A proposal for tsetse fly (*Glossina*) genome projects

Submitted April 30, 2010

by Serap Aksoy on behalf of International *Glossina* Genomics Community

Executive Summary

The *Glossina* (tsetse flies) are vectors of African trypanosomes, which are of medical and economic importance. 1.35 million DALYs are lost due to Human African Trypanosomiasis (HAT) and mortality related to HAT is ranked ninth out of 25 among the human infectious and parasitic diseases in Africa. Animal African Trypanosomiasis (AAT) causes approximately 3 million cattle deaths per year and farmers administer about 35 million doses of trypanocidal drugs at a cost of US$ 1-1.2 billion, resulting in total agricultural losses estimated at US$ 4.75 billion per year. Despite considerable research into trypanosomes, the toolbox for disease control is limited with neither vaccines nor effective and affordable drugs available in the near future. The African Union has made removal of trypanosomiasis via tsetse control a key priority for the continent. New and/or improved control tools will be developed through the expanded genomic resources proposed here.

This document proposes sequencing of 5 *Glossina* genomes (average size 400 Mb) and two related Dipterans, a non-vector obligate blood feeder (stable fly, *Stomoxys calcitrans*) and a non-blood feeding mechanical vector of numerous human pathogens (the house fly, *Musca domestica*). The data will complement and facilitate comparative analysis with the nearly complete *Glossina morsitans morsitans* genome and the available multiple *Drosophila* genomes. We propose deep sampling of two human vector species belonging to the Palpalis group and an animal vector species closely related to *G. m. morsitans*, followed by sampling at increasing evolutionary distances within *Glossina* and the related Dipterans. We also propose transcriptome sequencing for each species in support of genome annotation and functional comparative studies as a high priority. Sequencing of desirable populations at low coverage will enable SNP discovery projects to expand functional genomics studies and future field investigations.

Summary of Proposed Genomics Activities for 5 Genomes Cluster for genus *Glossina* and *related Diptera*

Activities	Colony (Source)	Parasites
1. *G. palpalis* WGS	*Bratislava*	Vector of *Tbg* in West Africa
G. fuscipes WGS	*Bratislava*	Vector of *Tbr* in East Africa
G. pallidipes WGS	*Johnson Ouma, Kenya*	Major vector of *Tbb*, minor vector of *Tbr*
G. brevipalpis WGS	*Otto Koekemoer, South Africa*	Vector of AAT (ancestral species)
G. austeni WGS	*Atway Masangi, Tanga*	Vector of AAT (phylogeny controversial)
2. *Stomoxys calcitrans WGS*	*Mike Lehane, UK*	
Musca domestica WGS	*Jeff Scott, Ithaca, USA*	
3. Transcriptome data on WGS projects		
4. Low coverage WGS of key field populations (10-12)		

Genomics information from the key *Glossina* species transmitting HAT and/or AAT will provide researchers with genes important for (1) vector competence, haematophagy and viviparity, (2) species specific sequences that could account for differences in their vectorial, host-seeking and discriminatory biology, (3) SNPs and genotyping capabilities to perform population level studies in support of vector control efforts as well as (4) genome wide association studies on phenotypes that are relevant to *Glossina*-trypanosome biology. This white paper has strong support from vector biologists and members of the tsetse, trypanosome community, house fly and stable fly communities in addition to the interest and commitment of geneticists, evolutionary and computational biologists whose contributions to this project will aid in the analysis of the data and quicken the pace of discovery.

1. Introduction

Tsetse flies (Diptera:Glossinidae) are vectors of pathogenic trypanosomes. Chief among these is *Trypanosoma (Trypanozoon) brucei rhodesiense* and *T. b. gambiense*. *T. b. gambiense* infection causes a chronic, slow wasting disease that ultimately causes death if untreated. This form of HAT occurs in West and central Africa (2). *T. b. rhodesiense* is zoonotic and causes a much more acute disease in humans that is rapidly fatal if not properly treated. *T. b. rhodesiense* occurs east of the Rift valley. 1.35 million DALYs are lost due to HAT and mortality related to HAT is ranked ninth out of 25 among the human infectious and parasitic diseases in Africa (3). Devastating epidemics in the 20th century resulted in hundreds of thousands of deaths in sub-Saharan Africa (4) but more effective diagnostics now indicate that data concerning sleeping sickness deaths are subject to gross errors due to under-reporting (5). With hindsight it is reasonable to infer that millions died from sleeping sickness during the colonial period. Loss of interest in, and funding for control programmes within the endemic countries resulted in a steep rise in incidence after the post-independence period of the 1960s. As surveillance and health care programmes in the affected regions were slowly restored, there has been a decline in reporting of new cases recently (6). However, lack of effective disease management tools cause sixty million people to live at risk for HAT in 37 countries covering ~40% of Africa (11M km^2).

In addition to the public health impact of HAT, nagana or Animal African Trypanosomiasis (AAT) caused by *T. b. brucei* and related trypanosomatids, *T. congolense* and *T. vivax*, limits the availability of meat and milk products in large regions of Africa. It also excludes effective cattle rearing from ten million square kilometers of Africa (7) with wide implications for land use; i.e., constraints on mixed agriculture and lack of animal labor for ploughing (8). The Programme on African Animal Trypanosomiasis (PAAT) estimate that AAT causes approximately 3 million cattle deaths per year and farmers are required to administer approximately 35 million doses of costly trypanocidal drugs. Economic losses in cattle production are estimated at US$ 1-1.2 billion and total agricultural losses caused by AAT are estimated at US$ 4.75 billion per year (9). In 2000 the African Union recognized trypanosomiasis as "one of Africa's' greatest constraints to socio-economic development" (10). This disease is having a major impact on the health and development prospects of large numbers of the most marginalized people in the world.

Despite the fact that trypanosomes form some of the best-studied organisms in biology, mammalian vaccines are not available and are unlikely to be developed due to the antigenic variation capacity of trypanosomes. Active surveillance and treatment of patients are essential for disease control. Unfortunately HAT treatment relies on dangerous and expensive drugs, which fortunately at the present time are provided free of charge by WHO thanks to a public-private partnership with Sanofi-Aventis (11-13). To further complicate control efforts, trypanosomes are increasingly becoming resistant to the drugs with at least 20% of patients not responding to the available drugs in the recent epidemic in Uganda (14, 15). The parasites *T. b. brucei*, *T. b. gambiense, T. congolense* and *T. vivax* genomes have been completed and several other kinetoplastid genomes are currently being sequenced (16-18). It is hoped that the new generation of genomics projects for African trypanosomes will lead to new drug targets for disease control. However, difficulties with the delivery of drugs to those in need of treatment in remote areas of subSahara will continue to remain paramount.

In addition to human-fly contact, animal reservoirs have been documented especially in the case of Rhodesiense disease in East Africa. Modeling control options (whether to target the animal reservoir, humans or vector) has shown that given the significance of the animal reservoir, vector

control is by far the most efficient method for bringing outbreaks under control (2). Vector control tools include synthetic pyrethroids applied as aerial sprays as was the case for the Okavango Delta in Botswana (19). Insectide pour-on applications are used on animals in farming communities. Although traps and targets can be effective in reducing local tsetse populations, they are not widely explored in HAT control due to lack of effective attractants for human disease transmitting tsetse species (20, 21). A genetic based vector suppression strategy, Sterile Insect Technique (SIT) can also be effective for tsetse control as has been successfully used to eliminate tsetse on Unguja Island, Zanzibar (22). Essential for the eventual success of vector control tools is the identification of genetically isolated populations to prevent recolonization of cleared areas with populations from neighboring areas.

