOY003300	TMP011448	GMOY009258
TMP005511	GMOY003449	TMP011617	GMOY009423
TMP005554	GMOY003491	TMP011786	GMOY009591
TMP005595	GMOY003531	TMP011993	GMOY009796
TMP005713	GMOY003645	TMP011994	GMOY009797
TMP005765	GMOY003696	TMP012225	GMOY010018
TMP005827	GMOY003757	TMP012368	GMOY010160
TMP005901	GMOY003829	TMP012387	GMOY010179
TMP006019	GMOY003945	TMP012389	GMOY010181
TMP006162	GMOY004085	TMP012391	GMOY010183
TMP006242	GMOY004163	TMP012412	GMOY010203
TMP006315	GMOY004236	TMP012414	GMOY010205
TMP006363	GMOY004282	TMP012491	GMOY010282
TMP006377	GMOY004296	TMP012534	GMOY010325
TMP006517	GMOY004434	TMP012770	GMOY010559
TMP006591	GMOY004507	TMP012772	GMOY010561
TMP006617	GMOY004532	TMP013100	GMOY010876
TMP006651	GMOY004566	TMP013137	GMOY010911
TMP006826	GMOY004737	TMP013254	GMOY011025
TMP006853	GMOY004763	TMP013269	GMOY011040
TMP006868	GMOY004778	TMP013384	GMOY011151
TMP006998	GMOY004905	TMP013603	GMOY011367
TMP007030	GMOY004935	TMP013701	GMOY011465
TMP007122	GMOY005027	TMP013886	GMOY011648
TMP007137	GMOY005042	TMP013948	GMOY011710
TMP007216	GMOY005119	TMP013958	GMOY011720
TMP007247	GMOY005150	TMP014030	GMOY011791
TMP007308	GMOY005208	TMP014133	GMOY011894
TMP007350	GMOY005248	TMP014187	GMOY011948
TMP007388	GMOY005285	TMP014204	GMOY011965
TMP007441	GMOY005336	TMP002718	GMOY012041

Appendix C

Expression profiling

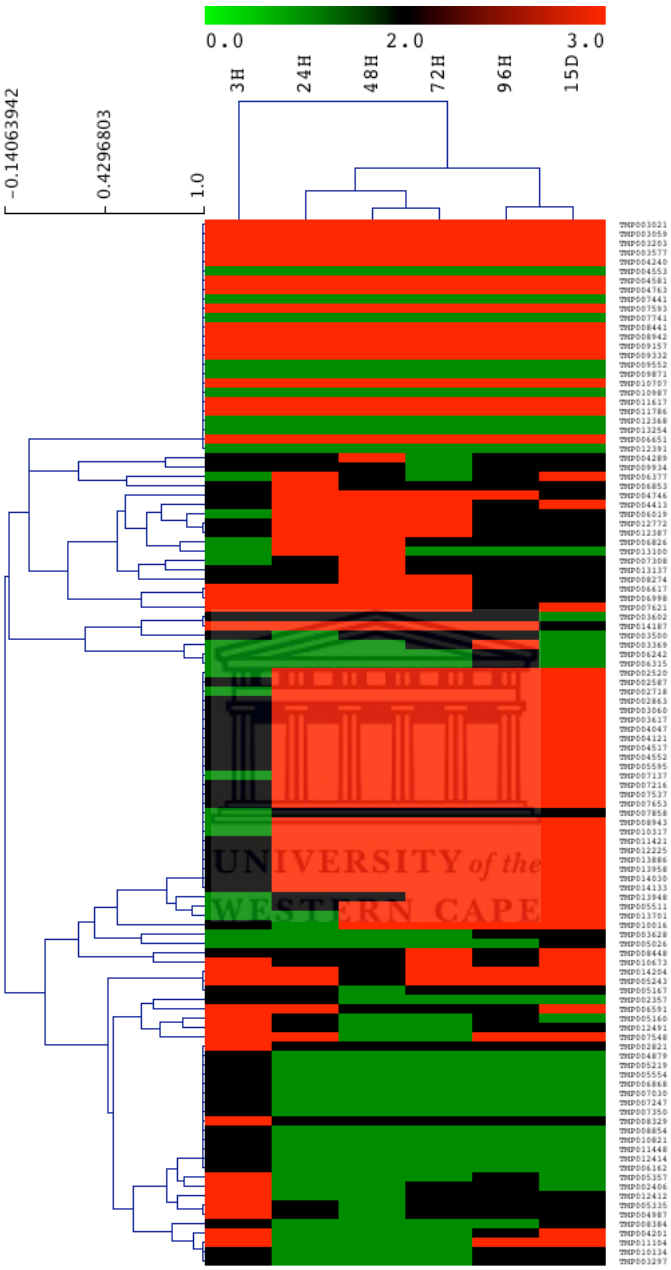


Figure C.1: The expression-based clustering analysis of *IRE*-regulated genes following blood-meal.

Table C.1: Messages differentially expressed at 3Hs post-blood meal.

Significantly Expressed Post-Blood Meal:

