Division of Collaboration and Education, Research Center for Zoonosis Control, Graduate School of Veterinary Medicine, Hokkaido University, Japan

Application of metagenomic approaches in the comprehensive detection and characterization of trypanosomes, microbiome and blood meal sources in tsetse flies

(メタゲノム的手法を用いたツェツェバエ内トリパノソーマ原虫、細菌叢、および吸血源の包括的解析)

Alex Samuel Kiarie Gaithuma

A Dissertation submitted at Hokkaido University in fulfilment of the requirements for doctor of philosophy (PhD) in Veterinary Medicine

Hokkaido University

Sapporo

Declaration

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

Alex Samuel Kiarie Gaithuma

Supervised by;

Professor Emeritus Chihiro Sugimoto

Hokkaido University, Graduate School of Veterinary Medicine, Research Center for Zoonosis Control, Division of Collaboration and Education, Japan

Associate Professor Junya Yamagishi

Hokkaido University, Graduate School of Veterinary Medicine, Research Center for Zoonosis Control, Division of Collaboration and Education, Japan

Dedication

I dedicate this work to my dear mum and dad for their unconditional love, support and prayers.

Table of contents

List of publications related to the dissertation

The contents of this dissertation are based on research conducted during the period October 2015 - September 2019 and part of the results published in the following publication.

Gaithuma AK, Yamagishi J, Martinelli A, Hayashida K, Kawai N, Marsela M, and Sugimoto C. A single test approach for accurate and sensitive detection and taxonomic characterization of Trypanosomes by comprehensive analysis of internal transcribed spacer 1 amplicons. *PLoS Negl Trop Dis*. 2019; 13: e0006842. Available: https://doi.org/10.1371/journal.pntd.0006842.

List of tables

List of figures

List of abbreviations

AAT	African animal trypanosomiasis
AITS	Amplification of internal transcribed spacer
AMPtk	Amplicon tool kit
ASV	Amplicon sequence variant
BSF	Bloodstream form
CB	Cattle blood
CSF	Cerebral spinal fluid
DADA	Divisive amplicon denoising algorithm
DNA	Deoxyribonucleic acid
FAO	Food and Agriculture Organization
HAT	Human African trypanosomiasis
iTOL	Interactive tree of life
ITS1	Internal transcribed spacer 1
LCBD	Local contribution to biodiversity
NTC	Negative template control
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
rRNA	ribosomal RNA
Tb	*Trypanosoma brucei*
Tbg	*Trypanosoma brucei gambiense*
Tbr	*Trypanosoma brucei rhodesiense*
Tc	*Trypanosoma congolense*
Tg	*Trypanosoma godfreyi*
Tv	*Trypanosoma vivax*
WBC	White blood cell
WHO	World Health Organization

1 General introduction

1.1 African trypanosomiasis

African trypanosomiasis is a neglected tropical disease endemic in rural areas of sub-Saharan Africa, that causes disease in humans (Human African trypanosomiasis or HAT) and animals (Animal African trypanosomiasis or AAT). Both pose a major health and economic burden in sub-Saharan Africa (WHO, 2015). African trypanosomiasis is caused by protozoan parasites of the genus *Trypanosoma*, transmitted by the bite of tsetse flies (genus *Glossina*). Trypanosomes belong to the order "kinetoplastida" due to the presence of a kinetoplast (elongated mitochondria lying close to the nucleus). Based on the mode of transmission by the insect vector, the genus *Trypanosoma* is divided into two groups: stercoraria and salivaria (Figure 1). The development of stercoraria parasites takes place in the intestinal tract of the insect vector and the vertebrate host is infected via feces. *Trypanosoma cruzi,* the causative agent of Chagas' disease, is an example of the stercoraria group. Salivarian parasites colonize the stomach of the insect vector. They never pass to the intestinal tract but they migrate towards the salivary gland and mature to vertebrate infectious forms. Two forms of the human disease exist (1) Rhodesian HAT caused by *T. b. rhodesiense*, an acute disease found in East and Southern Africa and transmitted by *G. morsitans,* and (2) Gambian HAT caused by *T. b. gambiense,* a chronic disease distributed in West Africa and is mainly transmitted by *G. pallidipes* (Médecins, 2004; WHO, 2013; Grébaut *et al.*, 2016). Uganda is the only country where both forms of the disease occur with the potential for overlapping infections (Berrang-Ford *et al.*, 2005). African animal trypanosomiasis refers to infections in many different animal hosts (bovine, buffalo, goat, sheep, camel, horse, pig and wild animals) and is caused by various trypanosome species including *T. brucei*, *T. congolense*, and *T. vivax*, *T. evansi* and *T. equiperdum* (Büscher, 2005). It is often treated as a single disease, but it is caused by

different trypanosome species and often mixed infections are involved. Additionally, different trypanosome species present varying clinical presentations. *Trypanosoma brucei*, *T. congolense*, and *T. vivax* cause a wasting disease called "nagana" that is responsible for most of the livestock losses in sub-Saharan Africa. Two species are distributed outside the African continent since they can be transmitted mechanically by biting insects mostly horseflies (*Tabanus* spp.) and stable flies (*Stomoxys* spp.). (1) *Trypanosoma evansi* which causes a disease called "surra" in Central and South America, the Middle East, and Asia, and (2) *T. vivax* (also transmitted by tsetse flies) is found in Central and South America. *Trypanosoma evansi* infection is a disease of economic importance since it affects the health of buffalo, cattle, and swine (Wuyts, Chokesajjawatee and Panyim, 1994)

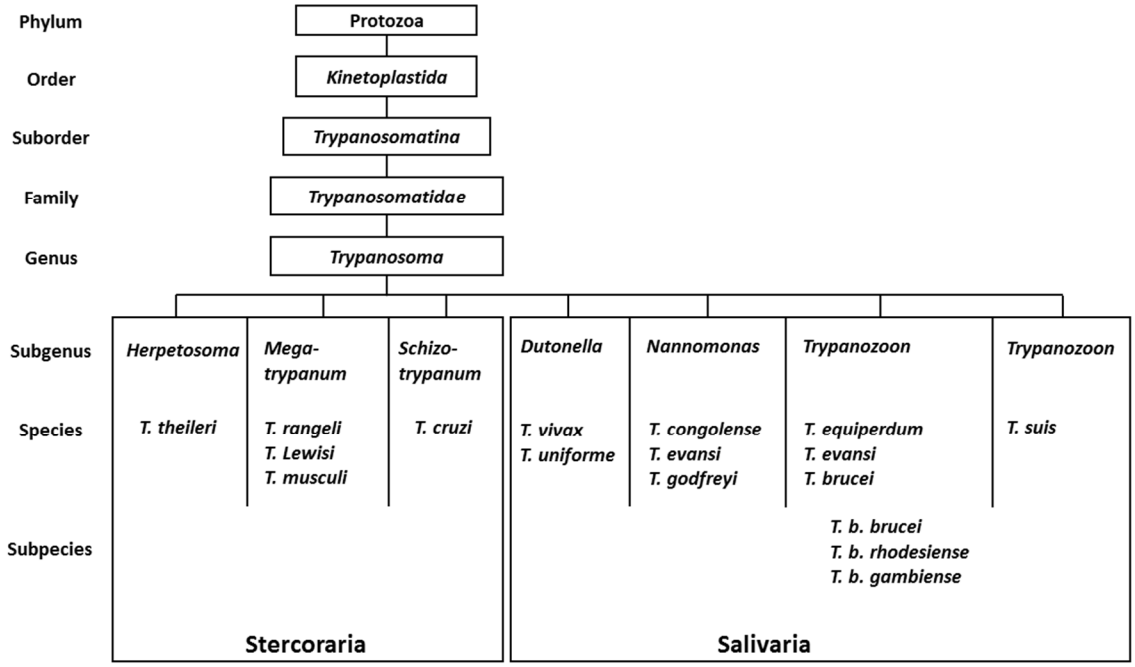

Figure 1: Table showing the classification of trypanosomes.

1.2 Life cycle of the African trypanosomes

Infection in the mammalian host begins when the metacyclic trypanosomes (infective stage), are injected intradermally by the tsetse fly (Figure 2). They transform rapidly into the bloodstream form (BSF), and divide by binary fission in the interstitial spaces in the skin. The buildup of cell debris and metabolic wastes sometimes leads to the formation of a "chancre" at the biting site. The parasites then undergo morphological changes and exhibit antigen variation of their surface coat, which harbor the Variant Specific Surface Glycoproteins (VSG). Trypanosomes are able to switch between about 107 VSG variants as a mechanism of evasion of the host's immune system. (Vickerman, 1965; Barry and McCulloch, 2001). The BSFs are heterogeneous consisting of the proliferative slender forms during the ascending phase of parasitemia and the non-proliferative stumpy forms at the peak of parasitemia (Vickerman, 1985). The stumpy forms ultimately establish themselves and form a parasite population that is preadapted for the transition to the procyclic forms (PCF), which occupy and proliferate in the midgut of the tsetse fly. During feeding, the tsetse fly takes up parasites with the blood meal from an infected host. The parasites undergo metabolic changes in the midgut of the fly where they lose their surface coat and transform into the proliferative PCFs which express their own surface proteins called Procyclic Acidic Repetitive Proteins (PARPs, or procyclins). To enable transmission from the tsetse fly to the mammalian host, the PCFs undergo two stages of differentiation in the fly: first, they establish in the fly midgut and then mature in the mouthparts or the salivary gland.

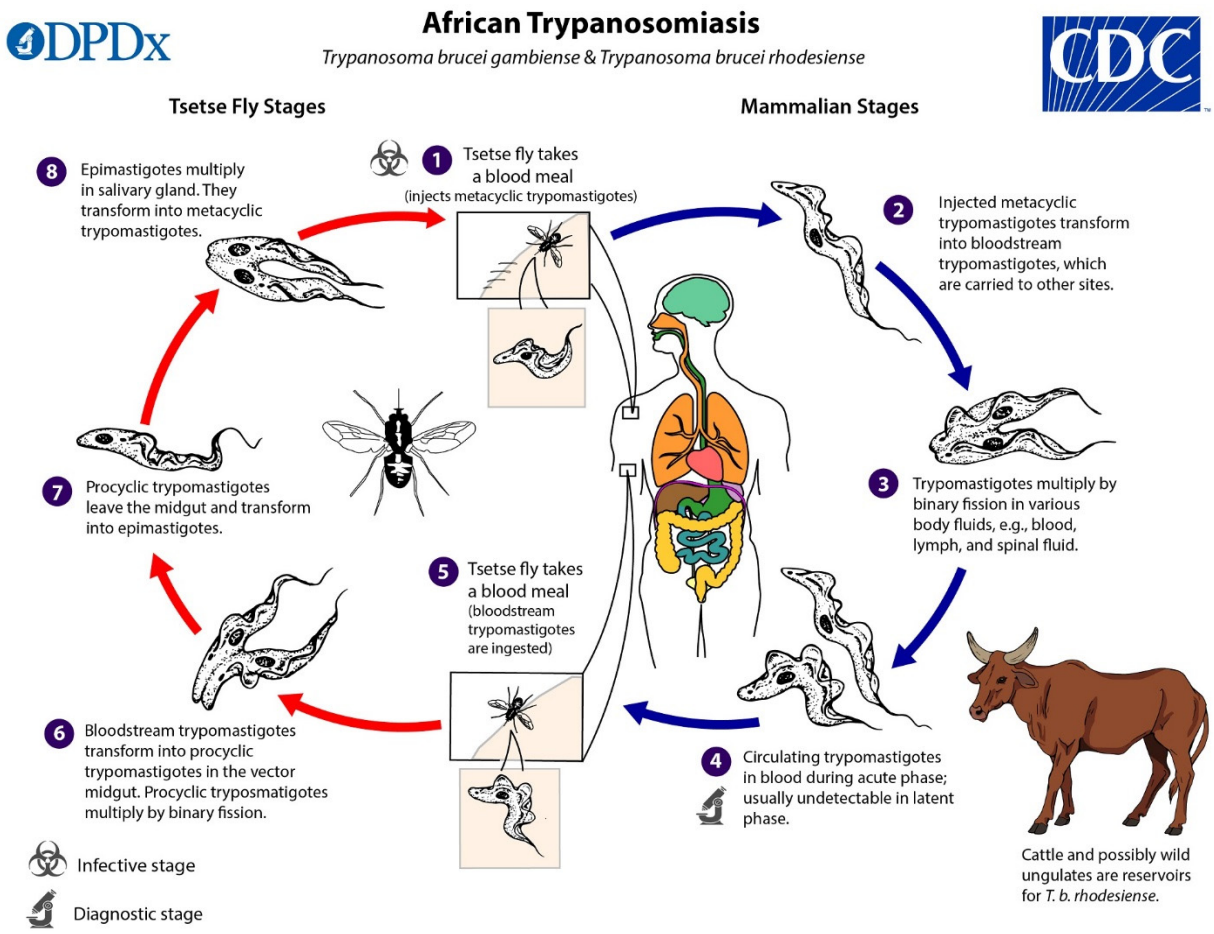

Figure 2: The life cycle of *Trypanosoma brucei* in a man and in the tsetse fly vector. (Source: https://www.cdc.gov/dpdx/trypanosomiasisafrican/index.html, accessed on 29/03/2019).

In each host, trypanosomes undergo many life cycle stages involving forms with discrete morphologies, patterns of gene expression, and proliferation status. In the human disease**,** the late (meningoencephalitic) stage of the infection coincides with the invasion of the central nervous system (CNS) by BSFs and is associated with psychiatric, motor and sensory disorders, along with sleep abnormalities (Kennedy, 2004; Sternberg, 2004; Brun and Blum, 2012). If untreated, late-stage patients progress to a final stage involving seizures, somnolence, coma, and death (Kennedy, 2004). In AAT, anemia is the most prominent disease-related feature, but trypanotolerance (the capacity of an animal to control severe anemia development) in cattle has been reported (Naessens, 2006). Diseased animals besides suffering from anemia, show degenerative, inflammatory and necrotic lesions in different organs due to an intensive invasion of lymphocytes, macrophages and plasma cells leading to characteristic wasting syndrome.

1.3 Distribution of African Trypanosomiasis

Human African trypanosomiasis was thought to have been largely conquered by the 1960s but re-emerged as a serious public health problem. The current distribution of the East African HAT is mainly focused in Uganda, Malawi, Zambia, and Tanzania while the West African HAT is focused mainly in the Congo, among other West African countries (Figure 3). Aggressive control of the disease and its vectors by various agencies in the last decade have reduced HAT cases dramatically to a prevalence of 50-70,000 cases (WHO, 2006). However, the disease burden is thought to be much more than it is due to underreporting. Lack of active surveillance, poor infrastructure, limited health facilities and low sensitivity of current diagnostic tools are leading causes of underreporting (Odiit *et al*., 2004). The distribution of African animal trypanosomiasis is poorly understood and has been mainly estimated by the distribution of tsetse flies and prevalence data (Muhanguzi *et al*., 2014). However, the mechanical transmission of *T. vivax* by biting flies indicates that tsetse distribution patterns cannot accurately predict the distribution of *T.*

vivax AAT. There is, therefore, need to generate and use up-to-date animal trypanosomiasis prevalence data.

If incidence alone is considered, the disease may appear as a minor problem compared with other diseases, but if disability-adjusted life years (Murray, 1994) is considered, and because of its severity (untreated HAT results in 100% mortality), then the social and economic impact of trypanosomiasis ranks third after malaria and schistosomiasis in affected parts of sub-Saharan Africa (Cattand *et al.,* 2001). Furthermore, animal reservoirs play an important role in the epidemiology of the East African HAT since *T. b. rhodesiense* infection can occur asymptomatically in animals (Sternberg, 2004).

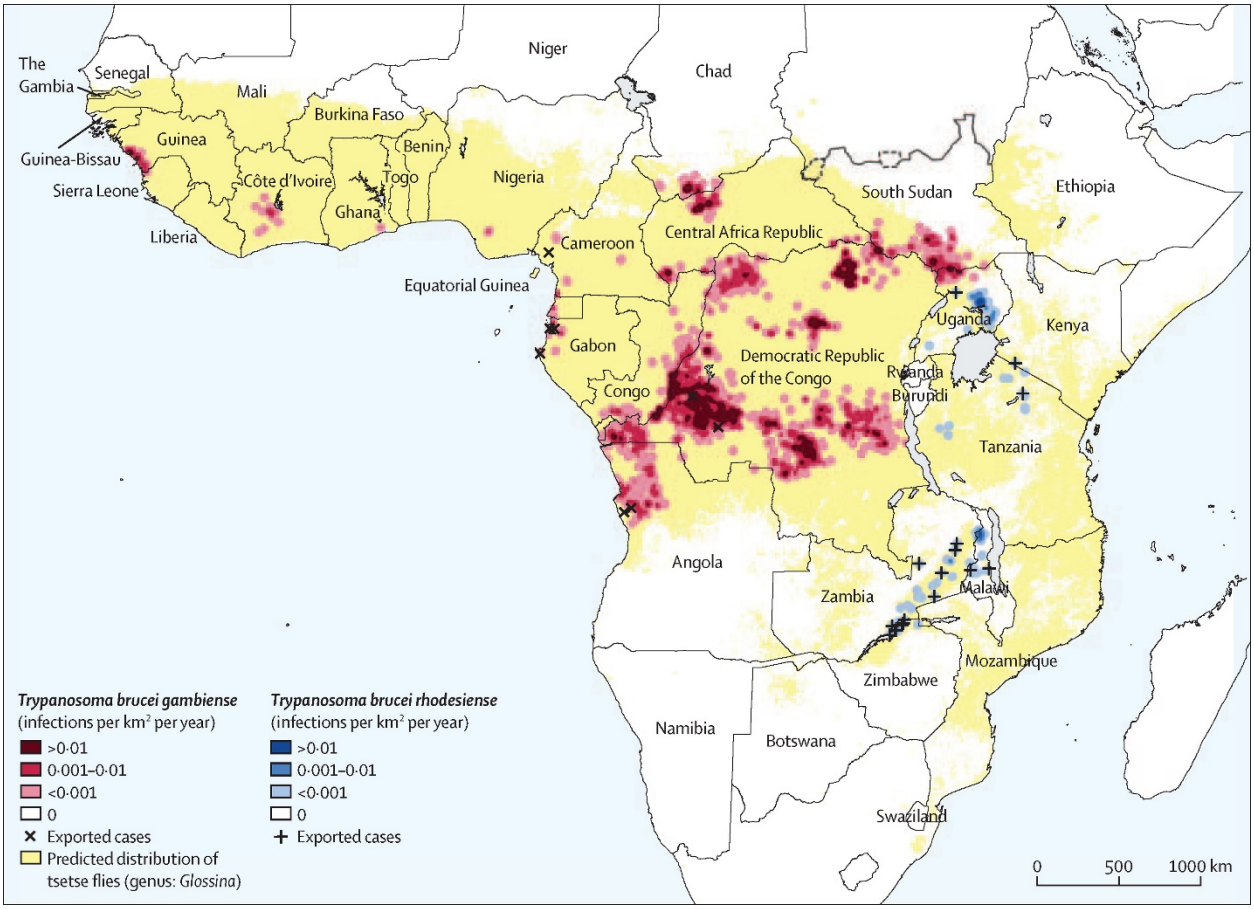

Figure 3: Distribution of HAT in sub-Saharan Africa. Infection data reported from 2010–14.

Source: (Büscher *et al*., 2017).

1.4 Control and diagnosis of Trypanosomiasis

Control of HAT depends on chemotherapy and vector control strategies. For many years, a limited number of drugs have been used but recently, new formulations have been developed thanks to increasing interest from the scientific community. The current treatment option for sleeping sickness, nifurtimox-eflornithine combination therapy, while effective, is burdensome for patients and health workers due to logistical challenges of hospitalization. Of note is the development and approval of an oral treatment regimen, fexinidazole, for the treatment of late-stage *T. b. gambiense*. The finding that the oral fexinidazole drug is effective and safe for the treatment of *T. b. gambiense* infection compared to nifurtimox-eflornithine combination therapy could be a key asset in the elimination of this fatal neglected disease (Mesu *et al.*, 2018). The choice of drug depends on the infecting parasite and the stage of the disease. In theory, a patient is considered cured when no trypanosomes are detected in the blood, lymph or cerebral spinal fluid (CSF) during this two year follow-up period and when white blood cell (WBC) counts stay or return to normal (WHO, 1998). For follow-up after treatment, patients' blood and CSF are re-examined on several occasions with longer follow-up periods of up to two years (Kennedy, 2008).