The broad goal of this proposal is to undertake comparative analysis of the genomes of a total of 5 different *Glossina* species that have been selected for their vectorial capacity and genetic relatedness and 2 related flies in family Muscidae that have been selected for their evolutionary relationship and the vast differences they display in their vectorial and overall biological traits. This is in addition to the nearly completed *Glossina morsitans morsitans* genome for which extensive genomic information is currently available. Unlike the genomes of some mosquitoes such as *Aedes* and *Culex* and ticks, the relatively small size of the tsetse genomes (approximately 400 Mb) and the muscid genomes (about 310 Mb) makes this project feasible. The insights gained from these comparative genomic studies will have several significant applications on HAT and AAT control, through a greater understanding of vectorial capacity, development of enhanced or new vector control tools and improved application strategies.

2. Background

2.1 *Glossina* are members of higher Diptera. Hippoboscoidea is a superfamily of Diptera that contains the Glossinidae (tsetse flies), the Hippoboscidae (louse flies), and the Streblidae and the Nycteribiidae (two families of bat flies). Molecular phylogenetic analysis supports the monophyly of the Hippoboscoidea as a whole (Fig 1)(1). Its placement within the Calyptraea remains inconclusive although Hippoboscoidea are placed deeply nested within the Calyptratae. *Musca domestica* and *Stomoxys calcitrans* are members of the superfamily Muscoidae within Calyptratae. The current consensus is that Hippoboscoidea is the sister group to Muscoidae. The proposed insects are related to *Drosophilatids* for which extensive genomics data is available, which should expedite genomic analysis. The relationship of *Glossina* to *Drosophila* is validated by the genetic synteny we have observed in the six tsetse BAC clones sequenced to data as well as in the overall gene similarity scores obtained in our EST projects. This close relationship will help the genome assembly processes. However, the biology of the tsetse and Drosophila, especially the obligate haematophagy and viviparous reproduction in tsetse is so different that it is bound to provide tremendous biological insights into the hugely important blood-sucking way of life. Information on the closely related blood-feeder (but non trypanosome vector) *Stomoxys calcitrans,* and a closely related non blood-feeder and non-trypanosome vector species, *Musca domestica* will provide many insights into the vectorial genetics of *Glossina*.

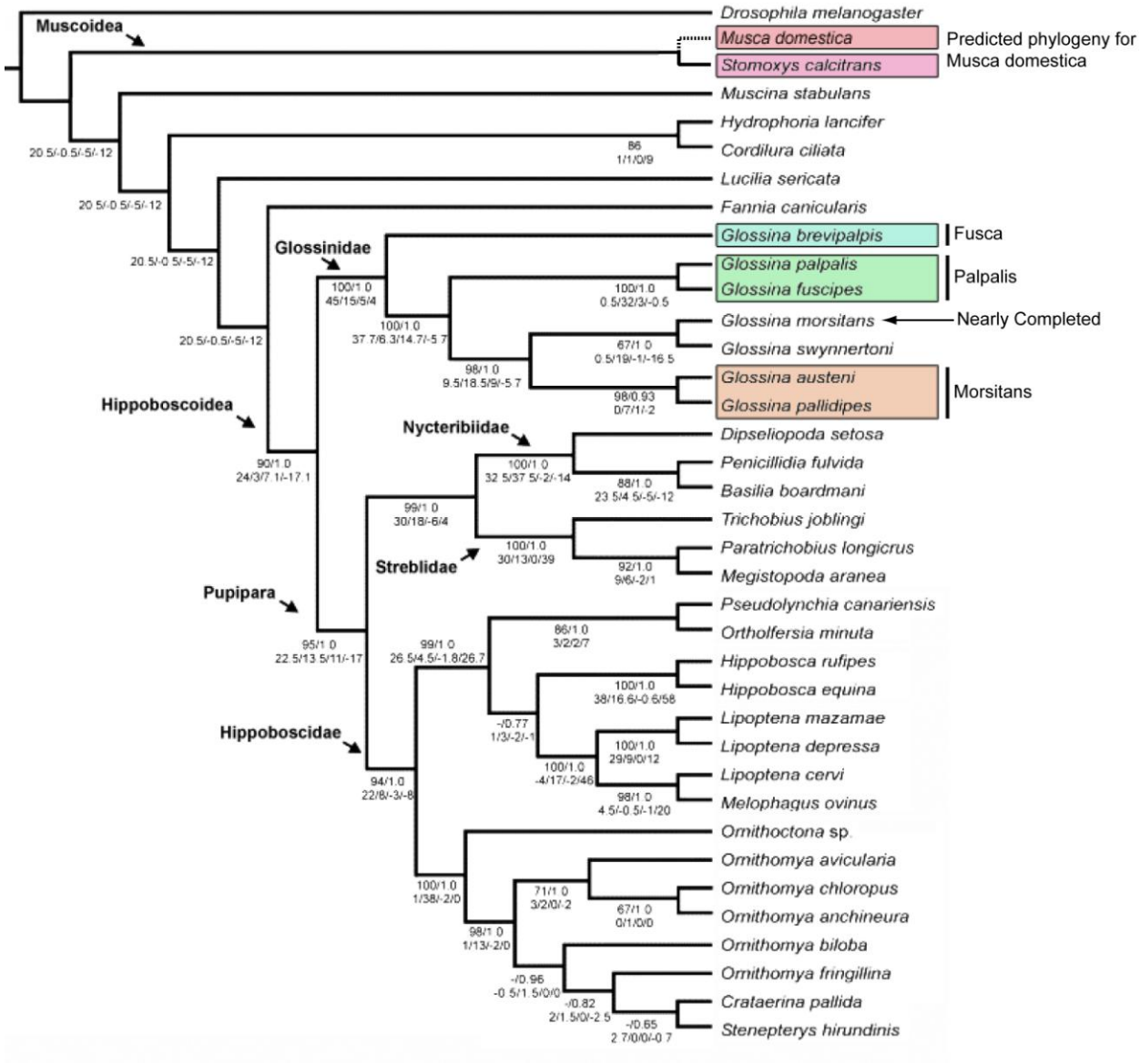

Fig 1. Maximum parsimony tree based on combined sequence data from CAD, COI, 16s, and 28s (numbers in first line are bootstrap support values and posterior probabilities; = bootstrap support <50; numbers in second line are PBS values for CAD/COI/16s/28s. Modified from (*1*). The relationship of the *Glossina* subgroups and associated species targeted for sequencing are identified by color. The positioning of *Stomoxys* and its predicted relationship with *Musca* are denoted.