Gene IDs	Functional Annotation	GO Category
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis
TMP003203	DNA replication licensing factor mcm7	Amino_acid_biosynthesis
TMP004746	myosin heavy chain type II	Amino_acid_biosynthesis
TMP007621	26s proteasome regulatory complex subunit psmd5	Amino_acid_biosynthesis
TMP014187	mediator 100kd	Amino_acid_biosynthesis
TMP006591	saccharopine dehydrogenase domain-containing protein	Amino_acid_biosynthesis
TMP009332	GF15465	Biosynthesis_of_cofactors
TMP003577	st7-like protein	Biosynthesis_of_cofactors
TMP010016	tpr repeat-containing protein	nucleotide biosynthesis
TMP006998	saccharopine dehydrogenase domain containing protein	nucleotide biosynthesis
TMP002520	nucleoporin	nucleotide biosynthesis
TMP004047	GJ14155	nucleotide biosynthesis
TMP003021	kelch-like protein diablo	nucleotide biosynthesis
TMP004517	importin9	nucleotide biosynthesis
TMP004763	spire	nucleotide biosynthesis
TMP008441	gamma-tubulin ring complex	nucleotide biosynthesis
TMP008943	Glucosidase	Cell_envelope
TMP006019	insulin-like peptide 1	Cell_envelope
TMP002863	angiotensin-converting enzyme	Cell_envelope
TMP007548	retinoid-and fatty acid-binding glycoprotein	Cell_envelope
TMP012387	cytochrome P450	Cell_envelope
TMP007537	integrator 4	Cell_envelope
TMP011786	disulfide isomerase	Cell_envelope
TMP004413	Wnk,isoformA	Central_intermediary_metabolism
TMP011617	E3 ubiquitin-protein ligase	Central_intermediary_metabolism
TMP007653	Abdominal-B protein	Central_intermediary_metabolism
TMP004240	elongation factor1 alpha	Central_intermediary_metabolism
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism
TMP013886	trna(uridine-2-o-)-methyltransferasetrm7	Central_intermediary_metabolism
TMP003059	serine/threonine-protein kinase	Energy_metabolism
TMP011421	tm2d1_drome	Energy_metabolism
TMP006617	20s proteasome regulatory subunit beta	Energy_metabolism
TMP012772	reverse transcriptase	Energy_metabolism
TMP010707	GH18161	Transcription and translation
TMP007137	hypothetical conserved protein	Transcription and translation
TMP007216	kinesin heavy chain	Transcription and translation
TMP002718	shaggy	Transcription and translation
TMP004121	zinc finger protein2	Transcription and translation
TMP004581	hypothetical protein	Transcription and translation
TMP012225	phd finger protein	Transcription and translation
TMP006853	visual system homeobox 1	Transcription and translation
TMP014030	GJ12585	Transcription and translation
TMP002587	alternative splicing factor SRP20/9G8	Transcription and translation
TMP005243	40S ribosomal protein	Transcription and translation
TMP007593	barren	Transcription and translation
TMP008942	eukaryotic Transcription and translation initiation factor 3	Transcription and translation
TMP009157	ubiquitin-conjugating enzyme9	Transcription and translation
TMP006826	ring-box protein2	Transcription and translation
TMP003617	host cell factor	Transcription and translation
TMP013958	PERQ amino acid-rich with GYF domain-containing protein	Transcription and translation
TMP003060	GK22170	Transcription and translation
TMP013100	sox box protein isoform a	Transcription and translation
TMP014204	60s ribosomal protein l23	Transcription and translation
TMP014133	unc-50	Transport

Significantly Repressed Post-Blood Meal:

Gene IDs	Functional Annotation	GO Category
TMP006377	coenzyme Q-binding protein	Amino_acid_biosynthesis
TMP007247	beaten path IIa	Amino_acid_biosynthesis
TMP007441	mitochondrial processing peptidase beta subunit	Amino_acid_biosynthesis
TMP012412	sticks and isoform a	Amino_acid_biosynthesis
TMP007030	ganglioside-induced differentiation-associated protein 1	nucleotide biosynthesis
TMP012414	Sticks and stones, isoform A	nucleotide biosynthesis
TMP008854	receptor tyrosine kinase	nucleotide biosynthesis
TMP007350	defective proboscisextension response	Biosynthesis_of_cofactors
TMP003297	TRP channel protein	Biosynthesis_of_cofactors
TMP005554	proton-associated sugar transporter a-like	Biosynthesis_of_cofactors
TMP002406	Cuticle protein	Cell_envelope
TMP005026	beta-carotene dioxygenase	Cell_envelope
TMP003500	beta-hexosaminidase fdl	Cell_envelope
TMP010987	sam-motif ubiquitously expressed punctatedly localized protein	Central_intermediary_metabolism
TMP003628	hypothetical protein	Central_intermediary_metabolism
TMP011448	brain chitinase and chia	Central_intermediary_metabolism
TMP004201	acetyl cholinesterse	Energy_metabolism
TMP009552	synaptotagmin	Energy_metabolism
TMP012368	GJ11024	Energy_metabolism
TMP005219	Unknown	Energy_metabolism
TMP005357	LIX1-like protein	Energy_metabolism
TMP009871	bruno-3 transcript	Energy_metabolism
TMP011104	serine-pyruvate mitochondrial	Energy_metabolism
TMP010821	Abdomnal A	Transcription and translation
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	Transcription and translation
TMP007741	meiosis-specific nuclear structura l	Transcription and translation
TMP008384	LIM domain protein	Transcription and translation
TMP006315	acylphosphatase	Transcription and translation
TMP013254	btk family kinase at isoform a	Transcription and translation
TMP010134	clock work isoformb	Transcription and translation
TMP003369	Serine/threonine-protein kinase	Transport
TMP006868	5-hydroxy tryptamine receptor1	Transport

Table C.2: Messages differentially expressed at 24H post-blood meal.

Significantly Expressed Post-Blood Meal:

Gene IDs	Functional Annotation	GO Category
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis
TMP003203	DNA replication licensing factor mcm7	Amino_acid_biosynthesis
TMP004746	myosin heavy chain type II	Amino_acid_biosynthesis
TMP007621	26s proteasome regulatory complex subunit psmd5	Amino_acid_biosynthesis
TMP014187	mediator 100kd	Amino_acid_biosynthesis
TMP006591	saccharopine dehydrogenase domain-containing protein	Amino_acid_biosynthesis
TMP009332	GF15465	Biosynthesis_of_cofactors
TMP003577	st7-like protein	Biosynthesis_of_cofactors
TMP010016	tpr repeat-containing protein	nucleotide biosynthesis
TMP006998	saccharopine dehydrogenase domain containing protein	nucleotide biosynthesis
TMP002520	nucleoporin	nucleotide biosynthesis
TMP004047	GJ14155	nucleotide biosynthesis
TMP003021	kelch-like protein diablo	nucleotide biosynthesis
TMP004517	importin9	nucleotide biosynthesis
TMP004763	spire	nucleotide biosynthesis
TMP008441	gamma-tubulin ring complex	nucleotide biosynthesis
TMP008943	Glucosidase	Cell_envelope
TMP006019	insulin-like peptide 1	Cell_envelope
TMP002863	angiotensin-converting enzyme	Cell_envelope
TMP007548	retinoid-and fatty acid-binding glycoprotein	Cell_envelope
TMP012387	cytochrome P450	Cell_envelope
TMP007537	integrator 4	Cell_envelope
TMP011786	disulfide isomerase	Cell_envelope
TMP004413	Wnk,isoformA	Central_intermediary_metabolism
TMP011617	E3 ubiquitin-protein ligase	Central_intermediary_metabolism
TMP007653	Abdominal-B protein	Central_intermediary_metabolism
TMP004240	elongation factor1 alpha	Central_intermediary_metabolism
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism
TMP013886	trna(uridine-2-o-)-methyltransferasetrm7	Central_intermediary_metabolism
TMP003059	serine/threonine-protein kinase	Energy_metabolism
TMP011421	tm2d1_drome	Energy_metabolism
TMP006617	20s proteasome regulatory subunit beta	Energy_metabolism
TMP012772	reverse transcriptase	Energy_metabolism
TMP010707	GH18161	Transcription and translation
TMP007137	hypothetical conserved protein	Transcription and translation
TMP007216	kinesin heavy chain	Transcription and translation
TMP002718	shaggy	Transcription and translation
TMP004121	zinc finger protein2	Transcription and translation
TMP004581	hypothetical protein	Transcription and translation
TMP012225	phd finger protein	Transcription and translation
TMP006853	visual system homeobox 1	Transcription and translation
TMP014030	GJ12585	Transcription and translation
TMP002587	alternative splicing factor SRP20/9G8	Transcription and translation
TMP005243	40S ribosomal protein	Transcription and translation
TMP007593	barren	Transcription and translation
TMP008942	eukaryotic Transcription and translation initiation factor 3	Transcription and translation
TMP009157	ubiquitin-conjugating enzyme9	Transcription and translation
TMP006826	ring-box protein2	Transcription and translation
TMP003617	host cell factor	Transcription and translation
TMP013958	PERQ amino acid-rich with GYF domain-containing protein	Transcription and translation
TMP003060	GK22170	Transcription and translation
TMP013100	sox box protein isoform a	Transcription and translation
TMP014204	60s ribosomal protein l23	Transcription and translation
TMP014133	unc-50	Transport