Current vector control interventions involve the use of insecticides which include ground spraying, sequential aerosol spraying technique, insecticide-treated targets or insecticide-treated animals, other-baited traps or screens, and the sterile insect technique. Some interventions that were conducted in the past such as bush clearing (tsetse habitat destruction) or elimination of wild animals (tsetse reservoir hosts) have been discarded for ecological and environmental concerns (WHO, 2017).

The World Health Organization (WHO) has an objective of the elimination of human African trypanosomiasis as a public health problem and the implementation of sustained surveillance in all disease-endemic countries. However, being a zoonosis, the total interruption of

transmission of East African HAT is considered not feasible and as with AAT, sustained reduction in disease incidence to a locally acceptable level is a more realistic target than elimination (Van den Bossche and Delespaux, 2011; Bouyer *et al.*, 2013). Sustaining control and surveillance includes active and passive case finding, diagnosis, treatment, follow-up, vector control and control of the animal reservoir. Parasite detection relies mainly on microscopic examination of blood smears. However, the Mini Anion Exchange Centrifugation Technique (mAECT) that involves separation of trypanosomes from the blood by anion-exchange chromatography concentrates trypanosomes which are normally in low numbers for microscopy to detect. This allows the detection of fewer than 30 trypanosomes/mL (Bonnet *et al.,* 2015). Another method of concentrating trypanosomes commonly used is Capillary Tube Centrifugation (CTC) that employs centrifugation of anticoagulated blood to concentrate trypanosomes. Antibody detection for West African HAT is done by Card-Agglutination Trypanosomiasis Test (CATT) for initial screening. It is a serological test that uses *T. b. gambiense* variable antigen type LiTat 1.3. However, false negatives have been reported for patients infected with strains of trypanosomes that do not express the LiTat 1.3 gene (Dukes *et al.*, 1992). For the East African HAT, no serological test for screening exists and the only method for diagnosis is by PCR detection targeting the serum-resistant antigen (SRA). Stage diagnosis in infected patients can only be done by examination of the CSF after lumbar puncture. However, the detection of trypanosomes in CSF by microscopy alone has limited sensitivity and has a poor reproducibility rate and concentration of trypanosomes is often carried out to increase sensitivity. The early diagnosis of HAT is crucial to avoid disease progression to the neurological stage, which requires potentially unsafe treatments. However, due to limited resources and inexperienced personnel, infected people are either misdiagnosed or die before diagnosis or treatment. Parasite and antibody detection is needed for adequate diagnosis in the field.

The diagnosis of trypanosomiasis in animals is mainly based on clinical suspicion only and diagnosis is not compulsory due to limited financial resources. The lack of encouragement for researchers to undertake validation of diagnostic tests for large scale implementation limits the diagnosis of AAT to epidemiology and surveillance purposes (Büscher, 2005). Additionally, the diagnosis of animals by suspicion leads to underreporting of incidence cases. Diagnostic tools are key elements for trypanosomiasis control since they provide the epidemiological data important for intervention strategies. Most of the available data has relied mostly on microscopy which has been shown to be less sensitive and accurate than DNA-based detection and identification methods (Njiru *et al.*, 2005). The use of DNA-based methods has the advantage of being more sensitive and able to identify trypanosomes to the subspecies level and to detect mixed infections. Development of new molecular tools that are field friendly and have the capacity for product development are needed since, in many settings, diagnosis continues to rely on insensitive and unsatisfactory parasitological or sero-diagnostic techniques. Recent advances in nucleic acid sequencing technologies present reliable and accurate diagnostic tools for diagnosis of trypanosomiasis owing to their increasing affordability and improved field applicability.

The major challenge in control of the zoonotic East African HAT and AAT is the wide variety of animal species involved, thereby complicating effective surveillance, prevention, and control efforts. Identification of reservoir hosts is labor-intensive due to several requirements such as the capture of potential wildlife hosts and experimental infections. Serological techniques were earlier used to identify the source of vertebrate blood (Stephen, 1999), but had a major setback in distinguishing closely related species. Recent developments in molecular biology have allowed a significant improvement in the efficacy and reliability of vertebrate host identification (Kent, 2009). Application of new tools for identification of tsetse fly vertebrate hosts will be important

for identifying the tsetse fly host preferences and predicting the trypanosome animal reservoirs in an endemic area.

The control African trypanosomiasis has been achieved largely by controlling the size of tsetse populations but recently, new disease control strategies that inhibit pathogen maturation within arthropod disease vectors are the focus of research. Enteric microbiota is increasingly being studied for use as 'probiotic' bacteria that could alter the physiology of the vector's gut to make the environment inhospitable to pathogens. Tsetse flies house a taxonomically diverse enteric microbiota that includes endosymbiotic *Wigglesworthia, Wolbachia* and *Sodalis* (Geiger *et al.,* as well other bacteria obtained from the environment. These microbes are intimately associated with many important aspects of tsetse fly biology. Thus, increasing the fundamental knowledge regarding how tsetse flies interact with this microbiota will allow the development of new and innovative control strategies aimed at reducing tsetse populations and/or tsetse vector competence.

This study aims to exploit targeted next-generation sequencing in the detection of trypanosome species, analysis of blood meal hosts and analysis of bacterial microbiome in wild-caught tsetse flies in order to provide useful data for prevention and control of trypanosomiasis and understanding tsetse fly biology and ecology. Diagnostic metagenomics has so far not been widely used in the detection of parasitic infections. The application of metagenomics is set to play a key role in future in the recovery of genomic-epidemiological data and determining the influence of parasites on the microbial ecology of the gut (Griffing *et al.*, 2011). The major challenge of metagenomics is laborious sample preparation and therefore new methods need to be developed to allow parasite detection in crude samples and/or without amplification. Recently, metagenomic applications which make use of next-generation sequencing (NGS) have been shown to be highly sensitive in the detection of low-frequency variants, have higher throughput with sample multiplexing and have a lower limit of detection of DNA. Since targeted approaches apply

selective amplification and sequencing of high-information-content regions of microbial genomes (without amplification of background DNA), they allow taxonomic identification at a significantly lower depth of sequencing and subsequently less complex computational analysis (Dekkera, 2018). Additionally, metagenomics applications are attractive because they are open-ended, assumption-free and a one-size-fits-all workflow that could be applied to any specimen to detect any kind of pathogen. The cost of sequencing continues to drop significantly and the availability of less laborious sample preparation methods offer a promising trend towards achieving field friendly tools. The development of point of care sequencing devices utilizing metagenomics is therefore feasible.

CHAPTER ONE

2 CHAPTER ONE: A new method for accurate and sensitive detection and taxonomic characterization of trypanosomes by comprehensive analysis of internal transcribed spacer 1 amplicons using next-generation sequencing

2.1 Summary

Development of better tools is required to monitor trypanosome genotypes circulating in both mammalian hosts and tsetse fly vectors in order to improve our knowledge on the epidemiological status of African trypanosomiasis. Such tools should enable researchers to determine the diversity of trypanosomes and to understand how environmental factors and ongoing control efforts affect trypanosome epidemiology in an area and ultimately parasite evolution. This study presents a single test approach for molecular detection of different trypanosome species and subspecies using newly designed primers to amplify the Internal Transcribed Spacer 1 (ITS1) region of ribosomal RNA genes, coupled to Illumina sequencing of the amplicons. The method's core outline follows the general outline of Illumina's widely used 16S bacterial metagenomic protocol which utilizes PCR and dual indexing to sequence amplicons from multiple samples in the same run. In this study, through the analysis of wild tsetse flies collected from Zambia and Zimbabwe, it was observed that conventional trypanosome species detection based on the band size in a gel is not always able to accurately distinguish between *T. vivax* and *T. godfreyi*. The newly developed method has increased specificity for the detection of trypanosomes at the species level (with the exception of the *Trypanozoon* subgenus) without the need for additional tests. It was also determined that *T. congolense* Kilifi subspecies was more closely related to *T. simiae* than to other *T. congolense* subspecies, agreeing with previous studies using satellite DNA and

18S RNA analysis. It was also observed that sequences matching *T. congolense* Tsavo (now classified as *T. simiae* Tsavo) clustered distinctly from other *T. simiae* Tsavo sequences. This results collectively suggest that the *Nannomonas* group is more divergent than currently thought and more studies are needed to clearly outline and understand their taxonomic classification. While the current taxonomic classification of trypanosomes does not list any subspecies for *T. godfreyi*, two distinct clusters were observed, indicating the presence of two distinct subspecies. This study presents a method that is simple but comprehensive in the identification of trypanosome species and subspecies using a single PCR test. This method is therefore recommended for the molecular detection, characterization, and classification of trypanosomes in field samples.

2.2 Introduction

Sleeping sickness cases have declined over the years with WHO estimating the number of actual cases to be below 20,000 reported in 2015 (WHO, 2018). The reduction of incidence cases is attributed to improved case detection, treatment and vector management (Simarro *et al.*, 2008). On the other hand, cases of African animal trypanosomiasis (AAT) are seldom reported and tabulated despite AAT being among the biggest constraints to sustainable livestock production in Africa. The parasites, *T. congolense* (Savannah) and *T. vivax*, are predominantly distributed in sub-Saharan Africa and are considered the most important animal trypanosomes (Morrison *et al.*, 2016) causing pathogenic infections (*Nagana*) in cattle, sheep, goats, pigs, horses, and dogs. *Trypanosoma brucei brucei* (and *T. b. rhodesiense*) is pathogenic to camels, horses, and dogs, and causes mild or no clinical disease in cattle, sheep, goats, and pigs (Van Den Bossche *et al.*, 2005; Gibson, 2007). Experimental infections with *T. godfreyi* in pigs shows a chronic, occasionally fatal disease (Adams *et al*., 2010; Auty *et al.*, 2012) while *T. simiae* causes a fatal disease in pigs and mild disease in sheep and goats. *Trypanosoma evansi* and *T. equiperdum* do not undergo a tsetse fly life cycle but are mechanically and sexually transmitted respectively while *T. vivax* can be transmitted mechanically by other biting insects (Gardiner and Mahmoud, 2012). The latter three species are therefore not limited to the distribution of tsetse flies and can be found outside sub-Saharan Africa (Lun and Desser, 1995; Lai *et al.*, 2008). *Trypanosoma evansi* infects camels, dromedaries, horses, and other equines as well as in a wide range of animals causing *Surra* disease, while *T. equiperdum* causes dourine in equines (Desquesnes *et al.*, 2013). Trypanosomes are maintained in both wild and domestic animals as reservoirs and their distribution is diverse and dynamic, making their control difficult.

The identification of trypanosome species and subspecies is important for instance, to interrogate their effect and contribution to livestock disease and elucidate assumed "strain"

differences in drug response among other factors. Morphological methods, which have been a gold standard for classifying trypanosomes for many years, have limited ability to distinguish between species because some trypanosomes share morphological features, are found in the same developmental sites, and mixed and immature infections also exist. A common marker to identify trypanosome species in hosts and vectors is the ribosomal RNA sequence region harboring internal transcribed spacer sequences located between the 18S and 5.8S ribosomal subunit genes which are about 100–200 copies (Desquesnes and Dávila, 2002). Amplification of the ITS1 region by PCR has been used widely to identify trypanosome species based on the amplicon size in a gel (Njiru *et al.*, 2005). However, analysis of ITS1 amplicons on agarose gels fails to distinguish between some species and/or genotypes such as *T. simiae* and *T. simiae* Tsavo. Another limitation with ITS1 based PCR is the specificity of detection, showing bias in the amplification and/or detection of some trypanosome species over others (Desquesnes *et al.*, 2001). Additionally, some ITS1 primers are prone to non-specific amplification in bovine blood samples (Tran *et al.*, 2014). To address some of these bottlenecks, fluorescent fragment length barcoding (FFLB) method based on length variation in regions of the 18S and 28S ribosomal RNA gene region has been developed for the detection of trypanosome species (Hamilton *et al.*, 2008). Fluorescently tagged primers, designed in conserved regions of the 18S and 28S ribosomal RNA genes, are used. Amplicons are then analyzed using an automated DNA sequencer and species are identified by size variation of the amplified fragments. The method has been shown to be sensitive in trypanosome species identification and has the capacity to detect new species through the identification of unique barcodes (Adams *et al.*, 2008; Garcia *et al.*, 2018). However, the method is laborious for large sample sizes since it requires the use of four PCR reactions per sample. A major problem with the use of ribosomal RNA genes is that they cannot be used to distinguish between *Trypanozoon* subspecies (*T. b. brucei*, *T. b. rhodesiense*, *T. b. gambiense*, *T. evansi,* and *T. equiperdum*) (Haag,

O'hUigin and Overath, 1998; Stevens *et al.*, 2001; Stevens and Brisse, 2004; Hamilton *et al.*, 2008). Currently, *Trypanozoon* subspecies are identified by specific PCR (Ngaira *et al.*, 2004; Njiru *et al.*, 2006; Li *et al.*, 2007; Picozzi *et al.*, 2008) and microsatellites markers (Li *et al.*, 2007; Simo *et al.*, 2011; Sánchez *et al.*, 2015).

For large sample sizes, undertaking multiple PCRs for each sample can be both expensive and laborious. Recently, the metagenomic approach utilizing next-generation sequencing (NGS) has been applied as a capable and convenient method for profiling bacterial and fungal communities. It has many advantages including high sensitivity to detect low-frequency variants, higher throughput with sample multiplexing, and the lower limit of the detection of DNA among others. The concept is not specific for profiling of bacterial and fungal communities but more versatile; however, relatively few studies have applied this technology for the diagnostics of protozoal infections, with the exception of *Plasmodium* in mosquitoes, (Paparini *et al.*, 2015; Barbosa *et al.*, 2017). Next-generation sequencing is suited in the analysis of the genetic diversity of trypanosome genotypes because of the depth of read coverage which enables high-resolution at a single nucleotide level. Despite the ITS1 region having been shown to differentiate between various trypanosome species, no study has used it as a target for DNA metabarcoding of trypanosomes. The determination of trypanosome sub-species and genotypes provides an opportunity to understand anthropogenic disturbance that may change genotypes of trypanosomes infecting human and livestock (Van den Bossche *et al.*, 2010).

In this study, the ITS1 region is used as a target for the DNA metabarcoding of trypanosome species. The objective of this study is to provide researchers with a tool that can be used to characterize various genotypes and subspecies of trypanosomes without the need for performing species-specific tests.

2.3 Materials and Methods

2.3.1 Sample collection

Three groups of tsetse flies collected at three different locations were analyzed (Figure 4). The first group was used for validation of the new method and consisted of 188 tsetse flies collected from Hurungwe in Zimbabwe between March and April 2014. The second group consisted of 200 tsetse flies from Rufunsa area (Zambia) near Lower Zambezi National park (surrounding farms and villages) collected in November and December 2013). Information on tsetse fly species and sex was not available for this group. The third group consisted of 85 flies caught in Kafue National park, Zambia in an area bordering between and public settlement area and wild animal reserve area, collected in June 2017. For this group, flies were sorted according to sex and species. Flies from all three groups had been collected using either custom-made mobile traps attached on slow-moving vehicles (Kafue and Rufunsa groups) or Epsilon traps (Hurungwe group). The flies were preserved individually in silica gel before crushing and DNA extraction.

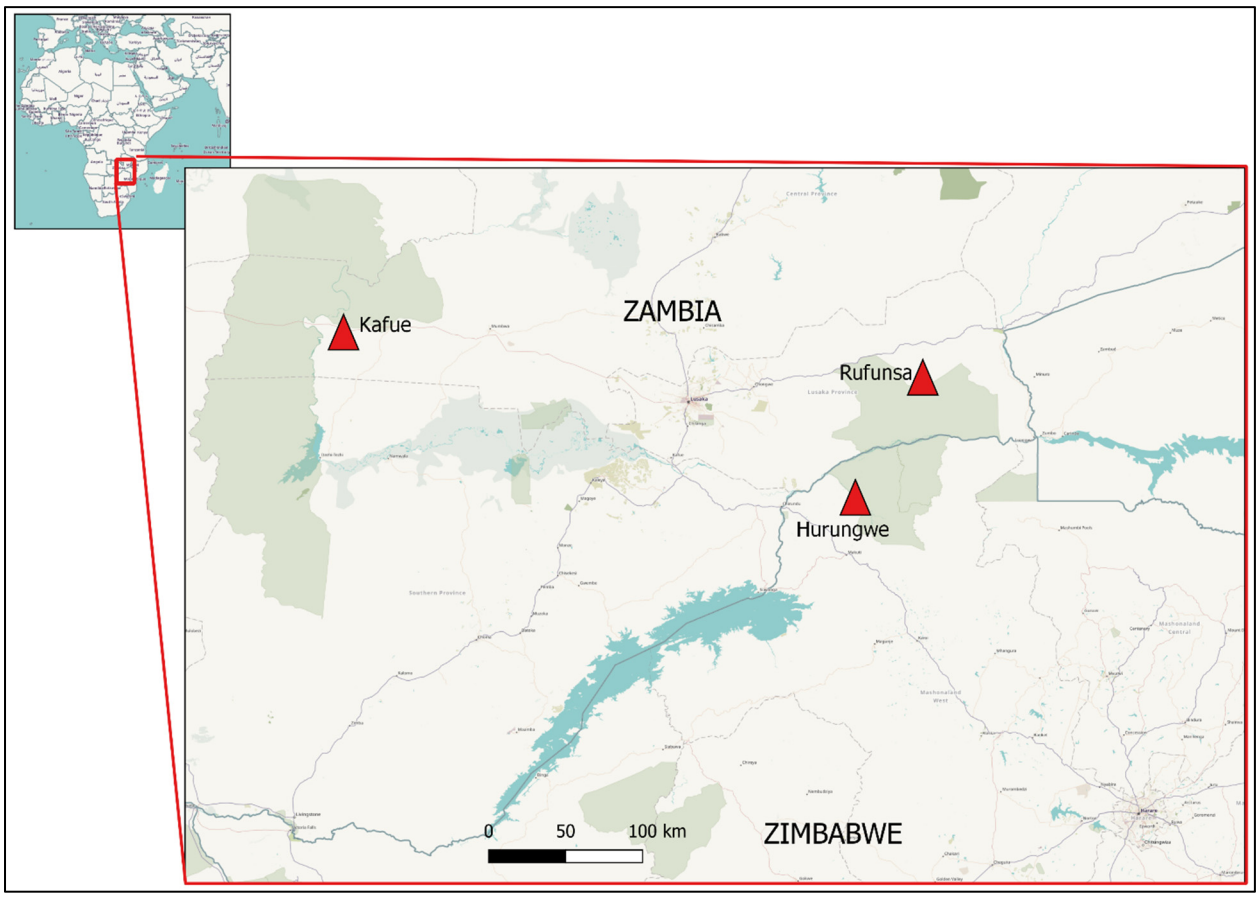

Figure 4. Tsetse fly collection sites from Zambia and Zimbabwe. The map was sourced from Open Access Maps (www.openaccessmap.org) and modified in Adobe Illustrator CC 2019 v23.0.1. Areas, where tsetse flies were caught, are marked with red triangles: Kafue in Zambia, Rufunsa in Zambia and Hurungwe in Zimbabwe.

2.3.2 Extraction of DNA

Extraction of DNA from all fly samples was done by briefly crushing dried flies in tubes containing stainless beads in smashing machine at 3,000 revolutions per minute for 45 sec. Isolation of DNA was performed using DNA isolation kit for mammalian blood (Roche USA) as per the manufacturer's protocol for the extraction of DNA from Buffy coat where red blood cell lysis step is bypassed allowing for lysis of all cells in the solution at once including trypanosomes by the white cell lysis buffer. The DNA samples were stored at -80ºC prior to further analysis.