Molecular taxonomy supports the monophyly of family Glossinidae, as the sister group to all Pupipara (Fig. 1). *Glossina* are free-living and only come into close contact with their host during feeding while the other three families are all genuine ectoparasites (i.e., species with a trophic and a spatial association to host) spending all or most of their adult life within the fur or among the feathers of their mammal (bats) and bird hosts. Strikingly, within Hippoboscoidea true ectoparasitism originated only once in the common ancestor of the Pupipara.

This specialization must have involved a change from a free-living, and blood-feeding fly (e.g., Glossinidae) to a fly with an obligate and close association with a particular vertebrate host. This specialization could also have contributed to the large observed difference in species diversity between the sister groups Glossinidae (22 spp.) and Pupipara (630 spp.) given that it has been postulated that specialization of feeding structures, host finding behavior, and population subdivision associated with parasitism can spur species diversification. Mammal feeding appears to be ancestral for the Hippoboscoidea.

2.2 Unusual reproductive biology of Hippoboscoidea.

One of the most remarkable morphological and physiological adaptations within Hippoboscoidea, including *Glossina*, is adenotrophic viviparity. A zygote develops and hatches in the female's reproductive tract and the larva feeds on "milk" produced by the female's reproductive accessory glands until it completes its development. The fully mature 3rd instar larva is deposited and quickly pupates within its last larval skin. The viviparous nature of *Glossina* reproduction means that each female can produce a total of about 8-10 progeny during her lifetime. Because of this low reproductive capacity, control methods relying on tsetse population reduction have been highly successful. Glossinidae, indeed all Pupiparia, are exclusively haematophagous. This highly restricted nutritional ecology supporting all developmental stages has resulted in obligate adaptations with symbiotic bacteria in tsetse. The endosymbionts provide nutritional supplements, in the absence of which females are rendered sterile. Tsetse's reliance on its obligate microbiota for reproduction provides a weak-link in its biology, and may generate alternative control strategies that should be further explored.

2.3 Three species groups of genus *Glossina*.

Within the *Glossinidae*, 33 extant taxa have been described of 22 species in 3 subgenera; Austenina Townsend, Nemorhina Robineau-Desvoidy, and Glossina Wiedemann that correspond to the Fusca, Palpalis, and Morsitans species groups respectively (described in (23). A new subgenus, Machadomia Dias 1987, has been proposed to incorporate the anomalous tsetse, *G. austeni* Newstead although its relationship with respect to the Palpalis and Morsitans complex flies remains controversial. Molecular taxonomy shows that two of the species groups (Palpalis- and Morsitans-species groups) are recovered as monophyletic with strong support (Fig 1). *Glossina brevipalpis* of the Fusca species group emerges as sister group to all remaining Glossinidae. The taxonomy of the bacterial obligate symbiont (genus *Wigglesworthia*) from different tsetse species also show the same relationships indicating concordant evolution between *Glossina* species and their endosymbionts (24).

Morsitans group flies are largely savanna and woodland inhabitants, although *G. pallidipes* may also be found in forests. Morsitans group taxa are adapted to drier habitats than the other two subgenera (25). Palpalis group flies tend to occur in riverine and lacustrine habitats. Fusca group flies largely inhabit moist forests of West Africa although *G. brevipalpis* occurs discontinuously in East Africa, Zaire, and Mozambique. The host-specificity of the different species groups vary, with the *palpalis* group flies displaying strong antrophilicity while the others are more zoophilic in preference.

2.4 Tsetse flies are vectors of pathogenic trypanosomes.

The principal vectors of HAT include *G. palpalis* s.l., *G. fuscipes* and *G. morsitans* s.l. The riverine habitats of Palpalis group flies and their adaptability to peridomestic environments along with human blood meal preferences make them excellent vectors for HAT. Other species

belonging to the Morsitans group (such as *G. pallidipes*) can also transmit human disease, but principally play an important role in AAT transmission. In particular, *G. pallidipes* has a wide distribution and has a devastating effect in East Africa.

2.5 Genetics of Vector Competence.

Trypanosoma brucei spp. salivary gland infection rates in tsetse flies are low, typically < 1% in the field and also in the laboratory when all flies are exposed to an infectious blood meal (26). Much higher infection rates with *T. brucei* may be expected given the great longevity of the flies (adult daily survival rates typically exceed 97% and half lives exceeding ~ 28 d) and prevalence of trypanosome-infected hosts. In contrast, the same tsetse species can transmit *T. congolense* and *T. vivax* more efficiently. Studies to date have shown that the pathway from ingestion of blood forms by tsetse flies to the production of infective metacyclic trypomastigotes of *T. brucei* in the tsetse salivary glands is long and complex, with many critical steps that developing trypanosomes must overcome. Once established in the midgut, procyclic trypomastigotes must transform to epimastigotes and proceed via the proventriculus and foregut to the host salivary glands where they develop to infective metacyclic trypanosomes. Only a small fraction of successful gut infections succeed in developing into infective metacyclic forms lodged in the vector salivary glands. A clearer understanding of this most unusual situation (completely different to sand flies, blackflies, mosquitoes etc) may provide key pointers to development of new control strategies based on genetic modification approaches.

A likely reason that tsetse infection rates are small is they have a robust innate immune system in which most ingested trypanosomes die in the fly midgut. Functional studies resulting from the available genomics data to date using a post-genomic tool RNAi (provided to flies by either injecting or feeding the corresponding dsRNA) already incriminate host proteins in parasite resistance. A novel protein EP-rich TsetseEP (27) and the antimicrobial peptides produced by the IMD immunity pathway (such as Attacins) have been implicated in parasite resistance in the midgut (28-31). In addition the obligate symbiont *Wigglesworthia* has been shown to influence tsetse's immunology by inducing the expression of the host protein Peptidoglycan Recognition Protein (PGRP-LB), which has putative antitrypanosomal actions in the midgut (32). Almost all investigations into tsetse's immunobiology have been performed on *G. m. morsitans*, for which we have some genomics information. Lack of available genomics resources has prevented our ability to similarly investigate host-parasite interactions in the HAT disease transmitting species of the Palpalis group. The availability of extensive transcriptome information on these species early in our proposed studies will make these studies feasible in these important vector species.

2.6 Tsetse population genetics.

Interest in tsetse population genetics has recently developed, fueled by the notion that genetic methods can promote effective area-wide population management by (1) better defining taxonomic units (33) and estimating degrees of isolation of populations of interest (34) or (2) modulating vector competence by genetic means (35). Only limited molecular markers are available for tsetse species but where applied, populations have been found to be highly structured, suggesting that control studies informed by population genetics data can succeed in limiting tsetse populations in selected areas.

3. Rationale

3.1. Multiple tsetse genomes will provide a framework to illuminate the genetic basis of vectorial capacity.

Two medically relevant questions (expanded upon below) in vector biology include: (1) Why do *Glossina* transmit African trypanosomes and not other genera? (2) Why are some *Glossina* species more efficient vectors of HAT than others? Answering these questions requires an in-depth understanding of the key traits that determine vectorial capacity. Our goal of facilitating comparative analysis of the multiple *Glossina* genomes outlined here would establish a rich data source and generate a framework for gathering this information and answering such questions.

(1) Why do Glossina transmit African trypanosomes and not other genera?