Significantly Repressed Post-Blood Meal:

Gene IDs	Functional Annotation	GO Category
TMP006377	coenzyme Q-binding protein	Amino_acid_biosynthesis
TMP007247	beaten path IIa	Amino_acid_biosynthesis
TMP007441	mitochondrial processing peptidase beta subunit	Amino_acid_biosynthesis
TMP012412	sticks and isoform a	Amino_acid_biosynthesis
TMP007030	ganglioside-induced differentiation-associated protein 1	nucleotide biosynthesis
TMP012414	Sticks and stones, isoform A	nucleotide biosynthesis
TMP008854	receptor tyrosine kinase	nucleotide biosynthesis
TMP007350	defective proboscisextension response	Biosynthesis_of_cofactors
TMP003297	TRP channel protein	Biosynthesis_of_cofactors
TMP005554	proton-associated sugar transporter a-like	Biosynthesis_of_cofactors
TMP002406	Cuticle protein	Cell_envelope
TMP005026	beta-carotene dioxygenase	Cell_envelope
TMP003500	beta-hexosaminidase fdl	Cell_envelope
TMP010987	sam-motif ubiquitously expressed punctatedly localized protein	Central_intermediary_metabolism
TMP003628	hypothetical protein	Central_intermediary_metabolism
TMP011448	brain chitinase and chia	Central_intermediary_metabolism
TMP004201	acetyl cholinesterse	Energy_metabolism
TMP009552	synaptotagmin	Energy_metabolism
TMP012368	GJ11024	Energy_metabolism
TMP005219	Unknown	Energy_metabolism
TMP005357	LIX1-like protein	Energy_metabolism
TMP009871	bruno-3 transcript	Energy_metabolism
TMP011104	serine-pyruvate mitochondrial	Energy_metabolism
TMP010821	Abdomnal A	Transcription and translation
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	Transcription and translation
TMP007741	meiosis-specific nuclear structura l	Transcription and translation
TMP008384	LIM domain protein	Transcription and translation
TMP006315	acylphosphatase	Transcription and translation
TMP013254	btk family kinase at isoform a	Transcription and translation
TMP010134	clock work isoformb	Transcription and translation
TMP003369	Serine/threonine-protein kinase	Transport
TMP006868	5-hydroxy tryptamine receptor1	Transport

Table C.3: Messages differentially expressed at 48H post-blood meal.

Significantly Expressed post-blood meal:

Gene IDs	Functional Annotation	GO Category
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis
TMP003203	DNA replication licensing factor mcm7	Amino_acid_biosynthesis
TMP004746	myosin heavy chain type II	Amino_acid_biosynthesis
TMP005511	atg5	Amino_acid_biosynthesis
TMP007621	26s proteasome regulatory complex subunit psmd5	Amino_acid_biosynthesis
TMP014187	mediator 100kd	Amino_acid_biosynthesis
TMP010016	tpr repeat-containing protein	nucleotide biosynthesis
TMP006998	saccharopine dehydrogenase domain containing protein	nucleotide biosynthesis
TMP002520	nucleoporin	nucleotide biosynthesis
TMP004047	GJ14155	nucleotide biosynthesis
TMP003021	kelch-like protein diablo	nucleotide biosynthesis
TMP004517	importin9	nucleotide biosynthesis
TMP004763	spire	nucleotide biosynthesis
TMP008441	gamma-tubulin ring complex	nucleotide biosynthesis
TMP009332	GF15465	Biosynthesis_of_cofactors
TMP003577	st7-like protein	Biosynthesis_of_cofactors
TMP008943	Glucosidase	Cell_envelope
TMP006019	insulin-like peptide 1	Cell_envelope
TMP002863	angiotensin-converting enzyme	Cell_envelope
TMP012387	cytochrome P450	Cell_envelope
TMP007537	integrator 4	Cell_envelope
TMP011786	disulfide isomerase	Cell_envelope
TMP004413	Wnk,isoformA	Central_intermediary_metabolism
TMP011617	E3 ubiquitin-protein ligase	Central_intermediary_metabolism
TMP007653	Abdominal-B protein	Central_intermediary_metabolism
TMP004240	elongation factor1 alpha	Central_intermediary_metabolism
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism
TMP013886	trna(uridine-2-o-)-methyltransferasetrm7	Central_intermediary_metabolism
TMP003059	serine/threonine-protein kinase	Energy_metabolism
TMP011421	tm2d1_drome	Energy_metabolism
TMP006617	20s proteasome regulatory subunit beta type psmb7 psmb10 pup1	Energy_metabolism
TMP007308	folylpolyglutamate synthase, mitochondrial like	Energy_metabolism
TMP012772	reverse transcriptase	Energy_metabolism
TMP010707	GH18161	Transcription and translation
TMP007137	hypothetical conserved protein	Transcription and translation
TMP007216	kinesin heavy chain	Transcription and translation
TMP002718	shaggy	Transcription and translation
TMP004121	zinc finger protein2	Transcription and translation
TMP004581	hypothetical protein	Transcription and translation
TMP012225	phd finger protein	Transcription and translation
TMP014030	GJ12585	Transcription and translation
TMP002587	alternative splicing factor SRP20/9G8	Transcription and translation
TMP007593	barren	Transcription and translation
TMP008274	CG15118	Transcription and translation
TMP008942	eukaryotic Transcription and translation initiation factor 3	Transcription and translation
TMP009157	ubiquitin-conjugating enzyme9	Transcription and translation
TMP006826	ring-box protein2	Transcription and translation
TMP003617	host cell factor	Transcription and translation
TMP013958	PERQ amino acid-rich with GYF domain-containing protein	Transcription and translation
TMP003060	GK22170	Transcription and translation
TMP004289	f-box protein	Transcription and translation
TMP013100	sox box protein isoform a	Transcription and translation
TMP013137	pentatricopeptide	Transcription and translation
TMP014133	unc-50	Transport