2.3.3 Primer design

For primer design, the following list of sequences was retrieved from NCBI: *T. brucei* (JX910378, JX910373, JN673391, FJ712717, AF306777, AF306774, AF306771 and AB742530), *T. vivax* (JN673394, KC196703 and TVU22316), *T. congolense* (JN673389, TCU22319, TCU22318, TCU22317 and TCU22315), *T. simiae* (JN673387 and TSU22320), *T. godfreyi* (JN673385), *T. evansi* (D89527), *T. otospermophili* (AB175625), and *T. grosi* (AB175624). Multiple alignments were done in Geneious 9.1.5 software (Biomatters Ltd, Auckland, New Zealand) using MAFFT with default settings (Figure 5). The ITS1 region was identified by original sequence annotations and by custom annotation using previous primers (CF and BR primers). Pairs of primers flanking the ITS1 region were picked manually based on the consensus alignment sequence. The final primer pair was manually edited to improve the range of trypanosome species and subspecies and named Amplification of ITS (AITS) forward (AITSF) and reverse (AITSR) (Table 1). Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast) was used to confirm the target species and check the melting temperature.

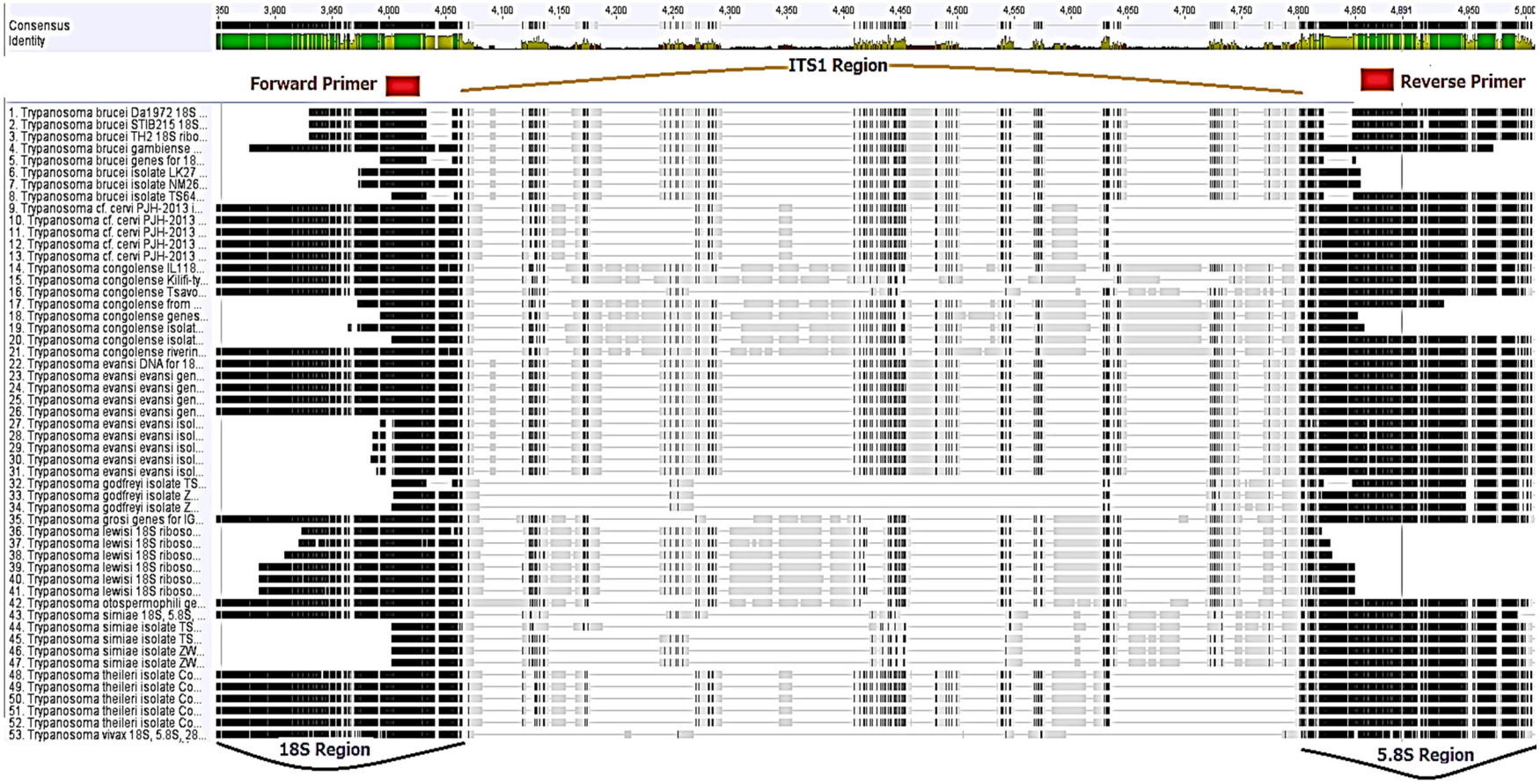

Figure 5. Alignment of ITS1 region from different trypanosome species showing the positions of the newly designed forward and reverse primer. The black and grey color represents conserved sequence regions while the white spaces represent mismatches. The lines in between these regions represent gaps. The top graph is the consensus sequence.

2.3.4 Experimental and *in silico* validation of primers

In silico testing

The sizes of expected target amplicons for the newly designed AITSF and AITSR primers were compared *in silico* with amplicons from three primers that are widely used for trypanosome species identification; CF/BR primers (Njiru *et al.*, 2005) and ITS1/ITS2 primers (Cox *et al.*, 2005). Simulate_PCR (Gardner and Slezak, 2014), a computer-based PCR simulation analysis algorithm, was used to extract ITS1 amplicons from NCBI *nt* database (downloaded on 3rd December 2017 from NCBI: ftp://ftp.ncbi.nlm.nih.gov/blast/db/) with each of the primer pairs to elucidate the scope of trypanosome species and subspecies detection and expected amplicon lengths. Simulate_PCR was run using the command with the minimum and maximum length of the amplicon was set to 100 and 750 b.p respectively, allowing one mismatch.

Sensitivity testing

To test the sensitivity of the AITSF/AITSR primers, PCR amplification on the ITS1 region of five trypanosome species was done and compared to results of PCR with CF/BR primers on the same templates. A Primer-BLAST was performed and subject sequences of different trypanosome species common to both AITSF/AITSR and CF/BR primers were picked. For each of the five trypanosome species, sequences with the 18S region and complete regions of ITS1 and 1.5S ribosomal RNA were chosen. Of these, one sequence was picked at random to represent each of the five trypanosome species resulting in the following sequences: *T. brucei* (AF306774), *T. simiae* (JN673387), *T. vivax* (KM391828), *T. congolense* (U22317) and *T. godfreyi* (JN673384). From this list, 18S to 5.8S ribosomal RNA region was obtained from each sequence, synthesized and cloned into a pGEMT-easy vector. Serial dilution of the vector stock solution was done to obtain working solutions of 10^1 to 10^7 ITS1 inserts per microliter. These dilutions were used as templates

for PCR reaction (1 µL per reaction) using either AITSF/AITSR or CF/BR primer sets. The PCR analysis was done in 10 µL primary reactions containing 0.5 µL of 10 µM of each of the forward and reverse primers, 10 µL of 2X Ampdirect® Plus buffer, 0.16 µL of 5 U/µL Taq polymerase (Kapa Biosystems, Boston, USA), 0.4 µL DMSO, and 1 µL extracted DNA as a template. An initial incubation at 95°C was done for 10 min, followed by 37 cycles of 95°C for 30 sec, annealing at 60 °C for 1 min for AITSF/AITSR primers and 58 °C for 1 min for CF/BR primers, extension for 72°C for 2 min, and a final extension at 72°C for 10 min. Amplicons were viewed on 1.5% Agarose gel under UV trans-illuminator (Atto Corporation, Tokyo, Japan).

2.3.5 Library preparation

For the detection of trypanosome species in tsetse fly samples, a two-step PCR protocol for the library preparation was adopted for multiplex PCR analysis using Illumina adapter sequences ligated to AITSF/AITSR primer set (Table 1).

Table 1. Primers used in this study.

Description	Primer name	Primer sequence (5'-3')
From Ref (Njiru *et al.*, 2005)	ITS1 CF	CCGGAAGTTCACCGATATTG
	ITS1 BR	TTGCTGCGTTCTTCAACGAA
New ITS1 forward primer	AITSF	CGGAAGTTCACCGATATTGC
New ITS1 reverse primer	AITSR	AGGAAGCCAAGTCATCCATC
Adapter sequence for the forward primer	Illumina adapter forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**NN** [AITSF]
Adapter sequence for the reverse primer	Illumina adapter reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NN** [AITSR]

[ ] indicates the position where the Illumina adapter was attached to the respective primer

For validation of the results, ITS1 PCR was done in duplicate for Rufunsa samples. Positive template controls comprising. *T. b gambiense, T. b rhodesiense,* and *T. congolense* DNA harvested from culture were included. To prepare an artificial mixture mimicking mixed infection, *T.b. gambiense and T. congolense* DNA was artificially mixed in equal proportions. These controls were processed as same as samples from PCR to sequencing. The first PCR reaction was performed using AITSF/AITSR primers (adapter ligated forms where Illumina adapter sequences) and sequencing libraries were prepared according to the Illumina MiSeq system instructions (Illumina, 2013).

The first PCR was done in 20 µL primary reactions containing 0.5 µL of 10 µM each of the AITSF and AITSR primers, 10 µL of 2X Ampdirect® Plus buffer, 0.16 µL of 5 U/µL Taq polymerase, 0.4 µL DMSO, and 1 µL extracted DNA as a template. An initial incubation was done at 95°C for 10 min, followed by 37 cycles of 95°C for 30 sec, annealing at 60 °C for 1 min, 72°C for 2 min, and a final extension at 72°C for 10 min.

The second PCR was done in 10 µL reactions containing 1 µL of 10 µM Illumina dual-index primer mix (i5 and i7 primers), 1.2 µL of 25 mM $MgCl_2$, 0.4 µL of 10 mM each of the dNTPs, 0.1 µL of 5 U/µL Taq polymerase, 4 µL 5X buffer, and 2 µL of template (1/60 diluted PCR product from first PCR). An initial incubation was done at 95°C for 3 min, followed by 11 cycles of 95°C for 30 sec, 61°C for 1 min, 72°C for 2 min and a final extension at 72°C for 10 min. A negative template control was included in each set of PCR reactions. All amplicons were run in 1.5% agarose gels viewed in a UV transilluminator. To enable the sequencing of all amplicons in this study in one run, different sets of dual index primers for each sample in the second PCR reactions were used.

2.3.6 Library sequencing

All the products from the second PCR were pooled into a single library by mixing equal volumes from each sample. The library pool was purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) by cutting out bands of similar sizes so that similar sized bands were all pooled into a single tube. This process also ensured purification from primer dimers and post PCR reagents. Quantification of each size separated library was done using a Qubit dsDNA HS assay kit and a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The concentration of the libraries was then adjusted to a final concentration of 4 nM using nuclease-free water and sequenced in the MiSeq platform (Illumina, San Diego, CA, USA). Sequencing was performed using a MiSeq Reagent Kit v3 (150-cycle) (for 300 base pairs, paired-end (Illumina, San Diego, CA, USA) with a PhiX DNA spike-in control added to improve the data quality of low diversity samples.

Raw data obtained from this study is available at Sequence read archive (SRA) database under the SRA accession number SRP159480.

2.3.7 Bioinformatics

Raw read data was analyzed using the Amplicon Tool Kit pipeline (AMPtk) (version 1.2.4) (https://github.com/nextgenusfs/amptk) as detailed in the workflow shown (Figure 6) with taxonomic identification was done using BLAST. All commands for analysis were run as a custom script (Appendix 1). Briefly, reads were processed by; 1) Trimming off primers, removal of sequences less than 100 b.p and pair-end merging of the reads. Custom merging parameters were used by editing the AMPtk file amptklib.py file where USEARCH options were changed to *fastq_pctid* 80, (minimum %id of alignment), *minhsp* 8, and *fastq_maxdiffs* 10. 2) Clustering using the denoising algorithm (Divisive Amplicon Denoising Algorithm) DADA2 using the *amptk dada2* command. This algorithm provides a clustering independent method that "corrects" or

"denoises" each sequence using statistical modeling of sequencing errors. It produces either the standard "inferred sequences" or amplicon sequence variants (ASVs) and also clusters the ASVs into biologically relevant operational taxonomic units (OTUs) using the UCLUST algorithm. 3) The ASV table was filtered to correct for index-bleed where a small percentage of reads bleed into other samples using the *amptk filter* command at 0.005, the default index-bleed percentage. 4) Additional filtering of the ASV table was done using the *amptk lulu* command which removes erroneous ASVs by high-throughput sequencing of amplified marker genes (Frøslev *et al.*, 2017). 5) Taxonomy was assigned to the final ASV table by BLAST (v2.6.0) (Camacho *et al.*, 2009) using custom options specified as shown in Figure 2.

To check the accuracy of the ASVs generated by this pipeline, analysis of FASTQ files generated *in silico* from sequences from a previous study (Auty *et al.*, 2012) was done using *ArtificialFastqGenerator* (Frampton and Houlston, 2012), which generated paired-end FASTQ files with 1000 reads per sequence. The pipeline was then run on the generated reads and the results compared to the species identity of parent sequences. All ASVs generated in this study are deposited in GenBank submission: SUB4757113 with accession numbers MK131764 - MK132190.

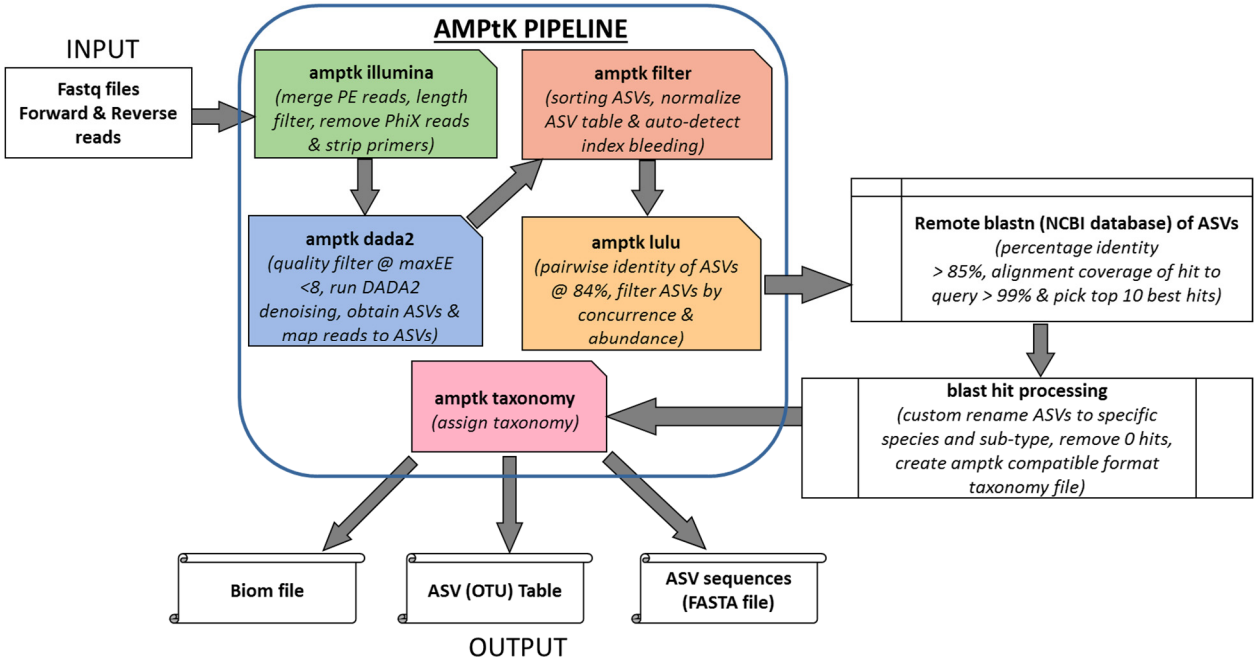

Figure 6. Workflow for read analysis using AMPtk pipeline. The raw reads are processed by quality filtering and merging followed by error correction using dada2. The low abundance ASVs are filtered out or merged forming the final OUT table. Taxonomy is then assigned using blast.

2.3.8 Phylogenetic and statistical analysis

A phylogenetic tree was created from the alignment of ASVs obtained from the analysis of all the tsetse fly samples. Alignments were made with MAFFT (Katoh and Standley, 2013) using *mafft-xinsi* (that allows for the prediction of RNA secondary structure in a multi-structural alignment) at 1,000 maximum iterations, leaving gappy regions, with 'kimura 1' option as score matrix. RAxML 8.0.26 was used for phylogenetic tree generation by maximum likelihood method using the 'GTRCATI' model at 10,000 bootstraps. The tree was visualized and annotated using iTOL (version 4) (Letunic and Bork, 2011). Statistical analysis and graphing of data were done in GraphPad Prism version 6.01 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com.

2.4 Results

2.4.1 Performance of new primers

Results obtained from *in silico* analysis showed that the AITSF/AITSR primers had a broad range similar to the ITS1/ITS2 primer set previously developed (Cox *et al.*, 2005) while the CF/BR primers previously developed to detect pathogenic Trypanosomes (Njiru *et al.*, 2005) had a range confined to the pathogenic trypanosomes only (Table 2). The sensitivity of newly designed AITSF/AITSR primers in the amplification of the ITS1 region of different trypanosome species was compared to the CF/BR primers. The evaluation was based on the visual sight of bands in a gel (the conventional method of analysis). Results showed that the AITSF/AITSR primers were comparatively more sensitive in the detection of *T. brucei*, *T. simiae* and *T. congolense* (Figure 7). The AITSF/AITSR primers detected 10^3 *T. brucei, T. simiae*, *T. vivax* and *T. congolense* and *T. godfreyi* ITS1 copies while CF/BR primers could detect 10^3 *T. godfreyi* and *T. vivax* ITS1 copies, 10^4 *T. simiae* and *T. congolense* ITS1 copies and 10^5 *T. brucei* ITS1 copies. Trypanosomes have 100 to 200 copies of ribosomal RNA genes (Desquesnes and M.R. Dávila, 2002), and thus the AITSF/AITSR primers can detect at least 100 trypanosomes of all five species.

Table 2. Amplicon sizes of new primers (ATSF/AITSR) compared to other primers (CF/BR and ITS1/ITS2) obtained by simulated PCR.