Palpalis group flies are the principal vectors of HAT while Morsitans and Fusca group flies are better at transmitting agents of AAT. Also, *Glossina* is in general more resistant to transmission of *T. brucei* complex parasites than transmission of *T. congolense* and *T. vivax*. This observation is not only present in field infections but also in laboratory challenges. The molecular basis of the infection barrier, *i.e* whether due to the lack of specific receptors, or to variable immunity, or to other mechanisms, and the variation in infection barriers among non-vector species remain unknown. Comparative genomics information of the different species of tsetse with differing vectorial capacity would be important for answering this question.

Information gained from the stable fly *Stomoxys calcitrans* and the house fly *Musca domestica* in the sister superfamily Muscoidea will provide invaluable genomics data to understand tsetse's vector competence genetics. *Stomoxys,* like tsetse is an obligate blood feeder and acts a mechanical vector of *Trypanosoma vivax* although it is unable to transmit human disease causing parasites. Although closely related to tsetse, *Stomoxys* has been shown to synthesize two unique defensins, Smd1 and Smd2, in the anterior midgut tissue of the bloodsucking fly, which may result in its non-vectorial status (36). Molecular information on the house fly is sparse to provide an initial comparison (37). Genome information on *Musca* will be of broad interest to vector biologists as it will advance our knowledge on its unusual mechanism for sex determination as well as mechanisms involved in insecticide resistance (*38*), which is a major public health concern.

Specific knowledge on *Glossina* evolution and vector competence will be obtained from genome comparisons involving the species targeted here with *Stomoxys, Musca* and the available *Drosophila* genomes in addition to those other Dipteran species (mosquitoes and sand flies) currently under investigation.

(2) Why are some tsetse species more efficient vectors than others?

This question acknowledges differences in vector ability between species. To transmit HAT parasites efficiently, the tsetse must have a high probability of feeding on humans. The genetic basis for the differences associated with host preference in different species complexes are unknown but are likely to involve odorant binding proteins and associated receptors. Comparative genomic analysis will identify these repertoires from the different species of tsetse proposed here. In addition tsetse's immune gene composition may vary between species and result in the resistance observed. Species-specific variations between tsetseEP proteins implicated in parasite resistance (*39*) and a trypanolytic lectin like protein has been uniquely identified in the Palpalis flies (*40*). Salivary gland protein profiles analyzed from different tsetse species vary (Aksoy unpublished). It is possible that this variability corresponds to either differential gene expression or the presence of different proteins such as receptors, which may be necessary for parasite maturation in salivary glands. Availability of the different genomes will allow for comparison of immunity genes and salivary gland genes as an initial attempt to understand the genetic basis of

susceptibility. In addition, a role for tsetse endosymbionts has also been implicated in parasite transmission. It is also possible that tsetse-endosymbiont interactions resulting in differential host immune regulation or nutritional environment in the midgut may also contribute to the differential vector susceptibility traits observed in field populations and between species. Specific answers to these questions will involve tsetse genome comparisons of factors important for olfactory physiology, comparisons of tsetse's endosymbiont genome sequences, which we will discover as part of this genomics proposal.

Knowledge on genes related to host trypanosome resistance mechanisms can be immediately used to generate refractory strains of tsetse to be used in SIT programs to increase the efficacy of their application in human disease endemic areas (*41*). A paratransgenic transformation strategy has been developed where antitrypanosomal genes can be expressed in the midgut in tsetse's commensal symbiont *Sodalis* (*42, 43*).

3.2 Extensive transcriptome data will improve annotation and identify genes and regulatory sequences that may allow some individuals within a species to be parasitized by trypanosomes

(1) Improved annotation of the Glossina genomes
Availability of extensive transcriptome information from the new species along with the draft genome of *G. m morsitans* will be essential for the annotation of the proposed genome sequences.

(2) Why are some individuals within a species parasitized by trypanosomes while others are resistant?
In the laboratory, only 5-10% of flies challenged with parasites by the same infected blood meal give rise to midgut parasite infections, while the rest are able to clear parasite infections. Young flies have been shown to be much more susceptible to infection than older adults (*44*). Parasite resistance in the midgut is typically associated with a parasite attrition process early during the infection, typically 3 days post parasite acquisition. The variability observed among individuals may arise from differential expression of tsetse's midgut proteins or may arise from differences in midgut endosymbiont densities or composition, which may influence host gene expression. Our studies have shown that those flies that are cleared of their obligate symbiont *Wigglesworthia* have much higher susceptibility to parasite infections and that differential expression of the host protein *PGRP-LB* is responsible for the difference (*45*). Studies have also noted a sex-bias associated with parasite maturation process in salivary glands as males are at least twice more susceptible to parasite infections than females (*46*). There are also recent studies showing that parasite infections in salivary glands change host gene expression profile in ways that enable parasite transmission processes in the mammal (*47, 48*). Comparative analysis of salivary gland transcriptomes of parasitized and normal flies can open up new investigations leading to control of parasite transmission in the human host. Specific answers to these questions will involve immune transcriptome comparisons of young and old adults, male and female salivary glands and normal and parasite infected salivary glands.

3.3 Population genetics
The African Union has pledged to eliminate trypanosomiasis from Africa altogether via tsetse eradication, the Pan African Tsetse and Trypanosomoses Eradication Campaign (PATTEC) (http://www.who.int/trypanosomiasis_african/partners/pattec/en/index.html), an effort funded by the African Development Bank. Key to the eradication strategy adopted will be the identification of isolated tsetse populations not prone to reinvasion following clearance. Many groups are now waiting for the genome sequences to provide the population genetics markers required to carry out

that work. A very important near-term benefit of genomics will be its impact on our knowledge of vector population biology. This information has the immediate potential to improve the efficacy and implementation of the current control programs on the ground (*34, 49*).

Population genetics analysis of *G. fuscipes* has indicated extensive genetic structuring in the field in East Africa (*50*). These populations now provide a unique opportunity to undertake a SNP based genomic scan analysis to understand tsetse's vector competence traits. Similarly there are *G. palpalis* populations in West Africa with differential transmission dynamics and host seeking behavior, which are ripe for genomic scans. Population genetic studies of *G. palpalis* in West Africa, using microsatellites, have shown structuring at microgeographic scales (*51*). When applied to tsetse control, these studies have shown that in some areas tsetse populations are genetically isolated [Guinea (*51, 52*); Senegal (manuscript submitted)], indicating that eradication campaigns may be feasible and sustainable in such areas.

3.4 Genomic information on olfactory biology from different tsetse species will enhance vector control tools

Unlike some other vectors, traps and targets have been developed to attract tsetse flies, in particular for the species in the Morsitans group. They are extensively used for control of AAT locally in farming communities. Knowledge on tsetse's olfactory physiology can result in enhancement of trapping technologies for other species, especially for the human disease transmitting Palpalis group (*53, 54*). The availability of more efficacious traps would then warrant their inclusion in the human disease control programs. At the core of attractants are tsetse's odorant binding proteins and olfactory receptors. Genomic information on odorant binding proteins and receptors from multiple species will pave the way for functional genomic studies to identify chemicals to be included as attractants in traps.