Significantly repressed post-blood meal:

Gene IDs	Functional Annotation	GO Category
TMP007247	beaten path IIa	Amino_acid_biosynthesis
TMP005335	zinc carboxypeptidase a1	Amino_acid_biosynthesis
TMP007441	mitochondrial processing peptidase beta subunit	Amino_acid_biosynthesis
TMP012412	sticks and isoform a	Amino_acid_biosynthesis
TMP007030	ganglioside-induced differentiation-associated protein 1	nucleotide biosynthesis
TMP012414	Sticks and stones, isoform A	nucleotide biosynthesis
TMP008854	receptor tyrosine kinase	nucleotide biosynthesis
TMP007350	defective proboscisextension response	Biosynthesis_of_cofactors
TMP003297	TRP channel protein	Biosynthesis_of_cofactors
TMP005554	proton-associated sugar transporter a-like	Biosynthesis_of_cofactors
TMP004987	cuticular protein 67B	Cell_envelope
TMP002406	Cuticle protein	Cell_envelope
TMP005026	beta-carotene dioxygenase	Cell_envelope
TMP007548	retinoid-and fatty acid-binding glycoprotein	Cell_envelope
TMP010987	sam-motif ubiquitously expressed punctatedly localized protein	Central_intermediary_metabolism
TMP003628	hypothetical protein	Central_intermediary_metabolism
TMP005160	homeobox protein	Central_intermediary_metabolism
TMP011448	brain chitinase and chia	Central_intermediary_metabolism
TMP004201	acetyl cholinesterse	Energy_metabolism
TMP009552	synaptotagmin	Energy_metabolism
TMP012368	GJ11024	Energy_metabolism
TMP005219	Unknown	Energy_metabolism
TMP005357	LIX1-like protein	Energy_metabolism
TMP009871	bruno-3 transcript	Energy_metabolism
TMP011104	serine-pyruvate mitochondrial	Energy_metabolism
TMP010134	clock work isoformb	Transcription and translation
TMP010821	Abdomnal A	Transcription and translation
TMP002357	dynein lightchain	Transcription and translation
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	Transcription and translation
TMP005167	arp 23 complex subunit arpc5	Transcription and translation
TMP007741	meiosis-specific nuclear structura l	Transcription and translation
TMP008384	LIM domain protein	Transcription and translation
TMP006315	acylphosphatase	Transcription and translation
TMP012491	amino-acylase-1	Transcription and translation
TMP013254	btk family kinase at isoform a	Transcription and translation
TMP003369	Serine/threonine-protein kinase	Transport
TMP006868	5-hydroxy tryptamine receptor1	Transport

Table C.4: Messages differentially expressed at 72Hs post-blood meal.

Significantly Expressed post-blood meal:

Gene IDs	Functional Annotation	GO Category
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis
TMP003203	DNA replication licensing factor mcm7	Amino_acid_biosynthesis
TMP004746	myosin heavy chain type II	Amino_acid_biosynthesis
TMP005511	atg5	Amino_acid_biosynthesis
TMP007621	26s proteasome regulatory complex subunit psmd5	Amino_acid_biosynthesis
TMP014187	mediator 100kd	Amino_acid_biosynthesis
TMP010016	tpr repeat-containing protein	nucleotide biosynthesis
TMP006998	saccharopine dehydrogenase domain containing protein	nucleotide biosynthesis
TMP002520	nucleoporin	nucleotide biosynthesis
TMP004047	GJ14155	nucleotide biosynthesis
TMP003021	kelch-like protein diablo	nucleotide biosynthesis
TMP004517	importin9	nucleotide biosynthesis
TMP004763	spire	nucleotide biosynthesis
TMP008441	gamma-tubulin ring complex	nucleotide biosynthesis
TMP009332	GF15465	Biosynthesis_of_cofactors
TMP003577	st7-like protein	Biosynthesis_of_cofactors
TMP008943	Glucosidase	Cell_envelope
TMP006019	insulin-like peptide 1	Cell_envelope
TMP002863	angiotensin-converting enzyme	Cell_envelope
TMP012387	cytochrome P450	Cell_envelope
TMP007537	integrator 4	Cell_envelope
TMP011786	disulfide isomerase	Cell_envelope
TMP004413	Wnk,isoformA	Central_intermediary_metabolism
TMP011617	E3 ubiquitin-protein ligase	Central_intermediary_metabolism
TMP007653	Abdominal-B protein	Central_intermediary_metabolism
TMP004240	elongation factor1 alpha	Central_intermediary_metabolism
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism
TMP013886	trna(uridine-2-o-)-methyltransferasetrm7	Central_intermediary_metabolism
TMP013948	coiled-coil-helix-coiled-coil-helix domain-containing protein2	Energy_metabolism
TMP003059	serine/threonine-protein kinase	Energy_metabolism
TMP011421	tm2d1_drome	Energy_metabolism
TMP006617	20s proteasome regulatory subunit beta type psmb7 psmb10 pup1	Energy_metabolism
TMP012772	reverse transcriptase	Energy_metabolism
TMP010707	GH18161	Transcription and translation
TMP007137	hypothetical conserved protein	Transcription and translation
TMP007216	kinesin heavy chain	Transcription and translation
TMP002718	shaggy	Transcription and translation
TMP004121	zinc finger protein2	Transcription and translation
TMP004581	hypothetical protein	Transcription and translation
TMP012225	phd finger protein	Transcription and translation
TMP014030	GJ12585	Transcription and translation
TMP002587	alternative splicing factor SRP20/9G8	Transcription and translation
TMP005243	40S ribosomal protein	Transcription and translation
TMP007593	barren	Transcription and translation
TMP008274	CG15118	Transcription and translation
TMP008942	eukaryotic Transcription and translation initiation factor 3	Transcription and translation
TMP009157	ubiquitin-conjugating enzyme9	Transcription and translation
TMP003617	host cell factor	Transcription and translation
TMP013958	PERQ amino acid-rich with GYF domain-containing protein	Transcription and translation
TMP008448	autophagy-related protein	Transcription and translation
TMP003060	GK22170	Transcription and translation
TMP014204	60s ribosomal protein l23	Transcription and translation
TMP014133	unc-50	Transport
TMP010673	Ferritin heavy chain	Transport

Significantly repressed post-blood meal:

Gene IDs	Functional Annotation	GO Category
TMP006377	coenzyme Q-binding protein	Amino_acid_biosynthesis
TMP007247	beaten path IIa	Amino_acid_biosynthesis
TMP007441	mitochondrial processing peptidase beta subunit	Amino_acid_biosynthesis
TMP007030	ganglioside-induced differentiation-associated protein 1	nucleotide biosynthesis
TMP012414	Sticks and stones, isoform A	nucleotide biosynthesis
TMP008854	receptor tyrosine kinase	nucleotide biosynthesis
TMP007350	defective proboscisextension response	Biosynthesis_of_cofactors
TMP003297	TRP channel protein	Biosynthesis_of_cofactors
TMP005554	proton-associated sugar transporter a-like	Biosynthesis_of_cofactors
TMP005026	beta-carotene dioxygenase	Cell_envelope
TMP007548	retinoid-and fatty acid-binding glycoprotein	Cell_envelope
TMP010987	sam-motif ubiquitously expressed punctatedly localized protein	Central_intermediary_metabolism
TMP003628	hypothetical protein	Central_intermediary_metabolism
TMP005160	homeobox protein	Central_intermediary_metabolism
TMP011448	brain chitinase and chia	Central_intermediary_metabolism
TMP004201	acetyl cholinesterse	Energy_metabolism
TMP009552	synaptotagmin	Energy_metabolism
TMP012368	GJ11024	Energy_metabolism
TMP005219	Unknown	Energy_metabolism
TMP005357	LIX1-like protein	Energy_metabolism
TMP009871	bruno-3 transcript	Energy_metabolism
TMP011104	serine-pyruvate mitochondrial	Energy_metabolism
TMP010134	clock work isoformb	Transcription and translation
TMP010821	Abdomnal A	Transcription and translation
TMP002357	dynein lightchain	Transcription and translation
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	Transcription and translation
TMP007741	meiosis-specific nuclear structura l	Transcription and translation
TMP008384	LIM domain protein	Transcription and translation
TMP006315	acylphosphatase	Transcription and translation
TMP009934	rhomboid	Transcription and translation
TMP004289	f-box protein	Transcription and translation
TMP012491	amino-acylase-1	Transcription and translation
TMP013100	sox box protein isoform a	Transcription and translation
TMP013254	btk family kinase at isoform a	Transcription and translation
TMP006868	5-hydroxy tryptamine receptor1	Transport

Table C.5: Messages differentially expressed at 96Hs post-blood meal.

Significantly Expressed post-blood meal:

Gene IDs	Functional Annotation	GO Category
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis
TMP003203	DNA replication licensing factor mcm7	Amino_acid_biosynthesis
TMP004746	myosin heavy chain type II	Amino_acid_biosynthesis
TMP005511	atg5	Amino_acid_biosynthesis
TMP014187	mediator 100kd	Amino_acid_biosynthesis
TMP010016	tpr repeat-containing protein	nucleotide biosynthesis
TMP002520	nucleoporin	nucleotide biosynthesis
TMP004047	GJ14155	nucleotide biosynthesis
TMP003021	kelch-like protein diablo	nucleotide biosynthesis
TMP004517	importin9	nucleotide biosynthesis
TMP004763	spire	nucleotide biosynthesis
TMP008441	gamma-tubulin ring complex	nucleotide biosynthesis
TMP009332	GF15465	Biosynthesis_of_cofactors
TMP003577	st7-like protein	Biosynthesis_of_cofactors
TMP008943	Glucosidase	Cell_envelope
TMP002863	angiotensin-converting enzyme	Cell_envelope
TMP007548	retinoid-and fatty acid-binding glycoprotein	Cell_envelope
TMP007537	integrator 4	Cell_envelope
TMP011786	disulfide isomerase	Cell_envelope
TMP011617	E3 ubiquitin-protein ligase	Central_intermediary_metabolism
TMP007653	Abdominal-B protein	Central_intermediary_metabolism
TMP004240	elongation factor1 alpha	Central_intermediary_metabolism
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism
TMP013886	trna(uridine-2-o-)-methyltransferasetrm7	Central_intermediary_metabolism
TMP013948	coiled-coil-helix-coiled-coil-helix domain-containing protein2	Energy_metabolism
TMP003059	serine/threonine-protein kinase	Energy_metabolism
TMP011421	tm2d1_drome	Energy_metabolism
TMP011104	serine-pyruvate mitochondrial	Energy_metabolism
TMP010707	GH18161	Transcription and translation
TMP007137	hypothetical conserved protein	Transcription and translation
TMP007216	kinesin heavy chain	Transcription and translation
TMP002718	shaggy	Transcription and translation
TMP004121	zinc finger protein2	Transcription and translation
TMP004581	hypothetical protein	Transcription and translation
TMP012225	phd finger protein	Transcription and translation
TMP014030	GJ12585	Transcription and translation
TMP002587	alternative splicing factor SRP20/9G8	Transcription and translation
TMP005243	40S ribosomal protein	Transcription and translation
TMP007593	barren	Transcription and translation
TMP008942	eukaryotic Transcription and translation initiation factor 3	Transcription and translation
TMP009157	ubiquitin-conjugating enzyme9	Transcription and translation
TMP003617	host cell factor	Transcription and translation
TMP013958	PERQ amino acid-rich with GYF domain-containing protein	Transcription and translation
TMP003060	GK22170	Transcription and translation
TMP014204	60s ribosomal protein l23	Transcription and translation
TMP003369	Serine/threonine-protein kinase	Transport
TMP014133	unc-50	Transport

Significantly repressed post-blood meal:

Gene IDs	Functional Annotation	GO Category
TMP007247	beaten path IIa	Amino_acid_biosynthesis
TMP007441	mitochondrial processing peptidase beta subunit	Amino_acid_biosynthesis
TMP007030	ganglioside-induced differentiation-associated protein 1	nucleotide biosynthesis
TMP012414	Sticks and stones, isoform A	nucleotide biosynthesis
TMP008854	receptor tyrosine kinase	nucleotide biosynthesis
TMP007350	defective proboscisextension response	Biosynthesis_of_cofactors
TMP005554	proton-associated sugar transporter a-like	Biosynthesis_of_cofactors
TMP005026	beta-carotene dioxygenase	Cell_envelope
TMP010987	sam-motif ubiquitously expressed punctatedly localized protein	Central_intermediary_metabolism
TMP011448	brain chitinase and chia	Central_intermediary_metabolism
TMP009552	synaptotagmin	Energy_metabolism
TMP012368	GJ11024	Energy_metabolism
TMP005219	Unknown	Energy_metabolism
TMP009871	bruno-3 transcript	Energy_metabolism
TMP010821	Abdomnal A	Transcription and translation
TMP002357	dynein lightchain	Transcription and translation
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	Transcription and translation
TMP007741	meiosis-specific nuclear structura l	Transcription and translation
TMP008384	LIM domain protein	Transcription and translation
TMP013100	sox box protein isoform a	Transcription and translation
TMP013254	btk family kinase at isoform a	Transcription and translation
TMP006868	5-hydroxy tryptamine receptor1	Transport

Table C.6: Messages differentially expressed at 15Days post-blood meal.