Trypanosome species/sub-species	**AITSF/R Amplicon lengths (b.p)**	**CF/BR Amplicon lengths (b.p)**	**ITS1/ITS2 Amplicon lengths (b.p)**
Trypanaosoma congolense other isolates	633 - 705	686 - 717	600 - 674
Trypanaosoma congolense riverine/forest-type	660	713	-
Trypanosoma congolense Kilifi-type	560	613	529
Trypanosoma cruzi	544 - 570	-	512 - 518
Trypanosoma rangeli	525	-	455 - 494
Trypanosoma cf. varani	511	-	480
Trypanosoma kuseli	504 - 506	-	473
Trypanosoma otospermophili	504	-	472
Trypanosoma lewisi	504	-	473 - 475
Trypanaosoma grosi	424 - 435	-	401 - 404
Trypanaosoma brucei/ evansi	415 - 431	468 - 484	384 - 394
Trypanosoma simiae	331 - 343	369-397	300 - 319
Trypanosoma cf. cervi	333 - 340	-	303 - 312
Trypanosoma avium	322	-	291
Trypanosoma congolense Tsavo-type	316	369	285
Trypanosoma theileri	269 - 350	-	238 - 265
Trypanosoma godfreyi	220	273	189
Trypanaosoma vivax	226-238	248 - 264	164 - 187

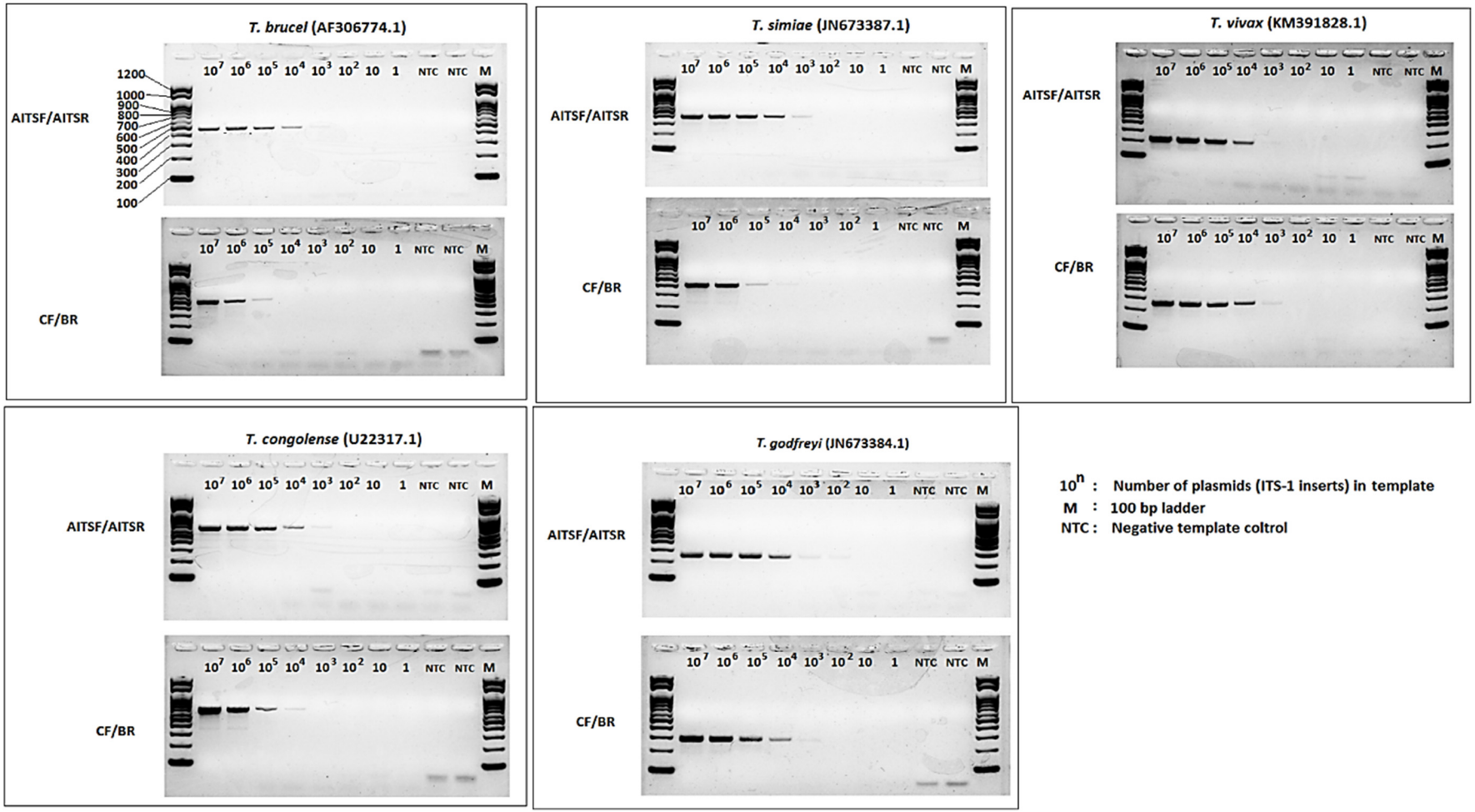

Figure 7. Gel images of products of AITSF/AITR and CF/BR primers showing the sensitivity of detection of trypanosome ITS1 inserts cloned in the pGEMT-easy vector. The AITSF/AITSR primers were comparatively more sensitive in the detection of *T. brucei*, *T. simiae*, *T. congolense,* and *T. godfreyi*. The sensitivity of the detection of T. vivax was similar for both AITSF/AITSR and CF/BR primers

2.4.2 Analysis of raw data from replicate analysis

Raw reads generated from amplicon sequencing were of moderately great quality where over 90% passed quality separating in all examples (with the exception of those from Hurungwe, Zimbabwe samples) (Table 3). The quantity of ASVs generated from replicate runs were marginally different showing slight contrasts in replicate runs. Only the forward reads were retained for downstream analysis during pre-processing, for the reads that did not merge either due to the amplicon being longer than 600 b.p (Figure 8) or because of low-quality bases. This did not affect the detection of trypanosome species. Both replicates had comparative outcomes with respect to trypanosome species detected by size examination (Figure 9A) and sequence analysis (Figure 9B). The result of identification for each of the trypanosome species and subspecies in replicate runs was comparable as shown by Fischer's exact test showing no significant difference ($P<0.05$) in the number of positive detections in replicate runs (Table 4).

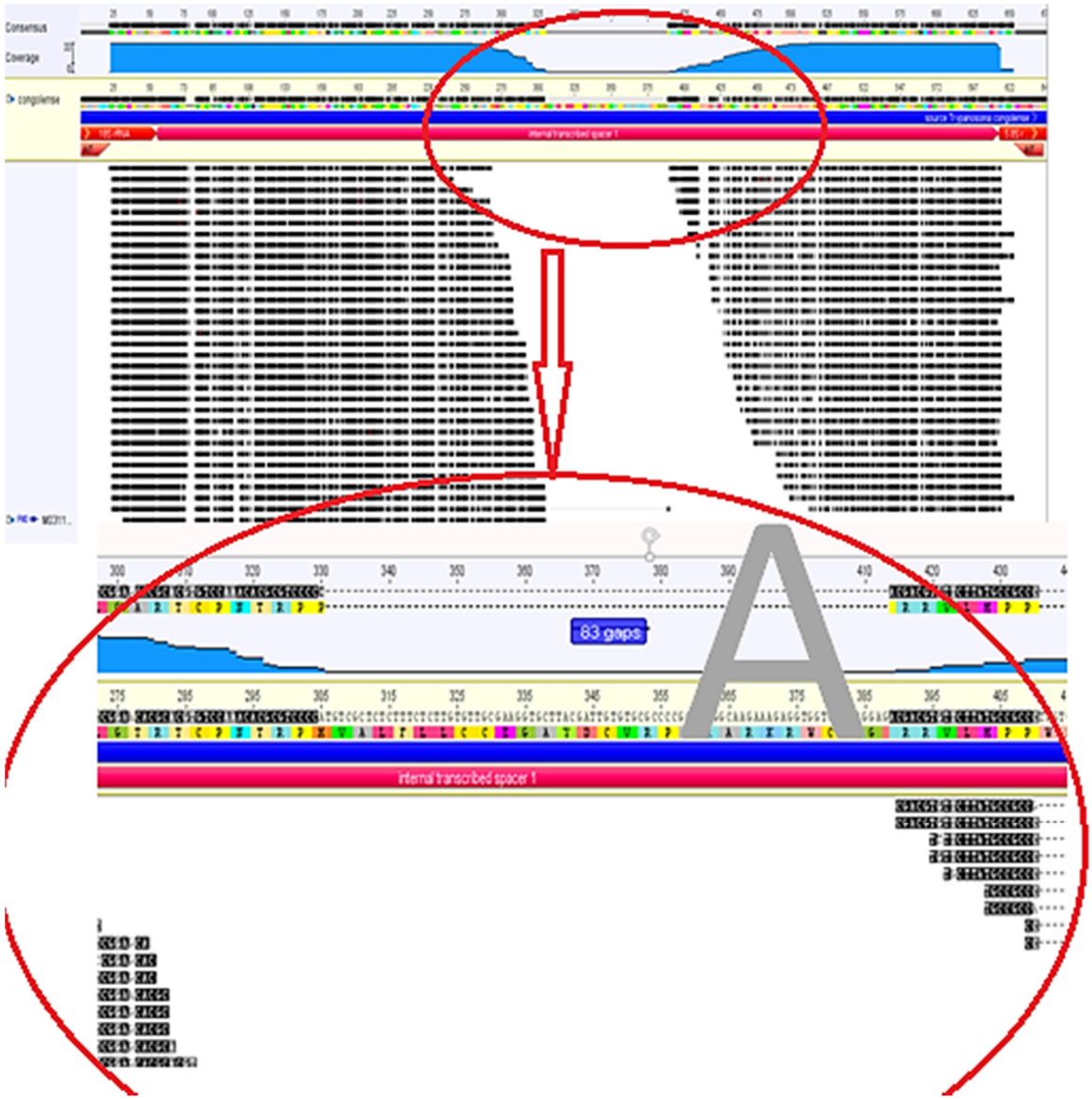

Figure 8. Pair-end reads from a *T. congolense* ITS1 amplicon (about 690 base pairs) longer than 600 base pairs mapped to a reference sequence. The amplicon is longer than 600 bases and therefore the reads cannot cover the whole length of the amplicon leaving a gap.

Table 3. Read data of all samples analyzed.

Source of sample	No. of samples	Total no. of reads	Reads after pre-processing (% of total)	Raw ASVs	OTUs (97% clustering of ASVs)	ASVs post-filtering
Rufunsa Run A	200	916,055	897,598 (99.8%)	269	89	174
Rufunsa Run B	200	1,289,667	1,248,934 (94.8%)	320	95	232
Kafue	85	483,589	454,799 (91.4%)	131	48	56
Hurungwe	188	29,798	11,247 (79.5%)	137	63	116

Amplicon sequence variants (ASVs) generated were filtered to remove underrepresented and/or artifact ASVs from the final taxonomy table.

Table 4. Statistical analysis of detection of individual trypanosome species in replicate runs.

Species/Sub-group	Run	Positive	Negative	Fisher's exact test P value Two-sided	P value summary (alpha<0.05)
T. brucei/T. evansi	Run A	8	192	0.4923	ns
	Run B	12	188		
T. congolense Forest	Run A	7	193	0.7996	ns
	Run B	9	191		
T. congolense Kilifi	Run A	12	188	0.6909	ns
	Run B	15	185		
T. congolense Savannah	Run A	14	186	1	ns
	Run B	15	185		
T. godfreyi	Run A	14	186	0.1062	ns
	Run B	6	194		
T. simiae	Run A	9	191	0.6549	ns
	Run B	12	188		
T. simiae Tsavo	Run A	9	191	1	ns
	Run B	8	192		
T. vivax	Run A	10	190	0.4092	ns
	Run B	15	185		

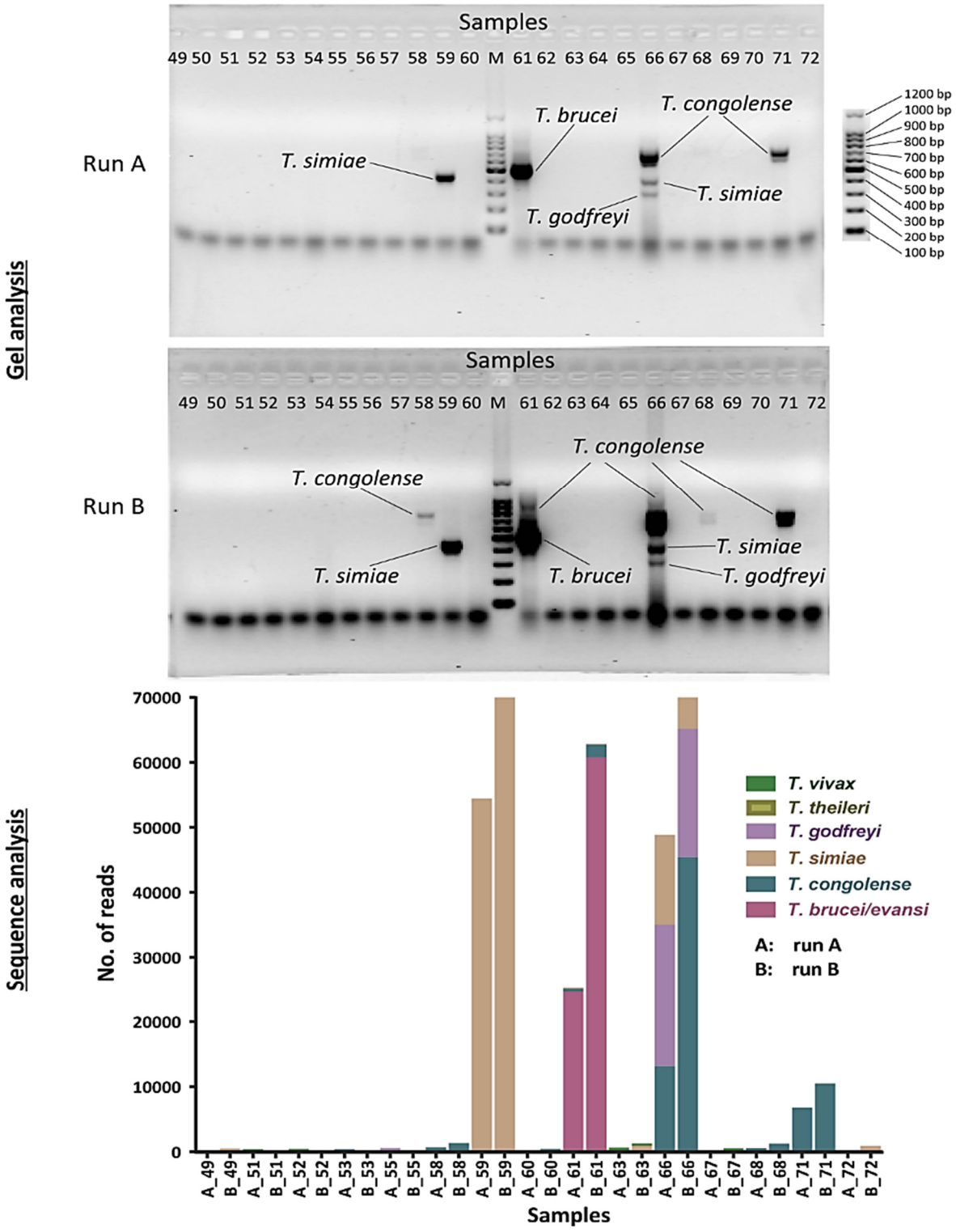

Figure 9. The results of representative replicates analysis. (A) Gel analysis of Rufunsa samples done in replicate showing matching bands per sample. (B) Amplicon sequence analysis of the same samples in showing the number of reads detected per species in each sample.

2.4.3 Validation of the pipeline and accuracy of detection

Results from analysis of simulation of test data demonstrated that ASVs produced by the pipeline were increasingly exact in effectively deciding the diversity of sequences in contrast to operational ordered units (OTUs) derived from clustering sequences at 97% identity (Figure 10). The specificity and accuracy of recognizing singular successions of the equivalent trypanosome species are reflected by the quantity of ASVs or OTUs representing each of the different species. For instance, only one OTU was generated for all three *Trypanosoma theileri* sequences, and three OTUs were generated for seven *Trypanosoma simiae* sequences, while the number of ASVs generated in each case represented each sequence accurately. The simulated data results indicated that read analysis using the AMPtk pipeline and ASVs instead of OTUs was suitable for sensitive identification of trypanosome reads.

OTU Taxonomy	AB175624 Trypanosoma grosi	AB175626 Trypanosoma kuseli	AB362411 Trypanosoma minasense	Y00055 Crithidia fasciculata	AY028450 Parabodo caudatus	JN673396 Trypanosoma theileri			JN673391 Trypanosoma brucei		JN673390 Trypanosoma brucei	U22320 Trypanosoma simiae	JN673386 Trypanosoma simiae	JN673387 Trypanosoma simiae	JN673382 Trypanosoma simiae Tsavo				JN673384 Trypanosoma godfreyi	JN673385 Trypanosoma godfreyi	MG255220 Trypanosoma vivax	JN673392 Trypanosoma vivax	JN673393 Trypanosoma vivax	JN673394 Trypanosoma vivax	U22317 Trypanosoma congolense Kilifi	U22319 Trypanosoma congolense Forest	JN673388 Trypanosoma congolense Savannah	JN673389 Trypanosoma congolense Savannah	JN673398 Trypanosoma sp. Z3206	JN673399 Trypanosomatidae sp. TS07016	JN673400 Trypanosomatidae sp. Z26907	JN673401 Trypanosomatidae sp. Z1505	JN673402 Trypanosomatidae sp. TS06050	JN673403 Trypanosomatidae sp. TS07116
OTU ID	OTU19	OTU1	OTU12	OTU20	OTU6	OTU28			OTU17		OTU18	OTU25	OTU22	OTU23	OTU24				OTU3	OTU2	OTU5	OTU9	OTU11	OTU8	OTU13	OTU15	OTU10	OTU4	OTU26	OTU23	OTU27	OTU7	OTU16	OTU14
ASV Taxonomy	AB175624 Trypanosoma grosi	AB175626 Trypanosoma kuseli	AB362411 Trypanosoma minasense	Y00055 Crithidia fasciculata	AY028450 Parabodo caudatus	JN673395 Trypanosoma theileri	JN673396 Trypanosoma theileri	JN673397 Trypanosoma theileri	AF306771 Trypanosoma brucei	JN673391 Trypanosoma brucei	JN673390 Trypanosoma brucei	U22320 Trypanosoma simiae	JN673386 Trypanosoma simiae	JN673387 Trypanosoma simiae	JN673379 Trypanosoma simiae Tsavo	JN673380 Trypanosoma simiae Tsavo	JN673381 Trypanosoma simiae Tsavo	JN673382 Trypanosoma simiae Tsavo	JN673384 Trypanosoma godfreyi	JN673385 Trypanosoma godfreyi	MG255220 Trypanosoma vivax	JN673392 Trypanosoma vivax	JN673393 Trypanosoma vivax	JN673394 Trypanosoma vivax	U22317 Trypanosoma congolense Kilifi	U22319 Trypanosoma congolense Forest	JN673388 Trypanosoma congolense Savannah	JN673389 Trypanosoma congolense Savannah	JN673398 Trypanosoma sp. Z3206	JN673399 Trypanosomatidae sp. TS07016	JN673400 Trypanosomatidae sp. Z26907	JN673401 Trypanosomatidae sp. Z1505	JN673402 Trypanosomatidae sp. TS06050	JN673403 Trypanosomatidae sp. TS07116
Representative Sequences / ASV ID	ASV19	ASV1	ASV12	ASV21	ASV6	ASV34	ASV33	ASV32	ASV20	ASV17	ASV18	ASV26	ASV23	ASV24	ASV28	ASV27	ASV25	ASV29	ASV3	ASV2	ASV5	ASV9	ASV11	ASV8	ASV13	ASV15	ASV10	ASV4	ASV30	ASV22	ASV31	ASV7	ASV16	ASV14
AB175624TrypanosomaGrosi	10000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AB175626TrypanosomaKuseli	0	9992	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AB362411TrypanosomaMinasense	0	0	9970	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y00055CrithidiaFasciculata	0	0	0	9999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AY028450BodoCaudatus	0	0	0	0	9793	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673395TrypanosomaTheileriPuku	0	0	0	0	0	9928	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673396TrypanosomaTheileriPuku	0	0	0	0	0	0	9998	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673397TrypanosomaTheileriPuku	0	0	0	0	0	0		9949	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AF306771TrypanosomaBrucei	0	0	0	0	0	0	0	0	9997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673391TrypanosomaBruceiSpottedhyena	0	0	0	0	0	0	0	0	0	10000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673390TrypanosomaBruceiZebra	0	0	0	0	0	0	0	0	0	0	9997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U22320TrypanosomaSimiae	0	0	0	0	0	0	0	0	0	0	0	9997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673386TrypanosomaSimiaeWarthog	0	0	0	0	0	0	0	0	0	0	0	0	9993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673387TrypanosomaSimiaeWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	9996	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673379TrypanosomaSimiaeTsavoWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673380TrypanosomaSimiaeTsavoWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9995	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673381TrypanosomaSimiaeTsavoWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673382TrypanosomaSimiaeTsavoWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673384TrypanosomaGodfreyiWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9942	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673385TrypanosomaGodfreyiWarthog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9954	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U22316TrypanosomaVivax	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9787	0	0	0	0	0	0	0	0	0	0	0	0	0
JN673392TrypanosomaVivaxGiraffe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9809	0	0	0	0	0	0	0	0	0	0	0	0
JN673393TrypanosomaVivaxWaterbuck	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9769	0	0	0	0	0	0	0	0	0	0	0
JN673394TrypanosomaVivaxCapeBuffalo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9782	0	0	0	0	0	0	0	0	0	0
U22317TrypanosomaCongolenseKilifi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9992	0	0	0	0	0	0	0	0	0
U22319TrypanosomaCongolenseForest	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9990	0	0	0	0	0	0	0	0
JN673388TrypanosomaCongolenseSpottedhyena	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9994	0	0	0	0	0	0	0
JN673389TrypanosomaCongolenseSavannahLion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9990	0	0	0	0	0	0
JN673398TrypanosomaSPHippopotamus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9994	0	0	0	0	0
JN673399TrypanosomatidaeSPSpottedhyena	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10000	0	0	0	0
JN673400TrypanosomatidaeSPCapeBuffalo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9996	0	0	0
JN673401TrypanosomatidaeSPHippopotamus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9992	0	0
JN673402TrypanosomatidaeSPWildebeest	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9933	0
JN673403TrypanosomatidaeSPWildebeest	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9956

Figure 10: Matrix comparison of ASVs and OTUs from simulated data. The ASVs were more specific in predicting correctly the trypanosome species of the original mock sequences than the operational taxonomic sequences OTUs.