3.5 Genomic information on tsetse's reproductive physiology will enhance vector control tools

Tools that aim to reduce tsetse populations have been highly effective due to tsetse's low reproductive capacity. Understanding the genetic processes and mechanisms that enable viviparity can lead to new methods to interfere in tsetse's reproductive capacity. In particular proteins expressed in tsetse's accessory glands (milk proteins) are essential for supporting larval development. To date, accessory gland transcriptome analysis has identified a number of proteins that are uniquely expressed in the milk and under transcriptional regulation in sync with larval developmental events (*43, 55, 56*). Studying the transcriptional regulation of these proteins and their promoter regions for identification of transcriptional factors involved in their regulation can lead to novel chemicals that can interfere with their synthesis.

3.6 Genomics information that will result from this project will expand the scientific community working on host-parasite interactions and HAT control.

Despite advances in the field of vector genomics, a key roadblock to advancing research in tsetse remains to be the small size of the tsetse research community. Most of the life cycle stages of African trypanosomes occur in tsetse fly vectors and remain unexplored due to lack of genomic resources to launch functional studies. We argue that more knowledge of the tsetse fly at the molecular level will attract new, high quality laboratories to study tsetse flies as has happened in the *Anopheles* field in the last decade. In particular, given the close evolutionary relationship *Glossina* has with *Drosophila* despite their vast physiological differences, we argue that the *Glossina* genomics resources will be very attractive to the *Drosophila* community. Including the house fly and stable fly genomes will immediately expand the pool of scientists that will draw

from this data. A larger scientific community will help generate resource development in addition to promoting training and capacity building in disease endemic countries.

In summary, the availability of comparative genomics information from the key *Glossina* species transmitting HAT and AAT will provide researchers with genes important for (1) vector competence, haematophagy and viviparity, (2) species specific sequences that could account for differences in their vectorial and host-seeking biology, (3) SNPs and genotyping capabilities to perform population studies in support of vector control efforts as well as (5) genome wide genome association studies on phenotypes that are relevant to *Glossina*-trypanosome biology.

4. Sequencing targets, priorities and considerations

4.1. Small genome size of target species

The genome size of different *Glossina* species was determined using flow cytometry with intercalating dyes (*57*), Table 2). In general they were found to be about 1.2-1.5 times the haploid genome size of *Drosophila virili*s, which is about 350 Mb in size.

Table 2. Genome size estimates of different *Glossina* species

Glossina species analyzed		Haploid genome size (pg or Gb)	Ratio *Glossina*/*D. virilis**
G. m. morsitans	Male	0.579 (0.590^)	1.546
	Female	0.613 (0.596^)	1.634
G. pallidipes	Male	0.509	1.356
	Female	0.533	1.422
G. p. palpalis	Male	0.482	1.285
	Female	0.479	1.278
G. fuscipes	Male	0.534	1.523
	Female	0.524	1.398

*The haploid genome size of *D. virilus* has been estimated as 0.34–0.38 pg
^Values independently determined by Dr. Spencer Johnston, Texas A&M.
All values determined using FaxCalibur flow cytometer by Biemont Christian, Christiane Nardon and Michèle Weiss, at Université Lyon, France

More than 2.4 million capillary shotgun reads have been produced from *G. m. morsitans*, which assemble into 26,000 scaffolds totaling 377 Mb in length. Within this preliminary assembly ~90% of reads can be placed and 50% of the genome is represented in scaffolds > 52 kb and contigs > 6.5 kb (scaffolds are contigs that linked together by sequences from 2 or more clones, which span a gap). Given that 2.4 million reads would represent only 2- 3x coverage, based on our original experimentally determined genome size for the genome of 570 Mb, the genome assembly is surprisingly good. We therefore believe that the coverage is in fact higher, and that the genome is smaller than we originally thought, close to about 389 Mb. Although it is possible that parts of the genome are refractory to cloning and capillary based sequencing, particularly as a whole genome amplification step was used to prepare the current sample, we feel that the most likely explanation is that the size determination by flow cytometry represents an overestimate. Thus, the relatively small size of the *Glossina* genome would mean that sequencing of additional five species should be an easily achievable goal. The genome size of the house fly has been determined to be 310 Mb (*37*). A preliminary analysis of the house fly transcriptome has recently been completed and this will greatly assist with the assembly and annotation of the house fly genome (Jeff Scott, personal communication).

4.2. de novo WGS assembly of *G. m. morsitans* genome

In addition to the sequence coverage we have described above, we are using 454 technology and aiming for a depth of coverage of 15 genome equivalents (a general requirement for assembling 454 data). In the case of *Glossina*, that could be as much as 8.5 Gb. Given that we are starting from draft coverage due our existing sequencing data (1.6 Gb, unassembled), we are producing additional sequence amounting to 6 Gb as a 50:50 mixture of shotgun and 3 kb paired end libraries for 15 runs, and will supplement this with ~600,000 reads (1 run) from a specialist 20 kb paired end library. So far approximately 2 Gb of new data have been generated, resulting in a marked improvement in the assembly statistics; using *Celera Assembler*, the assembly is 389 Mb and 50% of the data are now present in scaffolds > 308 kb and contigs > 11.2 kb. We plan to supplement these data with a single run from the Illumina platform but produced using 108 cycles of paired reads. This additional run will provide ~40Gb of data, approximately an additional 80-fold coverage, albeit in the form of short reads . The data will be assembled into longer *faux* reads for use in a hybrid assembly (e.g. using *Celera Assembler*) or will be used for iterative gap filling and error correction. We expect a draft assembly of the *G. m. morsitans* to be available this Fall.

We have also generated a BAC library for *G. m. morsitans (*funded through NIH/NHGRI (http://www.genome.gov/10001852). The desired average insert size of the library was around 120-140 kb, with an overall genome coverage of approximately 10x fold. The plates are available for public use (http://bacpac.chori.org). The finished sequences are available at ftp://ftp.sanger.ac.uk/pub/pathogens/Glossina/morsitans/. Annotation of the BAC clones have posed no unexpected problems. We have obtained 60,000 BAC-ends from this library and completed full sequence of six BAC clones. This paired BAC end sequence data will be important as sequence-tagged-connectors to assist in assembling scaffolds for the sequencing project.

4.3 Availability of Expressed-Sequence Tag (EST) libraries

Considerable progress has been made in generating ESTs and full-length gene sequences from *G. m. morsitans* as part of a gene-discovery effort (Table 3). This information has already provided insights into the key physiological processes using the *G. m. morsitans* system. Comparative analysis of the transcriptome of different tissues with their homologs in other diptera, *Drosophila*, *An. gambiae* and *Ae. aegypti*, has already shed light on the evolutionary processes that are conserved and that play a role in vector immunity as well as many new genes unique to the tsetse system (*26, 58*).

Similarly salivary gland transcriptome identified some conserved but many new genes that may play a role in trypanosome transmission process (*59*). Transcriptome analysis from reproductive tissue and intrauterine developmental stages identified 51 unique genes. Eleven of these unique proteins were homologous to uncharacterized putative proteins within the NR database suggesting the identification of novel genes associated with reproductive functions in other insects (hypothetical conserved). The analysis also yielded seven putative proteins without significant homology to sequences present in the public database (unknown genes). These proteins may represent unique functions associated with tsetse's viviparous reproductive cycle.