Significantly Expressed post-blood meal:

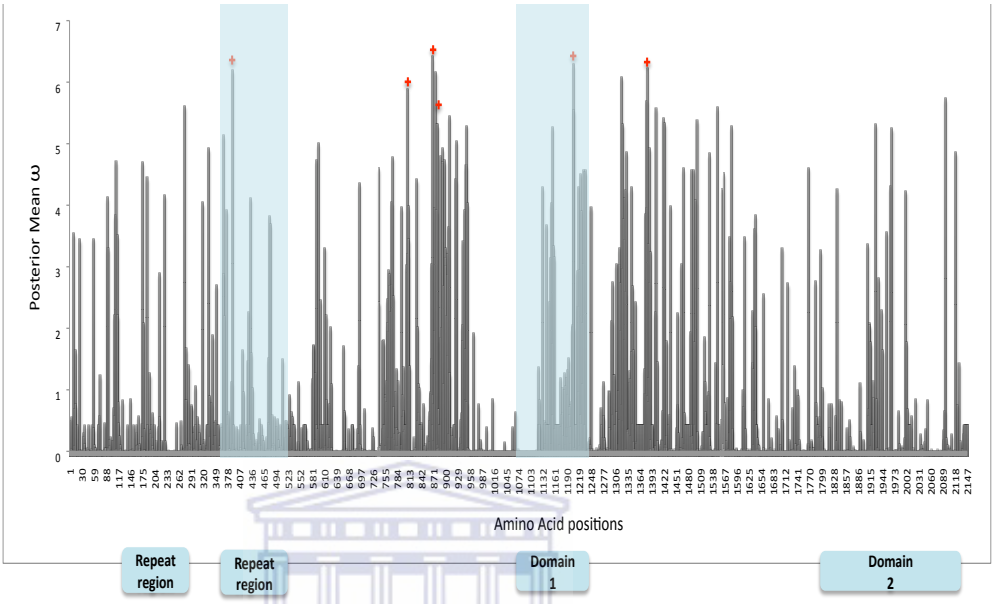
Gene IDs	Functional Annotation	GO Category
TMP005595	ATP-dependent RNA helicase dhx 36-like	Amino_acid_biosynthesis
TMP006377	coenzyme Q-binding protein	Amino_acid_biosynthesis
TMP004552	DNA polymerase epsilon	Amino_acid_biosynthesis
TMP003203	DNA replication licensing factor mcm7	Amino_acid_biosynthesis
TMP005511	atg5	Amino_acid_biosynthesis
TMP007621	26s proteasome regulatory complex subunit psmd5	Amino_acid_biosynthesis
TMP006591	saccharopine dehydrogenase domain-containing protein	Amino_acid_biosynthesis
TMP010016	tpr repeat-containing protein	nucleotide biosynthesis
TMP002520	nucleoporin	nucleotide biosynthesis
TMP004047	GJ14155	nucleotide biosynthesis
TMP003021	kelch-like protein diablo	nucleotide biosynthesis
TMP004517	importin9	nucleotide biosynthesis
TMP004763	spire	nucleotide biosynthesis
TMP008441	gamma-tubulin ring complex	nucleotide biosynthesis
TMP009332	GF15465	Biosynthesis_of_cofactors
TMP003577	st7-like protein	Biosynthesis_of_cofactors
TMP008943	Glucosidase	Cell_envelope
TMP002863	angiotensin-converting enzyme	Cell_envelope
TMP007548	retinoid-and fatty acid-binding glycoprotein	Cell_envelope
TMP007537	integrator 4	Cell_envelope
TMP011786	disulfide isomerase	Cell_envelope
TMP004413	Wnk,isoformA	Central_intermediary_metabolism
TMP011617	E3 ubiquitin-protein ligase	Central_intermediary_metabolism
TMP007653	Abdominal-B protein	Central_intermediary_metabolism
TMP004240	elongation factor1 alpha	Central_intermediary_metabolism
TMP010317	general transcription factor IIH subunit1	Central_intermediary_metabolism
TMP013886	trna(uridine-2-o-)-methyltransferasetrm7	Central_intermediary_metabolism
TMP004201	acetyl cholinesterse	Energy_metabolism
TMP013948	coiled-coil-helix-coiled-coil-helix domain-containing protein2	Energy_metabolism
TMP003059	serine/threonine-protein kinase	Energy_metabolism
TMP011421	tm2d1_drome	Energy_metabolism
TMP011104	serine-pyruvate mitochondrial	Energy_metabolism
TMP010707	GH18161	Transcription and translation
TMP007137	hypothetical conserved protein	Transcription and translation
TMP007216	kinesin heavy chain	Transcription and translation
TMP002718	shaggy	Transcription and translation
TMP004121	zinc finger protein2	Transcription and translation
TMP004581	hypothetical protein	Transcription and translation
TMP012225	phd finger protein	Transcription and translation
TMP014030	GJ12585	Transcription and translation
TMP002587	alternative splicing factor SRP20/9G8	Transcription and translation
TMP005243	40S ribosomal protein	Transcription and translation
TMP007593	barren	Transcription and translation
TMP008942	eukaryotic Transcription and translation initiation factor 3	Transcription and translation
TMP009157	ubiquitin-conjugating enzyme9	Transcription and translation
TMP003617	host cell factor	Transcription and translation
TMP013958	PERQ amino acid-rich with GYF domain-containing protein	Transcription and translation
TMP008448	autophagy-related protein	Transcription and translation
TMP003060	GK22170	Transcription and translation
TMP014204	60s ribosomal protein l23	Transcription and translation
TMP014133	unc-50	Transport
TMP010673	Ferritin heavy chain	Transport

Significantly repressed post-blood meal:

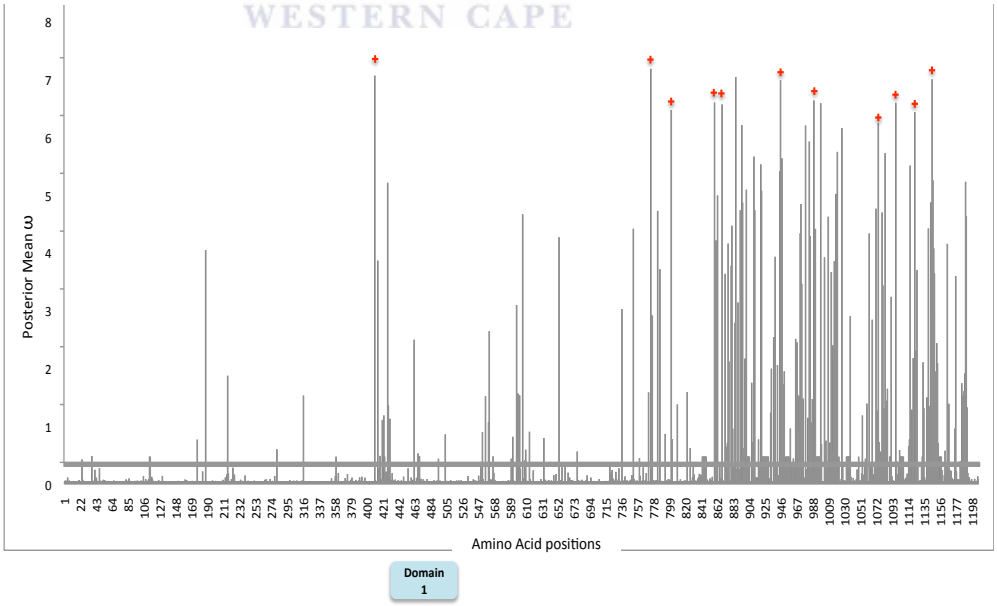
Gene IDs	Functional Annotation	GO Category
TMP007247	beaten path IIa	Amino_acid_biosynthesis
TMP007441	mitochondrial processing peptidase beta subunit	Amino_acid_biosynthesis
TMP007030	ganglioside-induced differentiation-associated protein 1	nucleotide biosynthesis
TMP012414	Sticks and stones, isoform A	nucleotide biosynthesis
TMP003602	phosphoinositide3-kinase	nucleotide biosynthesis
TMP008854	receptor tyrosine kinase	nucleotide biosynthesis
TMP007350	defective proboscisextension response	Biosynthesis_of_cofactors
TMP005554	proton-associated sugar transporter a-like	Biosynthesis_of_cofactors
TMP002406	Cuticle protein	Cell_envelope
TMP003500	beta-hexosaminidase fdl	Cell_envelope
TMP010987	sam-motif ubiquitously expressed punctatedly localized protein	Central_intermediary_metabolism
TMP005160	homeobox protein	Central_intermediary_metabolism
TMP011448	brain chitinase and chia	Central_intermediary_metabolism
TMP009552	synaptotagmin	Energy_metabolism
TMP012368	GJ11024	Energy_metabolism
TMP005219	Unknown	Energy_metabolism
TMP005357	LIX1-like protein	Energy_metabolism
TMP009871	bruno-3 transcript	Energy_metabolism
TMP010821	Abdomnal A	Transcription and translation
TMP002357	dynein lightchain	Transcription and translation
TMP004553	mitochondrial f1f0-atp synthase subunit cf6	Transcription and translation
TMP007741	meiosis-specific nuclear structura l	Transcription and translation
TMP006315	acylphosphatase	Transcription and translation
TMP013100	sox box protein isoform a	Transcription and translation
TMP013254	btk family kinase at isoform a	Transcription and translation
TMP003369	Serine/threonine-protein kinase	Transport
TMP006868	5-hydroxy tryptamine receptor1	Transport

Appendix D

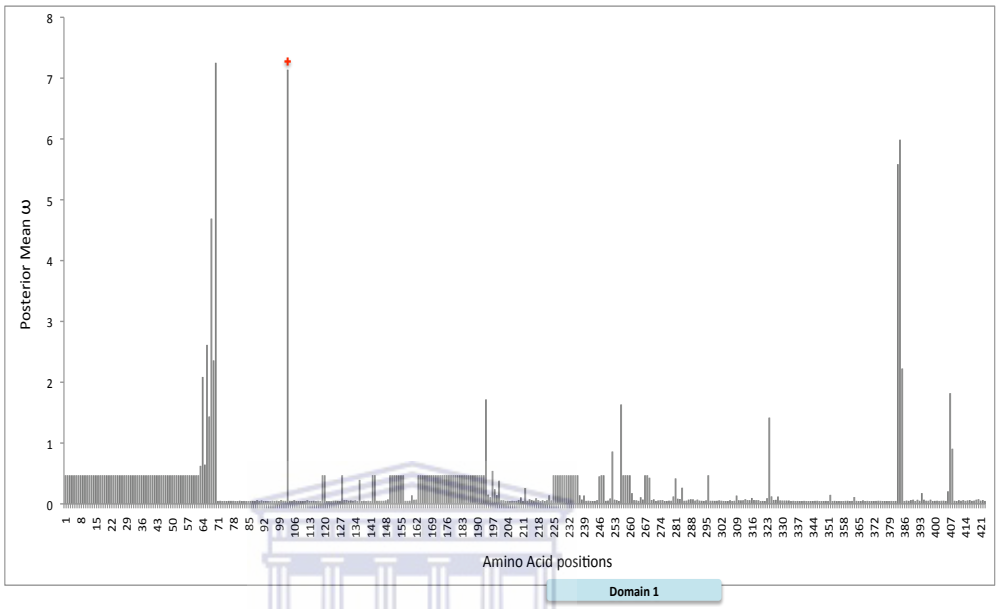
Evolution of IRE-regulated genes



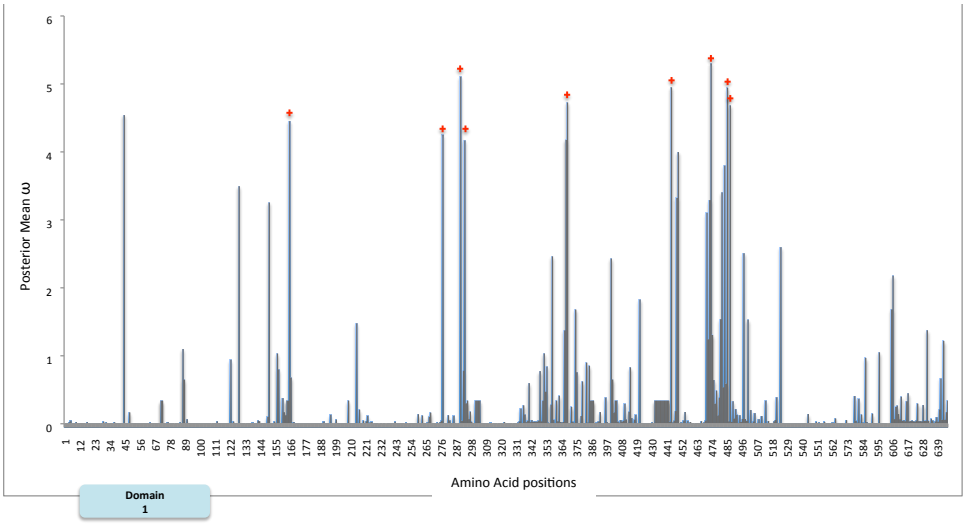
a)