2.4.4 The sensitivity of detection of amplicons is improved by sequencing revealing errors of detection in conventional ITS1 PCR-gel analysis

By comparing the results of the analysis of amplicon size by gel with those of sequence analysis, it was observed that sequencing increased the sensitivity of the detection of trypanosome DNA. Samples with bands barely visible after the 1st PCR became visible after the 2nd PCR and were confirmed as positive after sequencing (Figure 11A). It was also observed that some *T. godfreyi* and *T. vivax* amplicon bands were of a relatively similar size and it was difficult to distinguish the two by gel analysis alone (Figure 11B). In this Figure, sample number 10 has an ITS1 amplicon size of about 400 b.p similar to that of sample number 6 and 8. Sequence analysis showed that the band in sample number 10 was identified as *T. vivax* while bands observed in sample numbers 6 and 8 were identified as *T. godfreyi* despite their similar sizes. Mixed and single infections represented by multiple and single bands respectively were observed and confirmed by amplicon sequence analysis. Results for the second PCR using dual-index primers showed consistency with those of the first PCR. There were no bands visible outside the expected range indicating the absence of non-specific amplification in both PCR steps. The 1st PCR amplicons were slightly longer than expected sizes due to the adapter sequences (approx. 80 b.p) added to the primer, therefore the bands observed corresponded to *T. congolense (Kilifi/Forest and Savannah);* 650-800 b.p, *T. brucei;* 520-540 b.p, *T. simiae;* 440-500 b.p, *T. godfreyi;* 320-400 b.p, and *T. vivax;* 290-400 b.p.

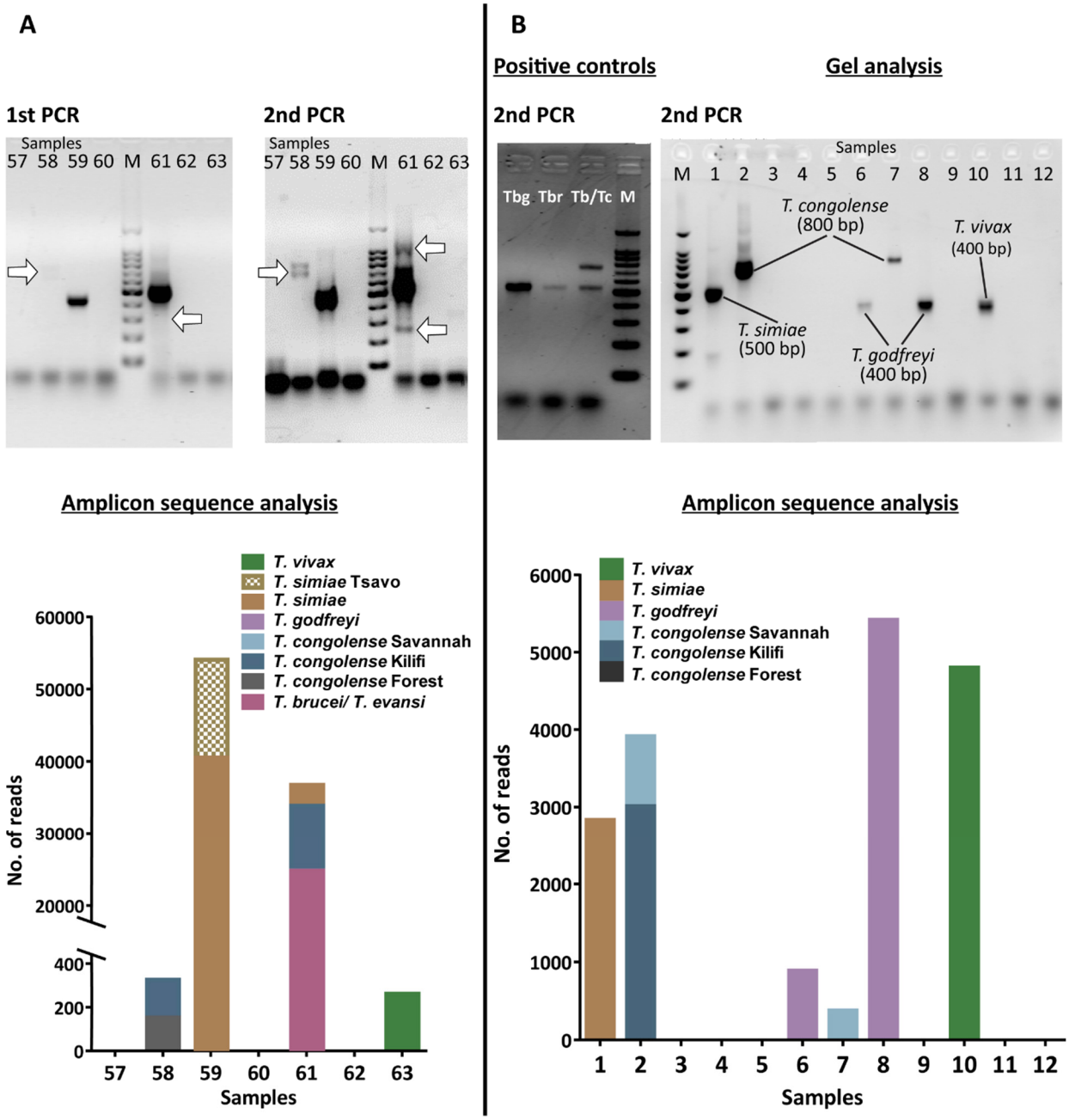

Figure 11. Comparison of gel and sequence analysis results. (A) Arrows showing bands are not visible after the 1st PCR become visible after 2nd PCR. (B) By gel analysis, amplicon bands of samples 5, 7 and 10 are indistinguishable by size and are deemed to be all *T. godfreyi* while sequencing reveals that the amplicon of sample 10 is, in fact, T. vivax. Positive controls comprise; Tbg (*T. b. gambiense*), Tbr (*T. b. rhodesiense*), Tb/Tc (an artificial mixture of equal amounts of *T. b. gambiense* and *T. congolense* DNA).

2.4.5 Trypanosome ITS1 sequences are able to distinguish between different trypanosome species and subspecies but not for the *Trypanozoon* subgenus.

Using phylogenetic analysis of ASV sequences, the accuracy to distinguish between trypanosome species and subspecies was analyzed. All ASVs were named after the area of the origin of the sample, ASV number, accession number and the taxonomic name of the top hit subject sequence after BLAST. It was observed that ASVs named after the same trypanosome species clustered together regardless of areas of sample collection. It was also observed that ASVs of the same trypanosome species sub-clustered into distinct subspecies (Figure 12A). The *Nannomonas* subgenus showed the highest diversity of sub-clustering where *T. simiae* clustered into two main clusters; *T. simiae* and *T. simiae* Tsavo. Two *T. simiae* Tsavo II ASVs from Kafue, with 91% and 97% identity to *T. congolense* Tsavo (Accession number U22318) which was recently reviewed and classified as *T. simiae* Tsavo (Gibson *et al.*, 2001; Gibson, 2003), clustered separately from the rest of the *T. simiae* Tsavo I ASVs from Rufunsa. It was also observed that *T. congolense* ASVs had the highest diversity and clustered into three main subspecies; Kilifi, Riverine/Forest, and Savannah. While *T. congolense* Savannah represented the most diverse group of the ASVs analyzed from the samples, *T. congolense* Kilifi clustered separately and far from *T. congolense* Savannah and Riverine/Forest subspecies. *T. godfreyi* showed sub-clustering into two main subspecies while *T. vivax* (belonging to the *Dutonella* subgenus) also clustered into two subspecies (Figure 12B). It was expected that the *Trypanozoon* subgenus (*T. brucei/T. evansi*) did not show any distinct sub-clustering.

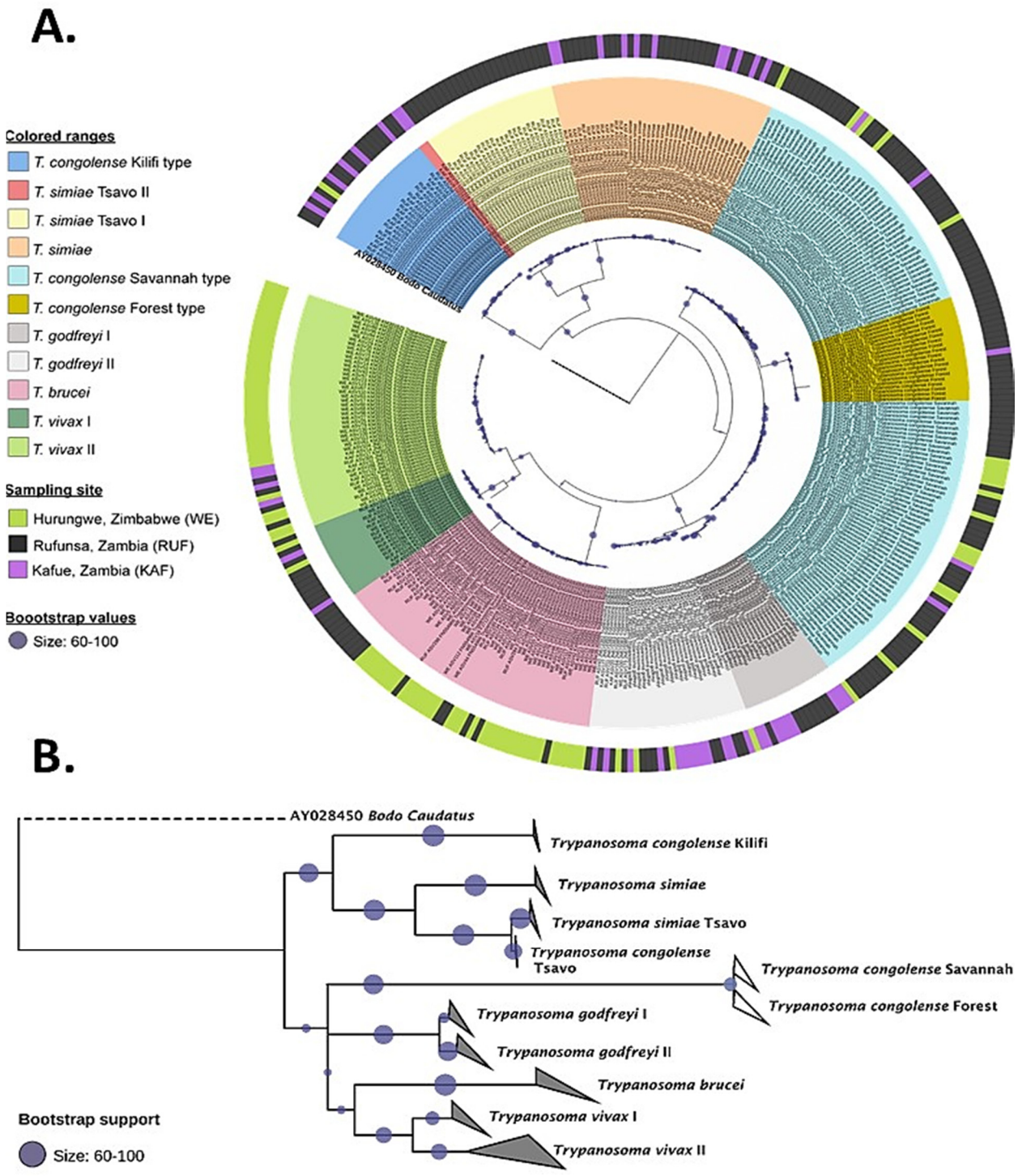

Figure 12. The distinct grouping of unique ASVs generated from amplicon sequence data. *Bodo caudatus* was included as an outgroup. (A) Individual trypanosome species and subspecies cluster into distinct clades shown in (B). ASV are named after their respective blast best hit matches.

2.4.6 Prevalence and distribution of trypanosome species in tsetse flies

The prevalence of trypanosome infection in tsetse flies collected in Rufunsa, Zambia, was 25.6%, Kafue, Zambia had a prevalence of 28.2%, and Hurungwe had a prevalence of 47.3%. The highest prevalence of *T. congolense* was seen in Rufunsa, Kafue had the highest prevalence of *T. godfreyi* (Table 5) and Hurungwe had the highest prevalence of *T. brucei/ T. evansi.* No *T. brucei/ T. evansi* was detected from flies collected in Kafue. Mixed infections were detected in flies from Rufunsa and Hurungwe while Kafue flies were predominantly infected with *T. godfreyi* (Figure 13) and no single infections with *T. congolense* or *T. godfreyi* were observed in flies from Hurungwe. It was observed that the infection rate in female flies (38.6%) was more than twice that of male flies (17.1%) from flies in Kafue (the only group that sex of flies was determined during collection). However, no *T. congolense* and *T. vivax* infections were detected in male flies.

Table 5. Prevalence of Trypanosome species infection in wild-caught tsetse flies.

Trypanosome species	Rufunsa (n=200)	Kafue (n=85)	Hurungwe (n=188)
Trypanozoon	6.0% (3.5% - 10.2%)	0.0% (0.0% - 4.3%)	45.7% (38.8% – 52.9%)
***T. congolense* Forest**	4.5% (2.4% - 8.3%)	1.2% (0.2% - 6.4%)	0.0% (0% - 2.0%)
***T. congolense* Kilifi**	7.5% (4.6% - 12.0%)	2.4% (0.7% - 8.2%)	4.8% (2.5% - 8.9%)
***T. congolense* Savannah**	7.5% (4.6% - 12.0%)	4.7% (1.9% - 11.5%)	39.9% (33.2% - 47.0%)
T. godfreyi	3.0% (1.4% - 6.4%)	16.5% (10.1% - 25.8%)	3.7% (1.8% - 7.5%)
T. simiae	6.0% (3.5% - 10.2%)	5.9% (2.5% - 13.0%)	1.1% (0.3% - 3.8%)
***T. simiae* Tsavo**	8.7% (4.5% - 16.2%)	2.4% (0.7% - 8.2%)	0.0% (0% - 2.0%)
T. vivax	7.5% (4.6% - 12.0%)	2.4% (0.7% - 8.2%)	29.2% (23.2% - 36.1%)
Trypanosoma (overall prevalence)	26.5% (20.9% -33.0%)	28.2% (19.8% - 38.6%)	47.3% (40.3% - 54.5%)

Confidence levels at 95% for apparent prevalence (Wilson) are shown in brackets.

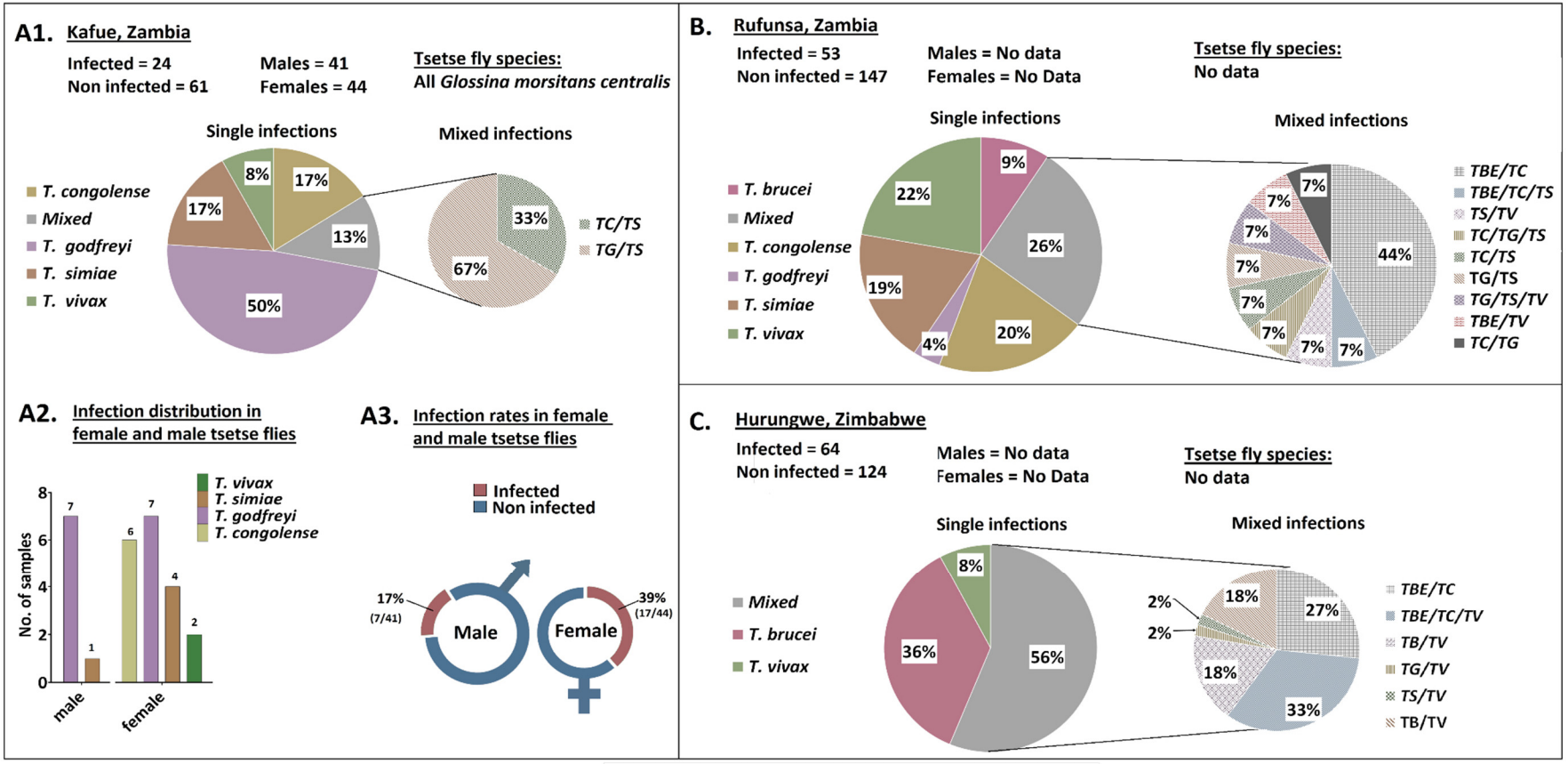

Figure 13. The distribution of trypanosome species amongst infected tsetse flies. (A1) Pie chart showing the prevalence of trypanosome species in wild-caught tsetse flies collected from Kafue, Zambia. (A2) Bar graph showing the Trypanosome species infecting male and female flies from flies collected in Kafue, Zambia. (A3) Graphical representation showing the infection rates of male and female flies collected in Kafue, Zambia. (B) Pie chart showing the prevalence of Trypanosome species in wild-caught tsetse flies collected from Hurungwe, Zimbabwe. (C) Pie chart showing the prevalence of Trypanosome species in wild-caught tsetse flies collected from Rufunsa, Zambia. TBE = *T. brucei*/*T. evansi*, TV = *T. vivax*, TS = *T. simiae*, TG = *T. godfreyi*, and TC = *T. congolense*.