Table 3. EST collections currently available.

Tissue source of library	# ESTs	Contigs	Investigator/status
Normalized midgut from naïve and trypanosome infected *G. m. morsitans*	21,427	3,230	Sanger, completed (*60*)

Normalized salivary gland from *G. m. morsitans*	27,426	3,660	Sanger, completed (*59*)
Normalized fatbody from naïve and immune challenged *G. m. morsitans*	20,257	3,658	TIGR, completed (*61*)
Reproductive organ of *G. m. morsitans*	3,438	1,435	Sanger, completed (*56*)
Head - *G. m. morsitans*	3,000	1,590	Sanger, completed
Antennae - *G. m. morsitans*	5,000		Sanger, completed
Adult male–full length *G. m. morsitans cDNAs*	10,000		RIKEN completed (5'-3-seq)
Adult female–full length *G. m. morsitans cDNAs*	10,000		RIKEN completed (5'-3-seq)
Larval full length *G. m. morsitans* cDNAs	10,000		RIKEN completed (5'-3 seq)
Pupal full length *G. m. morsitans* cDNAs	9,984		RIKEN completed (5'-3'seq)
Head - *G p. palpalis*	3,000		Sanger, completed
Head – *G. tachinoides*	3,000		Sanger, completed
Antennae – *G. pallipides*	3,000		Sanger, completed
Full length adult cDNAs *G. p. gambiensis*	10,000		Genoscope completed
Fatbody and gut – *G. palpalis*	10,000		Genoscope completed

Transcriptome analysis of antennal libraries and homology search of our genomic data with known proteins has identified a number of odorant binding proteins (*62*). While we have observed high homology to *Drosophila* sequences, tsetse transcriptomes also exhibit many unique genes that will be important in understanding its vectorial capacity. Transcriptomics information from the new *Glossina* species we are requesting will allow comparative analysis between the human and animal disease vector species and provide information on their host specificity. The availability of microarrays will now enable us to perform global expression studies to understand host-parasite interactions that result in the resistance phenotypes. The EST dataset are also essential for training gene-finding software and subsequent annotation of the full genomes.

4.4 Choice of species

Building on the draft *G. m. morsitans* genome and transcriptome data, additional *Glossina* genomes and transcriptome information obtained from the proposed species stand to advance the field dramatically (Table 4). The criteria used for selecting species for genome sequencing were (1) availability of biological materials, (2) vector status of the new species proposed relative to HAT and AAT, (3) degree of evolutionary distance from the draft genome of *G. m. morsitans*, (4) evolutionary positioning within genus *Glossina* and (5) comparative genomics to identify shared sequences, such as regulatory elements, or species-specific sequences that could account for differences in their biology. Collectively, this information will not only give a strong boost to the tsetse/HAT field, it will also advance work conducted with model organisms such as *Drosophila*, given its close phylogenetic positioning in Diptera. The selection of two closely related Muscid flies will further identify genes/pathways relevant for *Glossina*. In addition, both the house fly and stable fly are important insects of medical and agricultural relevance. Trachoma transmitted by *Musca* alone causes six million cases of childhood blindness each year (World Health Organization 2004). Most recently, house flies have been shown to transmit life-threatening antibiotic-resistant bacteria, which are an ever increasing problem in hospitals and other health care facilities (37). Hence information on these additional taxa will be vital for the larger vector biology community in addition to serving as excellent outgroups for understanding *Glossina* biology.

Table 4. Proposed 5 Genomes Cluster for Genus *Glossina* and 2 related *Diptera*

Activities	Colony (Source)	Trypanosome parasites
1. *G. palpalis* WGS	Bratislava	Vector of *Tbg* in West Africa
G. fuscipes WGS	Bratislava	Vector of *Tbr* in East Africa
G. pallidipes WGS	Johnson Ouma, Kenya	Major vector of *Tbb*, minor vector of *Tbr*

G. brevipalpis WGS *G. austeni WGS* *2. Stomoxys calcitrans WGS* *Musca domestica WGS* *3.* Transcriptome data on WGS species 4. Low coverage WGS of key field populations, (10-12 populations total)	Otto Koekemoer, South Africa Atway Masangi, Tanga Mike Lehane, UK Jeff Scott, Ithaca, USA	Vector of AAT, (ancestral species) Vector of AAT, (phylogeny controversial)

The species targeted for sequencing belong to the 3 species complexes identified for genus *Glossina*. We estimate that they span about 100 MY, although it is only a best estimate at this time. Of the 5 species in the Palpalis complex, *G. palpalis* in West Africa and *G. fuscipes* in Central and East Africa are the type members. Recently *G. fuscipes* has been involved in transmitting >90% of HAT if you overlay predicted fly distributions with incidence of sleeping sickness cases. Therefore inclusion of these species as important disease vectors is a top priority.

Glossina pallidipes is the major animal vector in East Africa with a wide continental distribution. Where there is HAT transmission, *G. pallidipes* also contributes to epidemics, especially through the animal reservoirs *T. b. rhod*esiense maintains. *G. pallidipes* is in the *morsitans* complex, same as the draft *G. m. morsitans* genome. Control of *G. pallidipes* relies on vector control through traps and targets for which population genetics information is a high priority. Hence genomic mining for microsatellites would be a high priority. *Fusca* complex flies are thought to be the ancestral species in genus *Glossina* and are also widespread vector of animal disease. Sequencing of *G. brevipalpis* will enhance our understanding of tsetse evolutionary biology. The phylogenetic positioning of *G. austeni* has been difficult to assign with respect to Palpalis and Morsitans complexes. It is an important vector with wide distribution all along the eastern side of Africa from Somalia to South Africa. In certain areas *G. austeni* has been found to be a more important vector of animal trypanosomes than *G. morsitans*, *G. pallidipes*, or *G. brevipalpis*. Although *G. austeni* was exterminated from the island of Zanzibar (22), it remains a major problem in most of its geographic range.

Stomoxys calcitrans (Family Muscadiae) is an excellent genome to complete from a comparative genomic perspective among taxa in higher Diptera. It is a very important ectoparasitic pest in stockyards in the USA. Also, despite being sympatric with tsetse and sharing the same host animals for blood feeding, it does not permit completion of the trypanosome lifecycle, but can only act as a mechanical vector for *T. vivax* only. Before undertaking the sequencing effort, we will analyze the genome size of *Stomoxys* to ensure that it is reasonably small.

Musca domestica (Family Muscadiae) is another pest species closely related to *Glossina*. Unlike *Stomoxys*, *Musca* does not feed on vertebrate blood, but acts a mechanical vector of important pathogenic microbes. Given its close positioning with *Stomoxys* and *Glossina,* genomic information as well as transcriptomics will help dissect the mechanistic basis of tsetse's vectorial capacity. It has a genome size of 310 Mb (*37*). Single lines have been reared for this species so that biological material is trivial to obtain and availability of genetic transformation methods will enable further functional studies.