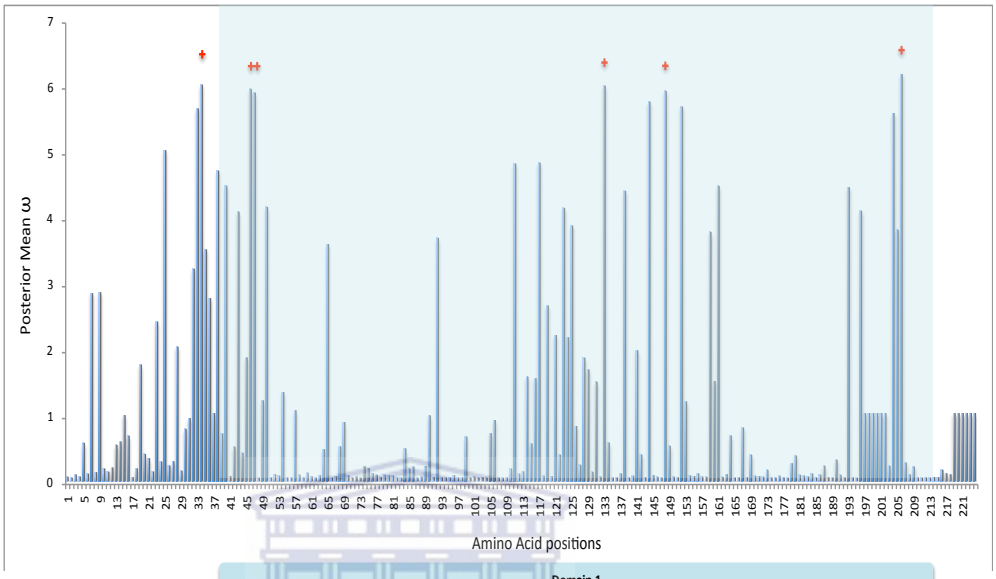
b)



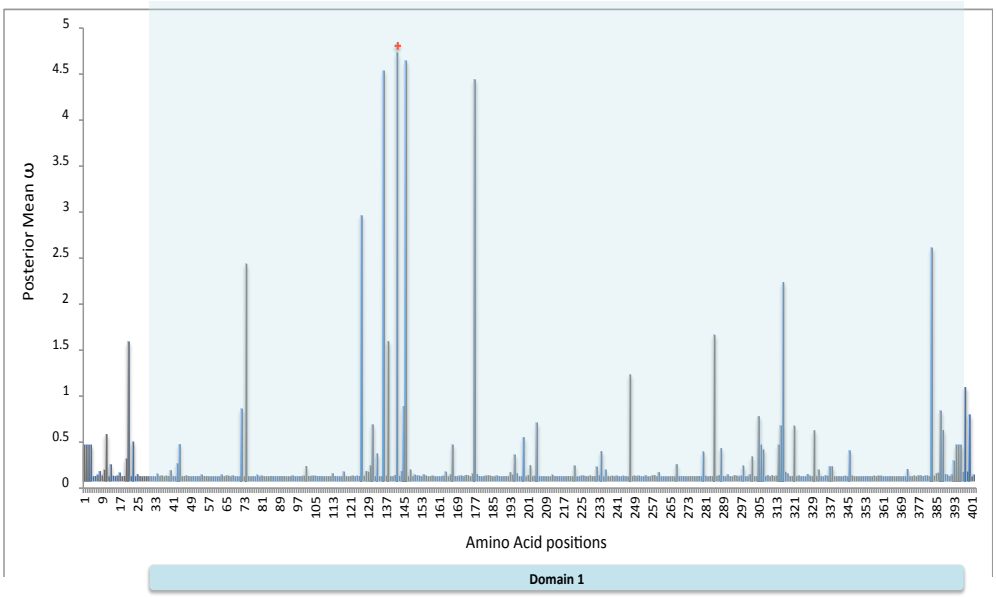
c)



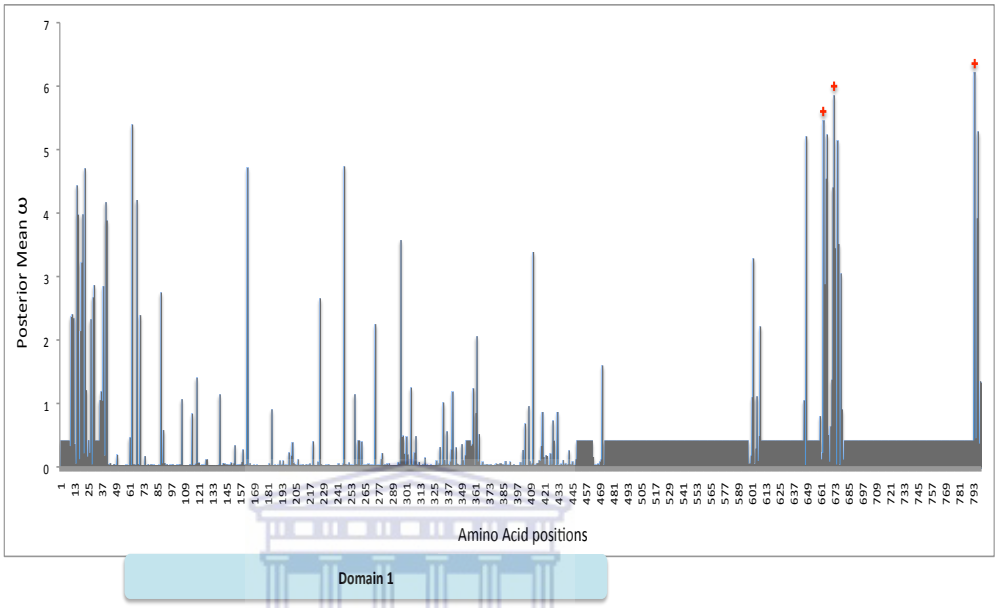
d)



e)



f)



g)

Figure D.1: Distribution of positively selected sites and their functional implications. Bar graph presenting positively selected amino acid sites indicated with "*". The blue bars in the x-axis illustrates the location of domains for each gene. Accordingly each figure represents the following: a) TMP002520, b) TMP008942, c)TMP009871, d) TMP010317, e) TMP010673, f) TMP011104, g)TMP011448.

UNIVERSITY of the
WESTERN CAPE