2.5 Discussion

This study reports a new and versatile approach for the detection of trypanosome DNA in multiple samples with high sensitivity and precision than standard PCR-gel approach. It was determined that conventional ITS PCR gel analysis is not an accurate way of determining the prevalence of trypanosome species infections because the identification of species by band size may lead to misidentification of some trypanosome species. With this new approach, it will be possible to detect novel trypanosome species and subspecies by sequence analysis on individual samples in studies involving large sample sizes (approximately 700-sample-mixed library can be sequenced in one run owing to the high repertoire of Illumina dual indexing primers). It will also allow researchers to make a better inference on trypanosome species present in an area. This approach is practical and, with the decreasing cost of next-generation sequencing, a cost-effective way to monitor large field samples from vectors and mammalian hosts.

Two distinct subgroups of *T. vivax* and *T. godfreyi* were identified. This is important since *T. vivax* is distributed outside Africa and can be transmitted both cyclically by tsetse flies and also mechanically while *T. godfreyi* etiology and pathogenesis are poorly understood. This method, therefore, creates a platform for more studies of these two trypanosome species. From this study, it was determined that *T. vivax* and *T. godfreyi* have similarly sized ITS1 amplicons making it difficult to distinguish between them based solely on gel band sizes. Sequencing and clustering of the reads effectively address this issue.

Previous studies have indicated that the ribosomal RNA genes are highly conserved in this subgenus and cannot tell apart the subspecies (Haag, O'hUigin and Overath, 1998; Stevens *et al.*, 2001). The method was not able to distinguish between the *Trypanozoon* species which are of great focus because 1) they cause HAT (*T. b. rhodesiense* and *T. b. gambiense*) and 2) their

distribution is not restricted to Africa (*T. evansi* and *T. equiperdum*). This partly explains the failure to distinguish between *Trypanozoon* species. Moreover, a study based on genome-wide SNP analysis of 56 *Trypanozoon* genomes, including eight *T. evansi* and four *T. equiperdum* revealed extensively similar genomes (Cuypers *et al.*, 2017). Thus, a single molecular test able to distinguish between members of the *Trypanozoon* species is yet to be developed thus, subspecies specific based tests remain obligatory for their identification. As part of this work, new primers to amplify the trypanosome ITS1 region were developed. They have high specificity to *T. brucei* compared to other widely used primers and cover a wider range of the *Trypanosoma* genus.

Phylogenetic analysis showed several interesting population substructures in the case of *T. simiae* and *T. congolense*. Within the *T. congolense* clade, Savannah and Riverine/Forest subspecies had more sequence similarity while the Kilifi type displayed more divergence. This agrees with a previous study that reported a 71% similarity in satellite DNA sequences of *T. congolense* Savannah and Riverine/Forest (Garside and Gibson, 1995) and that the Kilifi subspecies was as divergent from other *T. congolense* subspecies (Masiga *et al.*, 1992). The clustering of *T. congolense* Kilifi close to *T. simiae* species than other *T. congolense* subspecies is interesting since an earlier study had identified a new *T. congolense* Tsavo strain (Accession number U22318) (Majiwa *et al.*, 1993) which has since been classified as *T. simiae* Tsavo (Urakawa *et al*., 1998) showing that *T. congolense* and *T. simiae* are closer than previously thought. Two ASVs from Kafue area (classified as *T. simiae* Tsavo II in this study) that had 91% and 97% identity to the U22318 *T. congolense* Tsavo sequence and that clustered with *T. simiae* Tsavo rather than other *T. congolense* species sequences supporting the *T. simiae* Tsavo classification. However, they cluster separately from the other *T. simiae* Tsavo ASVs, suggesting that they may have a divergent genotype. Probably, there is a complex relationship between *T. congolense* and *T. simiae* species yet to be revealed.

The prevalence of trypanosome infection in tsetse flies from the sampled areas with single and mixed infection is consistent with previous studies (Lehane *et al.*, 2000; Peacock *et al.*, 2007; Barbosa *et al.*, 2017). This may be an important factor in the exchange of information between species within the tsetse fly. In flies collected in Kafue (all identified as *G. m. centralis*), the infection rate of female tsetse flies was twice more than that of male flies. This is in contrast to dissection data from the Tinde experiment where male *Glossina morsitans centralis* had a salivary gland infection rate (5.4%) of more than twice that of the female flies (2.1%) *(Maudlin et al.*, Other studies on *Glossina morsitans*, have reported high infection rates in female flies compared to males thus, more research is needed to find out the role of sex and infection rate in different *Glossina* species. (Peacock *et al.*, 2012; Isaac *et al.*, 2016). It is worth noting that the detection of trypanosome DNA is not a confirmation of infection of the tsetse fly since the possibility of detecting parasite DNA that was ingested with blood from an infected host cannot be ruled out.

In conclusion, this study reports a new approach for the detection and characterization of trypanosomes from field samples accurately (with the exception of *Trypanozoon* subgenus) and therefore infer the prevalence of infection more precisely using a single test without having to undertake satellite DNA analysis that requires species-specific primers. It takes advantage of deep sequencing enabling resolution at a single nucleotide level. This high resolution applied to the ITS1 region has not been studied before. However, this study reveals that this method is practical and sensitive enough for barcoding of African trypanosomes. With this method, it is possible to distinguish *T. godfreyi* from *T. vivax*, as well as highlight finer subpopulation structures within the *T. simiae* and *T. congolense* clades that reveals interesting questions regarding their classification. This method offers a tool for detection of trypanosomes in large sample size samples with high specificity in distinguishing between trypanosome subspecies. It also offers the potential of detection of novel trypanosomes subspecies and genotypes. Therefore, this method will contribute

to the understanding of the differences between *T. vivax*, *T. godfreyi* and *T. congolense* subspecies, such as the contribution of these different genotypes to livestock disease, the role of these genotypes in assumed "strain" differences in drug response, and whether these genotypes are to be correlated with the old morphological species designations. Importantly if these "strains" have the potential of evolving to new subspecies, that could pose new risks.

CHAPTER TWO

3 CHAPTER TWO: Analysis of mammalian blood meal sources and bacterial microbiome diversity in wild-caught tsetse flies using next-generation sequencing

3.1 Summary

Tsetse flies transmit trypanosomes which cause African trypanosomiasis endemic in 36 countries in sub-Saharan Africa. The East-African HAT is considered a zoonosis and the parasites (*T. b. rhodesiense)* are maintained in wild and domestic animal. Identification of animal reservoirs is, therefore, an important aspect in the control of the disease. Moreover, how blood meals from different animals affect the biology of tsetse flies is also not well understood. In this study, tsetse flies from two endemic foci, Kafue, Zambia and Hurungwe, Zimbabwe, were analyzed in order to identify mammalian blood meal sources. The bacterial community in each fly was also analyzed and the infecting trypanosome species in infected flies was determined as well. Mammalian DNA from various wild animals (warthog, kudu waterbuck bush pig, buffalo, African hunting dog) was detected as well as domestic animals (dog and cattle), small mammals (rat and bat) and humans. This is the first report of bats as a blood meal source of wild tsetse flies. More research is therefore needed to reveal the role of small mammals such as rodents and bats in trypanosome epidemiology. No significant change was seen in the bacterial microbiome with changes in blood meal sources or with trypanosome infection. This is in contrast with studies of other arthropod vectors (mosquitoes and ticks) that show microbiome changes with changes in blood meal diet.

3.2 Introduction

Tsetse flies inject trypanosome parasites through biting an infected host and feeding on infected blood. The trypanosomes then differentiate and multiply in the tsetse fly and are transmitted to another host when the tsetse fly feeds on a new host. Tsetse fly, therefore, serves as 'transporters' and also provide a niche where the trypanosomes can differentiate, multiply and become infective to new hosts. The ability of tsetse flies to acquire trypanosomes, favor their maturation, and transmit them to new hosts depends on factors inherent of the tsetse flies, trypanosome species, and other factors (Geiger *et al.*, 2015). Studies on tsetse fly have reported that infection of midgut with trypanosomes does not always result in mature infections (Moloo *et al*., 1986; Dukes *et al.*, 1989; Maudlin and Welbum, 1994),

Tsetse flies are hematophagous and survive almost solely on blood. As has been reported in mosquitoes and *Drosophila*, blood-derived factors remain active or become immunologically active in vector arthropods and can affect the insect physiology (Pakpour *et al*., 2013). Tsetse flies also have a host preference such that they prefer some mammalian hosts over others in the same ecosystem. This preference can differ greatly depending on tsetse fly species, wildlife animals present and geographical locations (Omolo *et al.*, 2009; Farikou *et al.*, 2010; Muturi *et al.*, 2011). It is presumed that tsetse flies may originally have been reptile feeders that later adapted to feeding on mammals and warthogs may have enabled them to leave their original habitat and invade the Savannah habitat (Ford, 1970). A review of the natural hosts of 17 tsetse species and subspecies from different regions of Africa concluded that *G. morsitans* prefer warthog as blood meal sources (Moloo, 1993). However, the *G. morsitans*, found mainly in woodland Savanna, has been shown to change behavior and feed on domestic animals, especially cattle, in the absence of wild animals. This has been reported in Sudan and Nigeria (FAO, 1982). Contrastingly, another study concluded that *G. morsitans* does not easily change its habitats to feed on man and domestic animals and is

one of the species that disappear when the human population grows (Stephen , 1999). On the other hand, the *G. palpalis* generally inhabit riverine vegetation and are closely associated with reptiles but are opportunistic feeders of mammals including humans (Stephen, 1999). Blood composition and source are therefore important factors that may modulate vector competence, through complex interactions. An earlier study found that delipidation of serum leads to a reduction of the maturation rate of *T. congolense* infections in *G. morsitans*, an effect also observed when complete removal of dietary serum from pig or cattle blood (Maudlin *et al.*, 1984). Another study found that blood from different sources resulted in differences in infection rates in *G. m. centralis* and *G. m. morsitans* pointing to the presence of species-specific factors in the blood that affects trypanosome survival (Mihok *et al.*, 1993). Blood meal analysis is an important aspect in the determination of tsetse fly vertebrate hosts, providing important information in the understanding of tsetse fly biology and ecology. Recently, new technological advances have enabled DNA-based methods which provide high specificity and accuracy in the identification of arthropod blood meals. Given that mitochondria contain independent genomes and are numerous in number per cell, mitochondrial genes have been reliable and popular targets for the identification of vertebrate species in arthropod blood meals. Two common mitochondrial gene markers, Cytochrome *b* (cyt*b*) and cytochrome *c* oxidase I (COI), are widely used for blood meal identification of mosquitoes, black flies, tsetse flies and fleas (Kent, 2009). The 12S gene is a small subunit of ribosomal DNA in mitochondrial DNA that is considered to be a conserved gene between taxa thus it has been used for phylogenetic analysis (Matthee and Davis, 2001; Vun *et al.*, 2011). More so, the 12S gene has an evolutionary rate that is almost similar to the evolutionary rate of mitochondrial DNA itself (Rosli *et al.*, 2011). Analysis of such gene targets is possible using recent techniques such as multiplex PCR and high-throughput sequencing enabling multiple amplicons from multiple samples to be analyzed in a single sequencing run in a less laborious and cost-effective approach.

Tsetse flies' natural refractoriness to trypanosome infections has been explained by several physiological factors. A previous study reported that tsetse fly's endogenous microbiome plays a role in the structural integrity of the peritrophic matrix (PM) during larval development (Weiss *et al.*, 2013). The PM which lines the fly midgut serves as a physical barrier that separates luminal contents from immune responsive epithelial cells and plays a role in parasite refractory phenotype during adulthood. A previous study suggests that *Sodalis* and *Wigglesworthia* (both maternally transmitted symbionts of tsetse flies) can modulate trypanosome development (Geiger, Ponton, and Simo, 2015). The artificial elimination of *Wigglesworthia* (an obligate symbiont of tsetse flies) from flies renders them sterile and compromise their immune system development (Weiss and Aksoy, 2011). This leads to increased fly susceptibility to trypanosome infection. Colonization with *S. glossinidius* (another tsetse fly endosymbiont) has also been shown to increase susceptibility for trypanosomes in tsetse flies (Welburn *et al.*, 1993). It has been reported that the ability of establishment of trypanosome in the tsetse fly midgut is significantly linked to the presence of some *S. glossinidius* genotypes (Geiger *et al.*, 2007). It is generally accepted that different *Sodalis* genotypes are associated with differing capacities for trypanosome establishment. Additionally increased density of the symbiont in the fly gut increases the susceptibility of infection with trypanosomes (Cheng and Aksoy, 1999). Another tsetse fly endosymbiont, *Wolbachia*, is maternally transmitted and is found in the reproductive tissues. It has been found to induce male-killing or cytoplasmic incompatibility (CI) in tsetse flies and mosquitos (Saridaki and Bourtzis, 2010; Alam *et al.*, 2011). The male-killing effect manifested was aborted embryos when an infected male is crossed with an uninfected female or a female infected with a different *Wolbachia* strain.

Wild tsetse flies also contain a variety of environmentally-acquired bacteria in their guts. The bacterial composition in tsetse fly gut significantly varies with geographic regions and tsetse

fly species. However, members of the phyla *Actinobacteria Bacteroidetes*, *Firmicuties,* and *Proteobacteria* are found consistently in tsetse species captured in geographically distinct areas (Geiger *et al.*, 2009; Lindh and Lehane, 2011; Aksoy *et al.*, 2014). They account for less than 1% of tsetse's entire enteric microbiota (Aksoy *et al.*, 2014) majority being the endosymbionts. The origin of environmentally-acquired gut bacteria and their physiological function in tsetse fly host remains to be revealed. In this study, high-throughput sequencing was applied to simultaneously detect and analyze trypanosome species, blood meal sources and microbial composition in tsetse flies. Since the transmission cycle of trypanosomiasis is dependent on the diversity and availability of vertebrates, both as blood meals for the vector and as potential hosts such an approach could be applied to study the different blood meal sources in a tsetse-infested area, the trypanosome species prevalence and bacterial microbiome present in tsetse flies, so as to understand tsetse fly ecology and trypanosome epidemiology in a holistic manner.

In this study, DNA metabarcoding approach is applied for simultaneous evaluation for trypanosomes, blood meal sources, and microbes in wild-caught tsetse flies. The main objective of this study is to identify tsetse fly host preference in specific areas and identify possible reservoirs of various trypanosome species in those areas. Another objective is to investigate whether feeding on different vertebrate hosts affects the bacterial microbiome of tsetse flies.

3.3 Materials and Methods

3.3.1 Sample collection

Tsetse flies were caught from two locations: 1) from Kafue national park in Zambia which is infested with only *G. m. centralis*, caught in July 2017 where a total of 85 tsetse flies were collected by handheld traps while flying around a slow-moving vehicle, 2) from Hurungwe game reserve in Zimbabwe caught in April 2014 using Epsilon traps for a previous study where 32 flies were analyzed (Figure 14). Flies collected from Kafue were sorted according to sex. For the flies caught in Zimbabwe, information on tsetse fly sex and species was not available. All flies had been preserved in silica gel then crushed in a smashing machine at 3,000 rpm for 45 sec. Isolation of DNA was done using the DNA Isolation kit for mammalian blood (Roche USA) as per the manufacturer's protocol for the extraction of DNA from buffy coat. The DNA samples were stored at -80ºC until analysis.

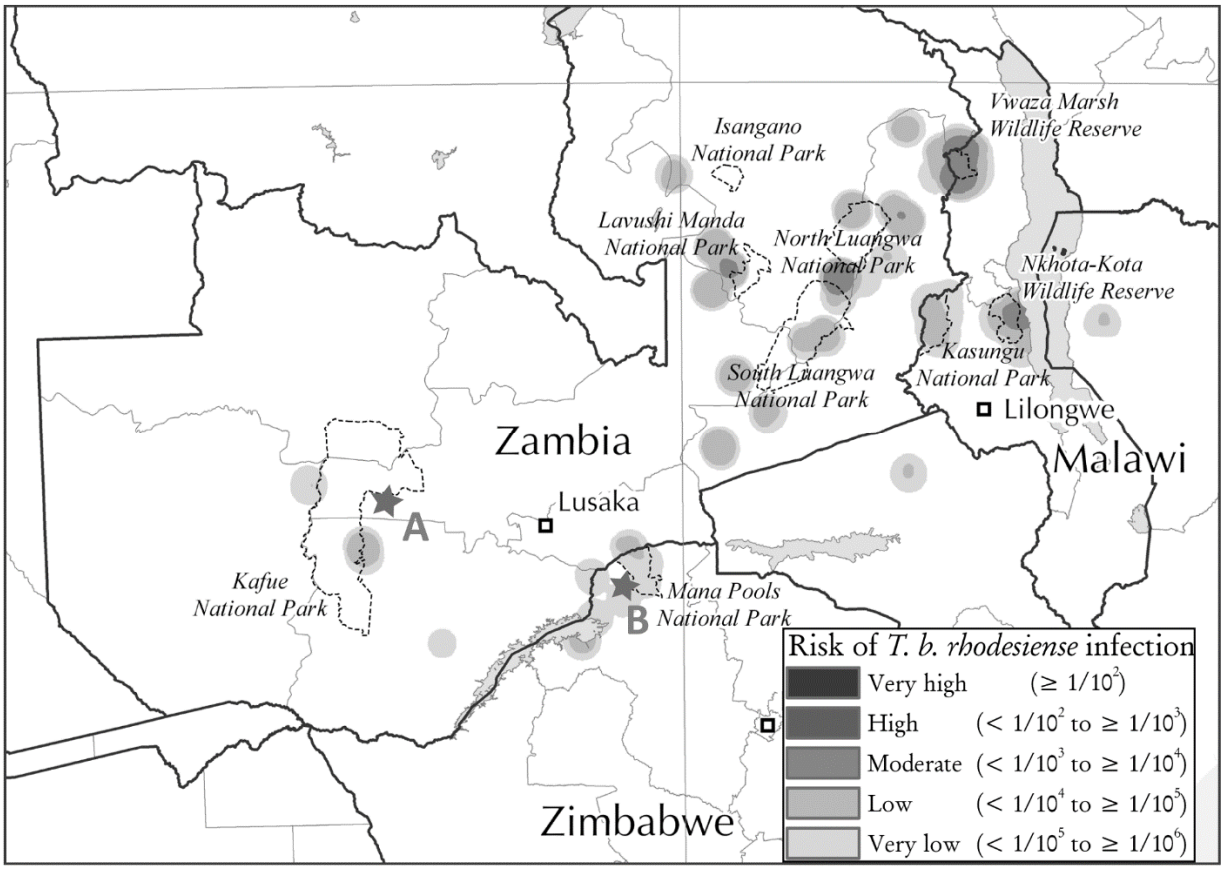

Figure 14. Map showing sample collection areas marked in stars (A) Kafue, Zambia, and (B) Hurungwe, Zimbabwe. The map was obtained from a previous study (Simarro *et al*., 2012) with Open Access permission https://www.plos.org/open-access and modified in Adobe Illustrator v23.0.1.

3.3.2 Laboratory-reared tsetse fly experiment

To determine the duration of residual DNA in a tsetse fly after feeding on blood, 45 teneral flies (flies that have not yet had their first blood meal) that had emerged two days earlier (sourced and maintained at the Yale School of Public health tsetse fly insectary) were fed on sheep blood on the first day of the experiment and transferred to fresh cages to avoid contamination with any blood on the surface of the cage. Each day flies were selected at random and immobilized by placing them on dry ice after which the mid-guts of half of the flies were extracted and the rest of the flies were crushed whole. From the midguts and crushed flies, DNA was extracted using Qiagen's DNeasy Blood and Tissue kit (Qiagen GmbH, Hilden, Germany) following manufacturers' protocol for the extraction of DNA from blood and stored at -20º until analysis. Detection of sheep DNA in all was done using sheep-specific PCR utilizing a sheep-specific reverse primer set (Table 6). The PCR reactions were carried out as specified in a previous study (Bakhoum *et al.*, 2016). The products were examined on 2% agarose gels.