Colonies necessary to obtain the biological material for WGS of the proposed species are available (sources indicated in Table 1). These colonies have been developed from few individuals and maintained as closed lines for many years. Given the low reproductive biology, we expect these to be highly inbred. For *Musca* and *Stomoxys*, single lines have been developed, which will be used for WSG analysis. The *G. m morsitans* WGS project involved sequencing from 3 individuals obtained from one mother. We will similarly prepare the DNA necessary for

sequencing for the species proposed here. Large colonies exist for all of the *Glossina* species we are targeting, hence these will be used for transcriptome analysis.

4.5 Suggested sequencing strategy and priorities

We propose 20-30x coverage of each genome, sufficient to allow generation of a high quality assembly. Sequencing all of the 5 *Glossina* species and the 2 related Dipterans as outgroup taxa will be necessary for reaching our proposal goals.

4.6 Transcriptome Sequencing

Transcriptome sequencing is essential for accurate annotation of the proposed genomes as well as for promoting functional genomics studies in the near future. We propose to sequence the transcriptome from each of our proposed species from normalized cDNA pools comprised of multiple developmental stages and trypanosome infected and uninfected tissues. The transcriptome analysis will be performed by RNA-Seq using whole male, whole female, whole parasite infected male, whole parasite infected female, normal and infected salivary glands, heads, female and male reproductive tissues and immature developmental stages (larva and pupae). We will perform the sequencing from proposed species as part of high priority studies early in the proposal. This will make functional studies feasible before the eventual completion of the genome sequences.

4.7 SNP and microsatellite discovery

Highly differentiated tsetse field populations have been identified in Uganda in the two disease territories that exhibit *T. b. gambiense* and *T. b. rhodesiense* infections, respectively (*50*). These populations now provide a unique opportunity to undertake a low coverage WGS projects to enable SNP based genomic scan analysis that can help understand tsetse's vector competence traits. For this we will analyze parasite infected and resistant individuals from a small number of populations (total of 10-12 populations) where the two forms of disease are prevalent to look for variations within a population and between the two belts. This analysis has the potential to identify loci associated with parasite transmission traits. These populations have already been genetically identified and biological materials to begin sequencing analyses are available. Similarly populations have also been characterized in West Africa with *G. palpalis* with differential transmission dynamics and host seeking behavior, which are ripe for genomic scans (*51*). When applied to tsetse control, these studies have shown that in some areas tsetse populations are genetically isolated, indicating that eradication campaigns may be feasible and sustainable in such areas. Availability of high quality genomic markers will help estimate parameters of key importance, such as effective population sizes for post control monitoring. In addition, SNPs can promote genetic association studies to identify target genes involved in vectorial competence, and to ultimately decipher the paradox of how tsetse of the *palpalis* group constitute the most dangerous vectors of HAT yet their field infection rates with *T. b. gambiense* are always lower than 1% within endemic areas. SNP analysis will also be instrumental in following the emergence of insecticide resistance in field populations given that control efforts are increasingly relying on the use of pesticides.

5. Management of the project – Bioinformatics and resource sharing

To date, a central site for the dissemination of knowledge concerning the *Glossina* spp projects has been maintained at SANBI (www.iggi.sanbi.ac.za). Processed EST data were submitted to the research community through the GeneDB and VectorBase. All sequences have been deposited in the appropriate database at GenBank. The first-pass annotation of new genome sequences, assembly and gene identification will be provided by the Sequencing Center assigned to this project. Finer annotation of the genomes will be performed in collaboration with VectorBase.

6. International *Glossina* Genome Initiative (IGGI) Community

The United Nation's Tropical Disease Research programme (TDR) brought together tsetse researchers from more than a dozen sleeping-sickness labs and genome centers (TIGR/USA, Sanger Institute/UK, RIKEN GCG/Japan, GENOSCOPE/France and South African Bioinformatics Institute, SANBI/South Africa) in an effort to review the prospects of genomics activities for the tsetse fly (*63*). The first meeting was held on 19–20 January in Geneva 2004. To date six additional meetings have been held in WHO/TDR offices in Geneva (February 4, 2005), at TIGR (December 2005), at Sanger Institute (December 2006), at SANBI, South Africa (November 2007), in Mombasa in association with the Annual East African Tsetse Network (EANETT) meeting (November 2008) and recently in Wash DC in association with the annual ASTMH meeting (Nov 2009). In 2007 IGGI organized a community transcriptome annotation jamboree at SANBI. Wide support for the jamboree (more than 40 investigators attended) indicated the positive effect that the International *Glossina* Genome Initiative (IGGI) is already having in attracting people into the area. IGGI also organized a number of Bioinformatics workshops to build capacity for tsetse genomics. To date three workshops have been organized at SANBI where over 30 African scientists were trained in the use of genomics data for functional analysis. Recently IGGI initiated a Functional Genomics Network where 4-5 promising junior African scientists will receive short-term training in functional genomics laboratories in order to establish long-term collaborations. IGGI members have a strong commitment to advancing the *Glossina* genomics and as an International Partnership work towards moving such proposals forward including this white paper application.

*Members of International *Glossina* Genomics Initiative (IGGI)

Serap Aksoy, Yale University School of Medicine, New Haven, USA
Adalgisa Cacconne, Yale University, New Haven, USA
Rita Rio, West Virginia University, Morgantown, WV, USA
Matt Berriman, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK
Neil Hall, University of Liverpool, UK
Masahira Hattori, University of Tokyo, Japan
Noboru Inoue, Obihiro University, Japan
Chihiro Sugimoto, Hokkaido University, Japan
Alan Christopher, SANBI, South Africa
Win Hide, Harvard University and SANBI, USA
Mike Lehane, LSTM, UK,
Joseph Ndung'u, FIND, Switzerland,
Philippe Solano, IRD/CIRDES, Burkina Faso
Patrick Wincker, Genoscope, France
Jan van den Abbeele, ITMA, Belgium
Loyce Okedi, National Livestock Research Institute, Uganda
Johnson Ouma, Trypanosomiasis Research Center, Kenya
Todd Taylor, RIKEN ACI, Japan
Daniel Masiga, ICIPE, Kenya
Atway Masangi, Tsetse Trypanosomiasis Research Institute, Tanga, Tanzania

Michael Gaunt, London School of Tropical Medicine, UK
Anna Malacrida, University of Pavia, Italy
Kostas Bourtzis, University of Ioannina, Greece
Adyl M.M.Abd-Alla, International Atomic Energy Agency, Austria
Peter Takac, Slovak Academy of Science, Bratislava, Slovakia
Jeff Scott, Cornell University, Ithaca, NY
Muscid and Stomoxys community members
Drosophila community members

References:

1. F. T. Petersen, R. Meier, S. N. Kutty, B. M. Wiegmann, *Mol Phylogenet Evol* **45**, 111 (Oct, 2007).
2. S. C. Welburn, E. M. Fevre, P. G. Coleman, M. Odiit, I. Maudlin, *Trends in parasitology* **17**, 19 (2001).
3. E. M. Fevre, M. Odiit, P. G. Coleman, M. E. Woolhouse, S. C. Welburn, *BMC Public Health* **8**, 96 (2008).
4. M. Lyons, *The Colonial Disease. A social History of Sleeping Sickness in Northern Zaire, 1900-1940*. (Cambridge University Press, Cambridge UK, 1992).
5. M. Odiit *et al.*, *Trop Med Int Health* **10**, 840 (Sep, 2005).
6. P. P. Simarro, J. Jannin, P. Cattand, *PLoS Med* **5**, e55 (Feb, 2008).
7. C. D. Steelman, *Annual Review of Entomology* **21**, 155 (1976).
8. A. M. Jordan, *Longman, London*, (1986).
9. L. T. Budd, *Tsetse and Trypanosomosis Research and Development since 1980: an Economic Analysis*. D. f. I. Development, Ed., Livestock Production Programme, Animal Health Programme, Natural Resources Systems Programme (UK, 1999), vol. 2.
10. J. P. Kabayo, *Trends Parasitol* **18**, 473 (Nov, 2002).
11. D. Butler, *Nature* **424**, 10 (Jul 3, 2003).
12. A. J. Nok, *Parasitol Res* **90**, 71 (May, 2003).
13. M. G. Etchegorry, J. P. Helenport, B. Pecoul, J. Jannin, D. Legros, *Trop Med Int Health* **6**, 957 (Nov, 2001).
14. B. M. Anene, D. N. Onah, Y. Nawa, *Veterinary parasitology* **96**, 83 (2001).
15. S. Geerts, P. H. Holmes, M. C. Eisler, O. Diall, *Trends in parasitology* **17**, 25 (2001).
16. M. Berriman *et al.*, *Science* **309**, 416 (Jul 15, 2005).
17. N. M. El-Sayed *et al.*, *Nucleic acids research* **31**, 4856 (Aug 15, 2003).
18. N. M. El-Sayed *et al.*, *Science* **309**, 404 (Jul 15, 2005).
19. P. M. Kgori, S. Modo, S. J. Torr, *Acta Trop* **99**, 184 (Oct, 2006).
20. L. L. Joja, U. A. Okoli, *Am J Public Health* **91**, 1583 (2001).
21. J. Lancien, *Ann Soc Belg Med Trop* **71 Suppl 1**, 35 (1991).
22. M. J. Vreysen *et al.*, *J. Econ. Entomology* **93**, 123 (2000).
23. E. S. Krafsur, *Infect Genet Evol* **9**, 124 (Jan, 2009).
24. X. Chen, S. Li, S. Aksoy, *Journal of Molecular Evolution* **48**, 49 (1999).
25. D. Rogers, T. Robinson, in *The Trypanosomiases,* I. Maudlin, P. Holmes, M. Miles, Eds. (CAB International;, Oxford, 2004), pp. 139–179.
26. S. Aksoy, W. C. Gibson, M. J. Lehane, *Advances in parasitology* **53**, 1 (2003).

27. L. R. Haines, S. M. Lehane, T. W. Pearson, M. J. Lehane, *PLoS pathogens* **6**, e1000793 (2010).
28. J. Wang *et al.*, *Insect Mol Biol* **17**, 293 (Jun, 2008).
29. D. Nayduch, S. Aksoy, *J Med Entomol* **44**, 660 (Jul, 2007).
30. C. Hu, S. Aksoy, *Mol Microbiol* **60**, 1194 (Jun, 2006).
31. Z. Hao *et al.*, *Proc Natl Acad Sci U S A* **98**, 12648 (Oct 23, 2001).
32. M. B. Wang, S. V. Wesley, E. J. Finnegan, N. A. Smith, P. M. Waterhouse, *RNA.* **7**, 16 (2001).
33. R. H. Gooding, E. S. Krafsur, *Annu Rev Entomol* **50**, 101 (2005).
34. P. Solano, S. Ravel, T. de Meeuˆs, *Trends in Parasitology* **in press**, (2010).
35. S. Aksoy, I. Maudlin, C. Dale, A. Robinson, S. O'Neill, *Trends in parasitology* **17**, 29 (2001).
36. M. J. Lehane, D. Wu, S. M. Lehane, *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11502 (1997).
37. J. G. Scott, N. Liu, M. Kristensen, A. G. Clark, *J Med Entomol* **46**, 175 (Mar, 2009).
38. D. M. Soderlund, *Pest Manag Sci* **64**, 610 (Jun, 2008).
39. L. R. Haines *et al.*, *Insect biochemistry and molecular biology* **35**, 413 (May, 2005).
40. L. U. Abubakar, W. D. Bulimo, F. J. Mulaa, E. O. Osir, *Insect biochemistry and molecular biology* **36**, 344 (Apr, 2006).
41. S. Aksoy, I. Maudlin, C. Dale, A. S. Robinson, S. L. O'Neill, *Trends Parasitol* **17**, 29 (Jan, 2001).
42. R. V. Rio, Y. Hu, S. Aksoy, *Trends Microbiol* **12**, 325 (2004).
43. G. M. Attardo *et al.*, *J Insect Physiol* **54**, 1236 (Aug, 2008).
44. S. C. Welburn, I. Maudlin, *Annals of tropical medicine and parasitology* **86**, 529 (Oct, 1992).
45. J. Wang, Y. Wu, G. Yang, S. Aksoy, *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12133 (Jul 21, 2009).
46. I. Maudlin, S. C. welburn, P. Milligan, *Acta tropica* **48**, 9 (1991).
47. G. Caljon *et al.*, *International journal for parasitology* **36**, 1025 (Aug, 2006).
48. G. Caljon, K. De Ridder, P. De Baetselier, M. Coosemans, J. Van Den Abbeele, *PloS one* **5**, e9671 (2010).
49. E. S. Krafsur, *Trends in parasitology* **19**, 162 (Apr, 2003).
50. J. S. Beadell *et al.*, *PLoS neglected tropical diseases* **4**, e636 (2010).
51. P. Solano *et al.*, *PLoS Negl Trop Dis* **3**, e392 (2009).
52. M. Kagbadouno *et al.*, *Parasite* **16**, 29 (Mar, 2009).
53. M. O. Omolo *et al.*, *PLoS Negl Trop Dis* **3**, e435 (2009).
54. J. B. Rayaisse, F. Courtin, M. Akoundjin, J. Cesar, P. Solano, *Parasite* **16**, 21 (Mar, 2009).
55. G. Yang, G. M. Attardo, C. Lohs, S. Aksoy, *Insect molecular biology*, (Jan 28, 2010).
56. G. M. Attardo, J. M. Ribeiro, Y. Wu, M. Berriman, S. Aksoy, *BMC genomics* **11**, 160 (2010).
57. S. Aksoy *et al.*, *Trends in parasitology* **21**, 107 (Mar, 2005).
58. M. J. Lehane, S. Aksoy, E. Levashina, *Trends in parasitology* **20**, 433 (Oct, 2004).
59. J. Alves-Silva *et al.*, *BMC genomics* **11**, 213 (Mar 30, 2010).
60. M. J. Lehane *et al.*, *Genome Biol* **4**, R63 (2003).
61. G. M. Attardo *et al.*, *Insect Mol Biol* **15**, 411 (Aug, 2006).
62. R. Liu *et al.*, *Cell Mol Life Sci* **67**, 919 (Mar, 2010).
63. D. Butler, *Nature* **427**, 384 (Jan 29, 2004).