3.3.3 Library construction by 12S and 16S rRNA gene PCR with the dual index for multiplex sequencing

Multiplex PCR was carried out following a dual PCR protocol. The first PCR amplifies the target region while the second PCR is for Illumina-based indexing to allow multiplex sequencing of pooled samples which have a unique pair of index sequences in both terminals. Primers used in this study are listed in Table 6. For blood meal source analysis, mitochondrial 12S ribosomal RNA gene-specific primers were used for the identification of mammalian species. For bacterial community analysis, primers specific to the 16S ribosomal RNA V3-V4 region were used. To identify trypanosome species in infected flies, primers specific to the internal transcribed spacer 1 (ITS1) were applied as described in chapter one.

Table 6. Primers used for PCR analyses.

Reference	Primer's target region	Primer name	Primer sequence (5'-3')
(Ushio *et al*, 2017)	Mitochondrial 12S ribosomal RNA gene	MiMammal-U forward	GGGTTGGTAAATTTCGTGCCAGC
		MiMammal-U reverse	CATAGTGGGGTATCTAATCCCAGTTTG
(Illumina, 2013)	Bacterial V3 and V4 region of 16S ribosomal RNA genes	16S Amplicon PCR forward Primer	CCTACGGGNGGCWGCAG
		16S Amplicon PCR reverse Primer	GACTACHVGGGTATCTAATCC
(Yin *et al.*, 2016)	Sheep specific reverse primer	Forward primer OVS-F	ACACAACTTCTACCACAACCC
		Forward primer OVS-R	AAACAATGAGGGTAACGAGGG
(Gaithuma *et al.*, 2019)This study	Internal transcribed spacer 1 of trypanosome ribosomal RNA genes	AITSF forward Primer	GGCGTGAATAGTACTAGTAGCATGAGGATGA
		AITSF forward Primer	AGGAAGCCAAGTCATCCATC
	Illumina Multiplexing adapter for PE library	Forward primer adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNN[**Forward primer**]
		Reverse primer adapter	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNN[**Reverse primer**]

[ ] the position of attachment of respective primer to the multiplexing adapter

The first PCR was done in 20 µL primary reactions containing 0.5 µL of 10 µM of each of the forward and reverse primers, 10 µL of 2X Ampdirect Plus buffer (Shimadzu, Kyoto, Japan), 0.16 µL of 5 U/µL KAPA Taq polymerase (Kapa Biosystems, Boston, USA), 0.4 µL of DMSO, and 1 µL of sample DNA. Initial denaturation of template DNA was carried out at 95°C for 10 min followed by 37 cycles of 95°C for 30 sec, annealing for 1 min at 60 °C for blood meal PCR, 55 °C for bacterial microbiome PCR and 60 °C for trypanosome ITS1 PCR. An initial extension was done at 72°C for 2 min, and a final extension at 72°C for 10 min. After the first PCR, Agencourt AMPure XP beads (Beckman Coulter, California, USA) were used to purify the

amplicon from free primers and PCR reagents and amplicons inspected on 2% agarose gels. A negative template control was included in each set of PCR reactions.

The second PCR was done in 10 µL reactions containing 1 µL of 10 µM Illumina dual-index primer mix (i5 and i7 primers), 1.2 µL of 25 mM $MgCl_2$, 0.4 µL of 10 mM of dNTP mix, 0.1 µL of 5 U/µL KAPA Taq polymerase (Kapa Biosystems, Boston, USA), 4 µL of 5X buffer, and 2 µL of template (purified amplicon from first PCR diluted 60 times). Denaturation of the template DNA was done at 95°C for 3 min, followed by 11 cycles of 95°C for 30 sec, 61°C for 1 min, an initial extension at 72°C for 2 min, final extension at 72°C for 10 min. To enable sequencing of all amplicons in one run, a unique set of dual index primers was used for each sample for each PCR analyses (blood meal, bacterial 16S, and ITS1). Clean-up of the PCR amplicons was done using Agencourt AMPure XP beads and products were validated by electrophoresis using 2% agarose gels.

3.3.4 Sequencing

Equal volumes of each of the PCR products from the 2nd PCR were pooled into one library which was then purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) and amplicons of similar sizes were pooled together. Quantification of each of the libraries was done using a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The concentration of each library was adjusted to a final concentration of 4 nM using nuclease-free water and applied to the MiSeq platform (Illumina, CA, USA) and sequencing performed using a MiSeq Reagent Kit v3 for 2x300 paired-end. A PhiX DNA spike-in control was added to improve the data quality of low diversity samples. For trypanosome species detection assay, positive template controls comprising an artificial mixture of *T. congolense* and *T. b. rhodesiense* DNA from cultured parasites was also included in the sequencing library.

3.3.5 Read analysis

For trypanosome species identification, ITS1 reads were processed using the workflow outlined in chapter one. For bacterial microbiome assay and mammalian species detection, a modified amplicon toolkit (AMPtk, version 1.2.5) pipeline (Palmer *et al.*, 2018) (Figure 15). Briefly, raw reads were pre-processed by trimming off primers, removal of sequences less than 100 b.p, and merging pair-end reads. Read merging was done using USEARCH for trypanosome species and 16S bacterial microbiome sequence analysis with customized parameters: *fastq_pctid* set to 80, (minimum %id of alignment), *minhsp* set to 8, and *fastq_maxdiffs* set 10 to limit the number of mismatches in the alignment to 10, and *fastq_pctid* set to 300. For blood meal analysis, VSEARCH was used for merging the long overlap region between the paired reads (~270bp), allowing for merging of staggered read pairs. Clustering was done using AMPtk inbuilt DADA2 denoising algorithm, to give amplicon sequence variants (ASVs). Operational taxonomic units (OTUs) were also obtained by clustering ASVs at 97% identity. Downstream processing of ASVs (for ITS1 and bacterial 16S RNA amplicons) or OTUs (blood meal amplicons) was done by filtering out underrepresented sequences that may result from index-bleed where a small percentage of reads bleed into other samples. Only ASVs or OTUs that were represented by more than 10 reads in a sample were considered as a positive outcome. An additional clustering was done using LULU command to remove erroneous molecular ASVs or OTUs. Taxonomy assignment of 12S rRNA amplicons for mammalian species detection was by BLAST (v2.7.0) (Camacho *et al.*, 2009) with custom options (Figure 15) The taxonomy of an OTU was assigned from the top hit subject sequence with a sequence identity of at least 97% and E-value threshold of 10^{-5}. For 16S bacterial microbiome sequence analysis, taxonomy was assigned using SILVA release 132 SSU reference database in qiime2 (https://view.qiime2.org).

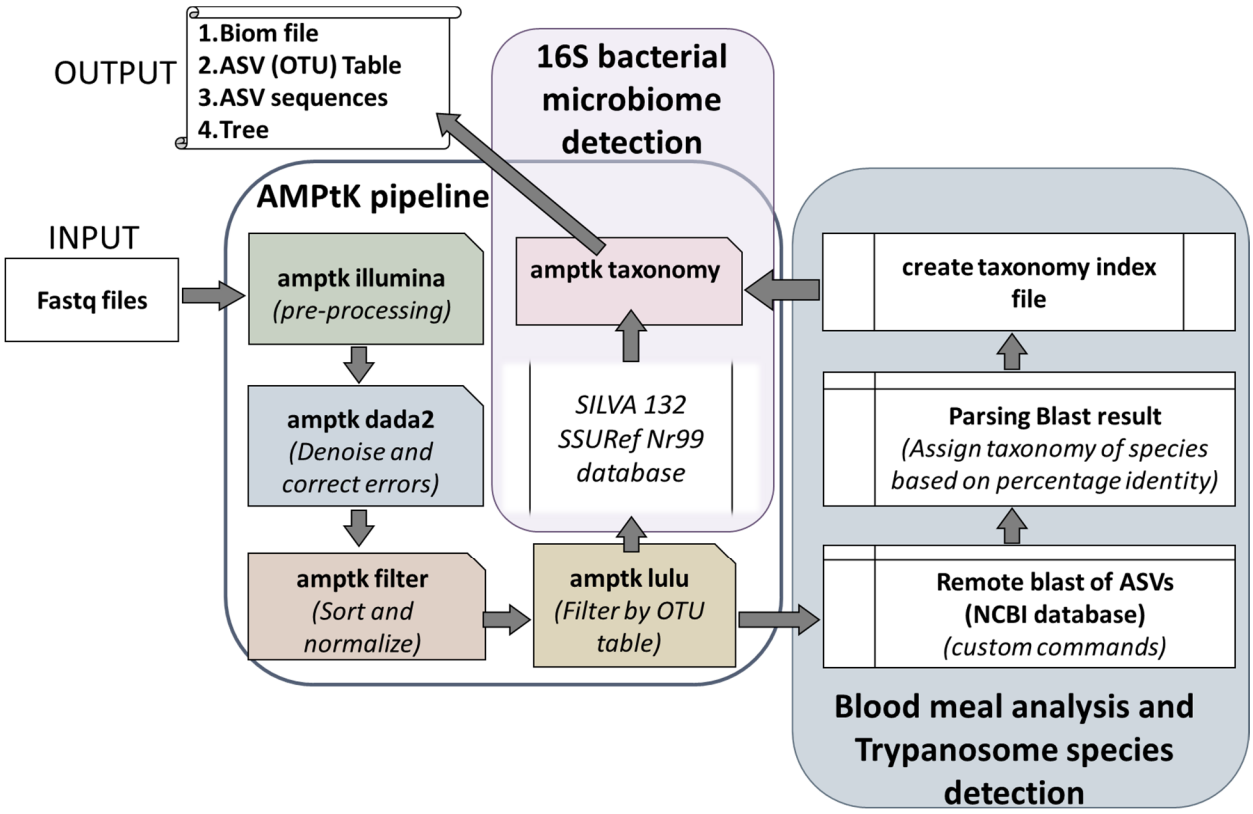

Figure 15. The AMPtk pipeline workflow for read analysis. The raw reads were processed by quality filtering and merging followed by error correction using dada2. The low abundance ASVs were filtered out or merged forming the final OTU table. Taxonomy was then assigned using the blast database for trypanosome species and vertebrate host species and SILVA database for bacteria species.

3.3.6 Data analysis and statistics

Downstream analysis and visualization of plots were done in R v.3.5.1 using Phyloseq package and qiime2. Normalization of the read counts was done by dividing each read count with the total library size. Alpha diversity was estimated using Shannon similarity index and significance between groups tested using the Kruskal–Wallis non-parametric rank test. The shared taxonomic composition across samples was done by beta diversity analysis using Bray-Curtis distance measure (a quantitative measure of community dissimilarity). Abundance plots for all bacterial genera per sample were also generated. Abundance differences of the top three genera were tested using Mann-Whitney unpaired rank test in GraphPad Prism 8 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

3.4 Results

3.4.1 Blood meal host DNA can be detected for a long time after feeding.

Following initial feeding with sheep blood, sheep DNA was detected from all flies (both whole flies and dissected midguts) sacrificed two days post sheep blood feeding (Figure 16). At 5 days post sheep blood feeding, sheep DNA could be amplified from both gut and whole flies in 40% and 50% of the flies respectively (Table 7).

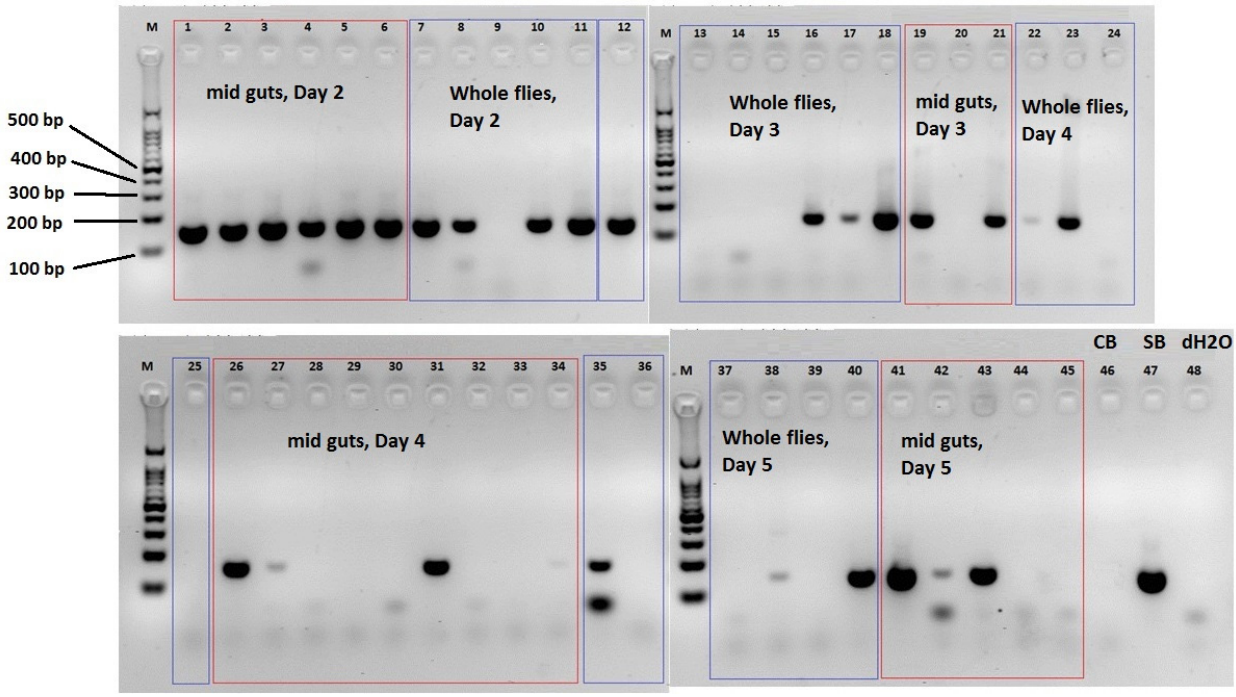

Figure 16. Detection of sheep DNA from whole flies and midguts from tsetse flies fed with sheep blood and maintained with cattle blood. Sheep DNA was detected in at least 5 of flies that were sampled on each day. The expected amplicon was 145 base pairs. M= DNA marker.

Table 7. Persistence of sheep DNA in tsetse flies after initial feeding with sheep blood.

	Days post-feeding			
	2	**3**	**4**	**5**
Gut	6/6	2/3	4/9	3/5
	100%	**66%**	**44%**	**60%**
Whole flies	4/5	4/7	2/4	3/6
	80%	**57%**	**50%**	**50%**

3.4.2 Bats and rats among other mammals are sources of blood meal for wild tsetse flies.

Analysis of the mitochondrial 12S ribosomal RNA revealed a variety of both domestic and wild animals. In flies from Kafue, Zambia (all *G. m. centralis*), DNA from human, cattle, dog, bush pig, African buffalo, warthog, greater kudu, rat, and bat was detected (Table 8) while in flies from Hurungwe, Zimbabwe DNA from waterbuck, bush pig, warthog, and greater kudu was detected. This is the first report of the detection of the bat and rat DNA from wild tsetse flies. Individual files also had DNA from single and multiple mammalian species (Figure 17 A). Majority of the reads obtained from Kafue samples were from human and black rat while in Hurungwe samples, most of the reads were from warthog (Figure 17 B). Of the flies that mammalian DNA was detected, 36% of tsetse flies caught in Kafue, Zambia, and 34% of tsetse flies caught in Hurungwe, Zimbabwe contained DNA from single mammalian species in individual flies compared to 31% of tsetse flies caught in Kafue, Zambia, and 3% of tsetse flies caught in Hurungwe, Zimbabwe flies which contained DNA from multiple mammalian species. The OTUs generated from the blood meal PCR had a high percentage identity to the best hit sequence for the mammalian species assigned (Table 9). Majority of the flies (n=30, 94%) caught in Hurungwe, Zimbabwe had mouse and human DNA detected (Table 9). The presence of mouse DNA in all wild caught tsetse flies is highly unlikely from wild-caught flies. This result was treated as possible contamination with extraneous human and mouse DNA since they had been collected earlier for a previous study and extra precautions to avoid contamination with mammalian DNA. For these samples, any human and mouse DNA reads detected was not considered as a true blood meal source and bacterial microbiome analysis was not done for these samples. However, these samples were analyzed to test our sequence analysis pipeline based on the detection of DNA from wild animals. Extra precautions were applied during the collection of flies in Kafue, to avoid contamination of DNA from extraneous sources. This included wearing clean gloves, use of

sanitized tweezers while handling the flies and processing each fly individually to avoid cross-contamination, ensuring that all flies were processed immediately after being caught.

Table 8. Blood meal sources and trypanosome species identified in tsetse flies collected in Kafue Zambia and tsetse flies collected in Hurungwe Zimbabwe (species undetermined).

Host DNA species detected	Kafue, Zambia (*n*=85)			Hurungwe, Zimbabwe (*n*=32)		
	No. of flies (%)	Trypanosome positive flies	Trypanosome sp. detected in flies	No. of flies (%)	Trypanosome positive flies	Trypanosome sp. detected in flies
African buffalo *(Syncerus caffer)*	2 (2%)	1	*Tg*	-	-	-
African hunting dog *(Lycaon pictus)*	2 (2%)	2	*Ts*	-	-	-
Black rat *(Rattus rattus)*	27 (32%)	9	*Tg, Ts, Tc*	-	-	-
Bush pig *(Potamochoerus larvatus)*	2 (2%)	2	*Tg, Ts, Tc*	1 (3%)	0	-
Cattle *(Bos taurus)*	3 (4%)	3	*Tg, Ts*	-	-	-
Common warthog *(Phacochoerus africanus)*	5 (6%)	3	*Tg, Ts, Tc*	8 (25%)	3	*Tb, Tc, Tv*
Dog *(Canis lupus)*	5 (6%)	2	*Tg*	-	-	-
Greater kudu *(Tragelaphus strepsiceros)*	3 (4%)	1	*Ts*	3 (9%)	1	*Tb, Tc, Tv*
Human *(Homo sapiens)*	43 (51%)	14	*Tg, Ts, Tc, Tv*	*30 (94%)		
Mouse *(Mus musculus)*	-	-	-	*30 (94%)		
Straw-colored fruit bat *(Eidolon helvum)*	2 (2%)	0	-	-	-	-
Waterbuck *(Kobus ellipsiprymnus)*	-	-	-	1 (3%)	0	-

*detection of human and mouse DNA in this sample was considered to be as a result of contamination, *Tb = Trypanosoma brucei*, *Tc = Trypanosoma congolense*, *Ts = Trypanosoma simiae*, *Tg = Trypanosoma godfreyi*, and *Tv = Trypanosoma vivax*.

Table 9. Taxonomy of operational taxonomic units (OTUs) derived from read analysis.

Kafue, Zambia*					
Query coverage	**Percent identity**	**OTU ID**	**Subject Accession number**	**Subject common name**	**Description**
100	100	OTU1	LC422285	Black rat	*Rattus rattus* HS2324 mitochondrial gene for 12S rRNA, partial sequence
100	99.405	OTU6	JQ235547	African buffalo	*Syncerus caffer* isolate 655 mitochondrion, complete genome
100	100	OTU8	KT598692	African hunting dog	*Lycaon pictus* mitochondrion, complete genome
100	97.605	OTU11	GQ338939	Bush pig	*Potamochoerus larvatus* 12S ribosomal RNA gene, partial sequence; mitochondrial
100	100	OTU7	6NF8_A	Cattle	*Bos taurus* Chain A, 28S ribosomal RNA, mitochondria
100	100	OTU4	GQ338942	Common warthog	*Phacochoerus africanus* 12S ribosomal RNA gene, partial sequence; mitochondrial
100	100	OTU5	LC422289	Dog	*Canis lupus familiaris* SDO-17misc-63 mitochondrial gene for 12S rRNA, partial sequence
100	99.405	OTU3	JN632708	greater kudu	*Tragelaphus strepsiceros* isolate PHC11 mitochondrion, complete genome
100	100	OTU2	KY751400	Human	*Homo sapiens* mitochondrion, complete genome
100	100	OTU9	FN674024	Human	*Homo sapiens* nuclear sequence of mitochondrial origin
100	100	OTU13	AC090204	Human	*Homo sapiens* chromosome 8, clone RP11-11N9, complete sequence
100	100	OTU15	FN673842	Human	*Homo sapien*s nuclear sequence of mitochondrial origin
100	100	OTU16	AL590407	Human	Human DNA sequence from clone RP11-522L3 on chromosome X, complete sequence
100	99.383	OTU35	AL954650	Human	Human DNA sequence from clone RP11-523M19 on chromosome 1, complete sequence
100	98.837	OTU12	JN398171	Straw-colored fruit bat	*Eidolon helvum* 12S ribosomal RNA gene, partial sequence; tRNA-Val gene, complete sequence; and 16S ribosomal RNA gene, partial sequence; mitochondrial
Hurungwe, Zimbabwe¶					
100	97.605	OTU4	GQ338939	Bush pig	*Potamochoerus larvatus* 12S ribosomal RNA gene, partial sequence; mitochondrial
100	100	OTU3	GQ338942	Common warthog	*Phacochoerus africanus* 12S ribosomal RNA gene, partial sequence; mitochondrial
100	99.405	OTU11	JN632708	Greater kudu	*Tragelaphus strepsiceros* isolate PHC11 mitochondrion, complete genome
100	98.81	OTU7	JN632651	Waterbuck	*Kobus ellipsiprymnus* isolate Niger mitochondrion, complete genome

* Only OTUs represented by more than 10 reads in a sample were considered valid.

¶ Only OTUs represented by more than 10 reads in a sample were considered and human and mouse OTUs were considered as contamination and removed from the analysis.

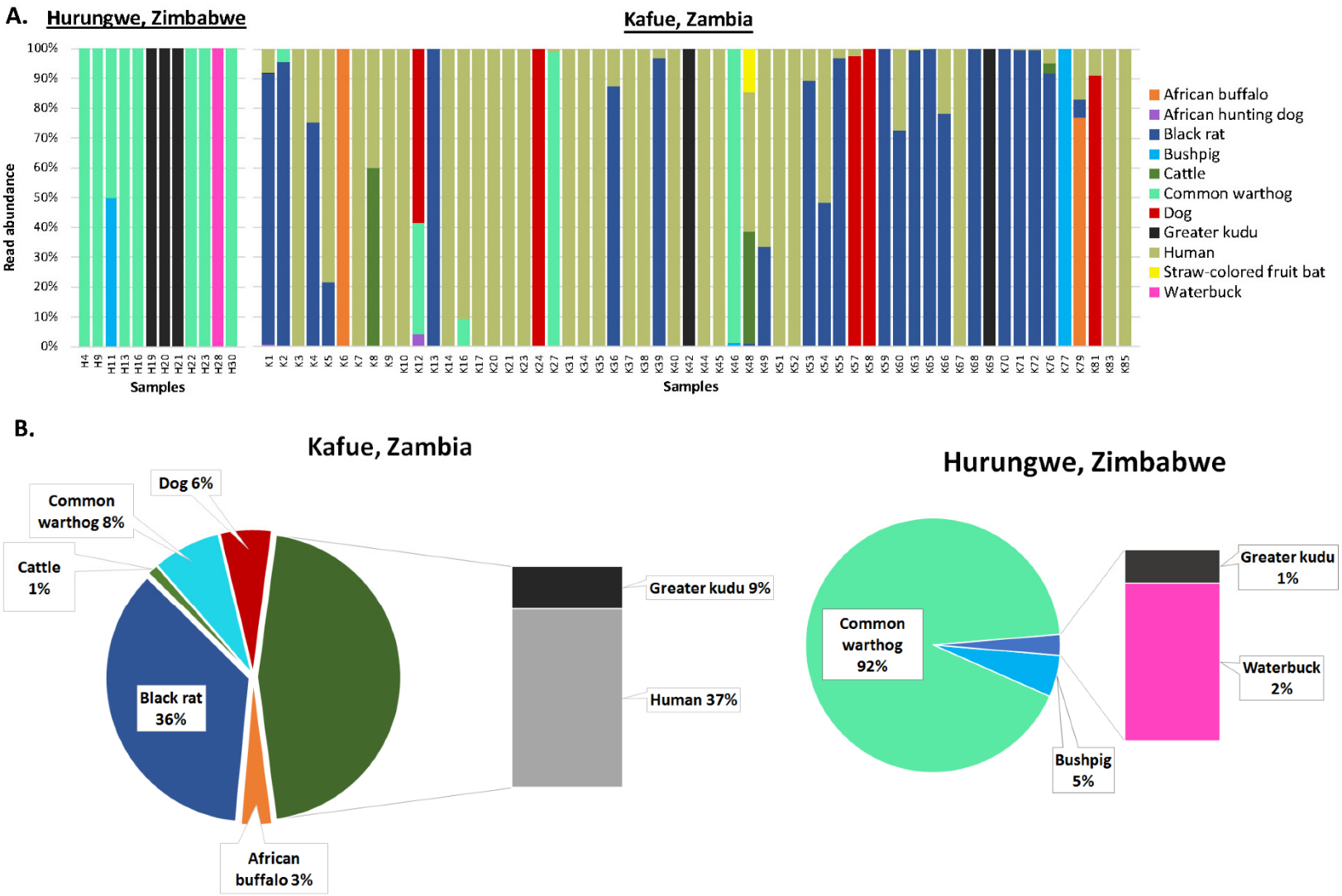

Figure 17. Blood meal sources identified in individual wild tsetse flies caught in Kafue, Zambia, and Hurungwe, Zimbabwe. (A) Proportions of reads and (B) Proportion of the number of reads of each detected mammalian species in Kafue and Hurungwe samples.

3.4.3 Tsetse flies have a limited diversity of the bacterial microbiome

Analysis of the bacterial microbiome in wild-caught tsetse flies caught in the Kafue area (*n*=85) revealed a limited diversity of bacterial communities as shown by alpha diversity analysis with Shannon index (Figure 18). There was no significant difference in diversity observed across fly groups (mixed/single/no blood meal host, trypanosome infected/non-infected flies). However, there was a significant difference ($P<0.05$) in alpha diversity between female and male flies. This difference was due to read an abundance of *Wolbachia* in male and female flies (Figure 19). Beta diversity analysis by non-metric Multi-Dimensional Scaling (NMDS) showed no distinctive bacterial species abundance for any particular group of flies (Figure 20). This data also indicated that the majority of the bacteria detected are tsetse fly endosymbionts: *Wigglesworthia, Wolbachia* and *Sodalis*.

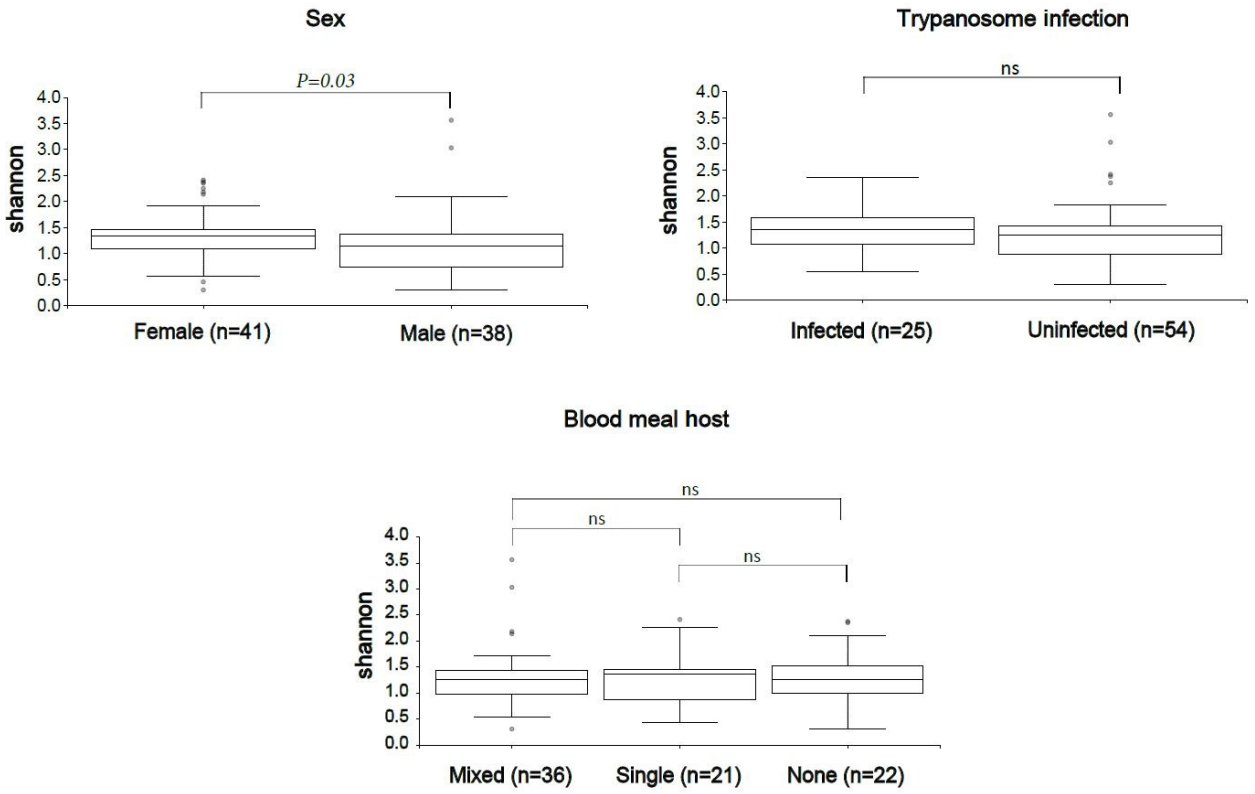

Figure 18. Plots showing alpha diversity of tsetse fly bacterial microbiome diversity estimated Shannon similarity index across different groups. The line in the middle of each bar represents the mean and outer lines represent the upper and lower limits. Dots represent samples outside the limits. (*ns= not significant*, Kruskal–Wallis non-parametric rank test).

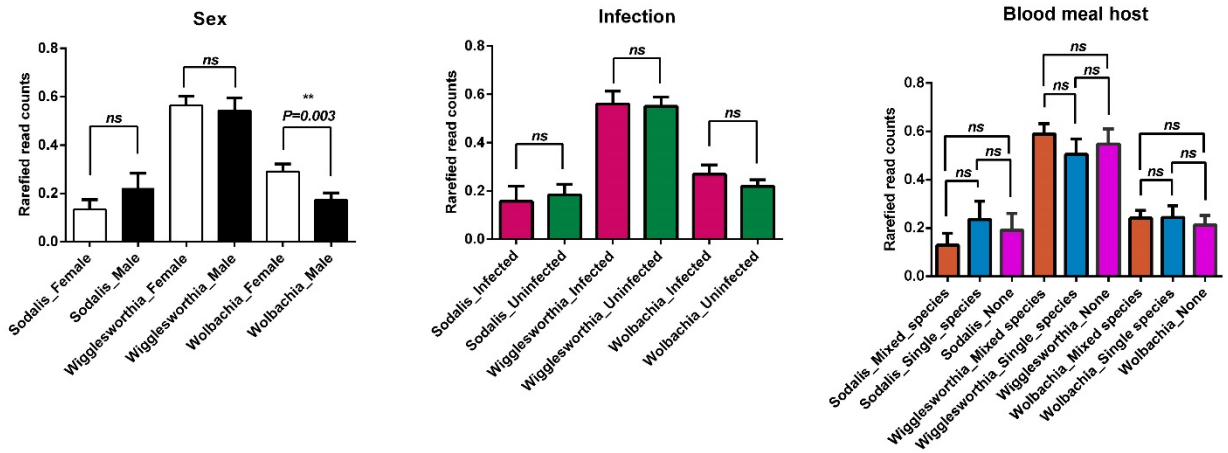

Figure 19. Plots showing the abundance of *Sodalis*, *Wolbachia,* and *Wigglesworthia*. Rarefied read counts obtained for the genus *Sodalis*, *Wolbachia* and *Wigglesworthia* are plotted for each of the groups are (A)female and male flies, (B) trypanosome infected and uninfected flies, and (C) flies in which blood meal host was detected or not detected. (* indicates significance *P<0.05,* Mann-Whitney unpaired rank test).

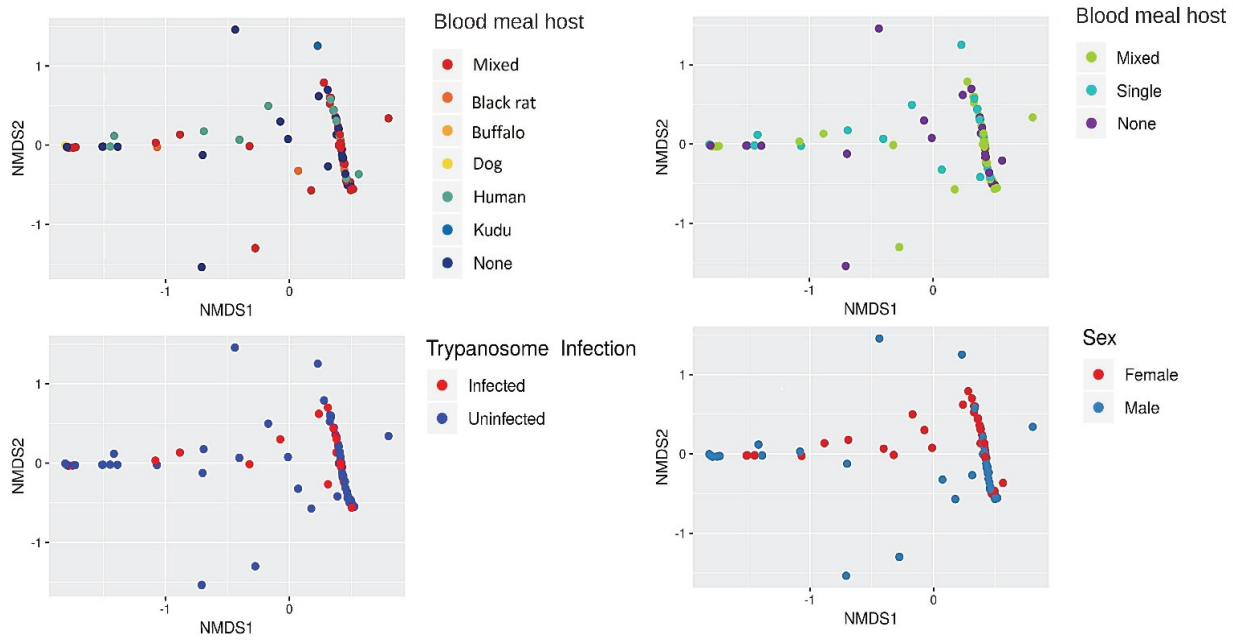

Figure 20. Non-metric multi-dimensional scaling diversity (NMDS) ordination plots showing beta diversity analysis plots for different groups of flies. Each dot represents a sample from an individual fly. The colors represent the respective group variables.

Most of the non-symbiotic bacteria were detected in the trypanosome negative flies which included *Yersinia, Staphylococcus, Pseudomonas, Lactobacillus, Micrococcus, Sphingomonas, Nosocomiicoccus,* and *Corynebacterium* among others (Figure 21). Some soil and other environmentally derived bacteria were also identified, e.g. *Tumebacillus*, *Novosphingobium*, and *Rhizobium* pointing to the likelihood of contamination with bacteria from the environment.

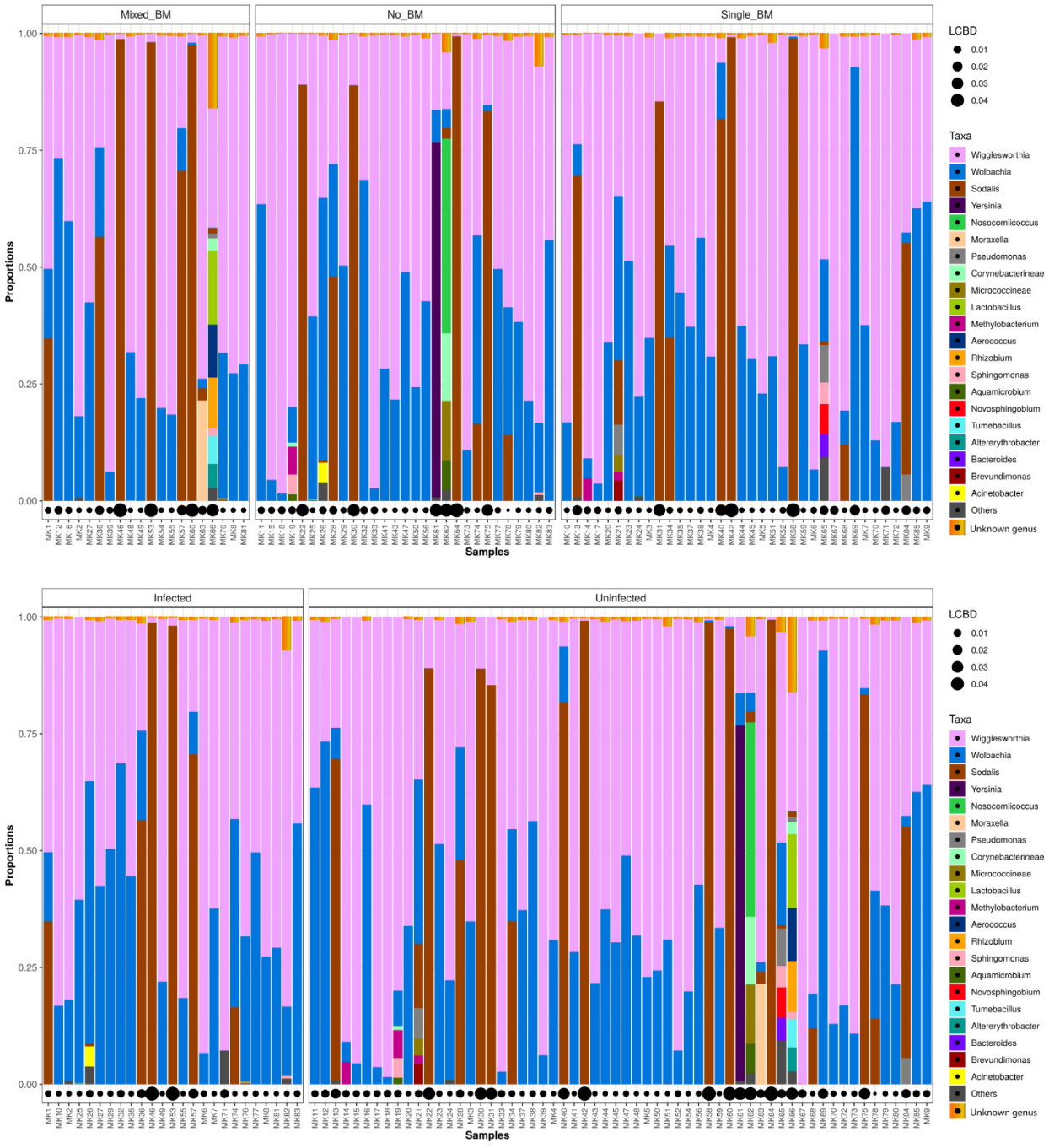

Figure 21. Top twenty most abundant bacteria (genus level). Top: samples grouped according to blood meal sources. Bottom: samples grouped according to trypanosome infection. Unknown genus represent sequences that could not be taxonomically identified at genus level and others represent the sum of all the other genus identified. *LCBD= Local contribution to biodiversity*.

3.5 Discussion

The results obtained in this study showed the broad range of mammalian hosts that tsetse flies feed on in a small geographic area. The flies caught in this study were collected in areas that interface human settlements and areas set aside for wild animals. The preferred host for tsetse flies caught in Hurungwe, Zimbabwe was warthog agreeing with a study reviewing of 47, 697 blood meal analyses from 38 African countries for the period 1953 to 1991, which found that *G. morsitans* generally prefer warthog as blood meal sources (Moloo, 1993). However, flies from Kafue, Zambia (*G. m. centralis*) showed preference to the human host. This could be explained by the fact that the flies were collected in an area located in the interphase of game reserve and human settlement area further supporting the submission that tsetse flies are able to change feeding preferences in the absence of wild animals or the presence of an alternative host (Clausen *et al.*, 1998; Stephen, 1999). Results on the wildlife animals identified in this study i.e. warthog, bush pig, African buffalo, African hunting dog, and greater kudu in agreement with previous studies (Clausen *et al.*, 1998) including observations made in Luangwa Valley, Zambia that reported similar mammalian hosts. This indicates that wild animals (mainly found in woodland Savannah and bushy areas) are a major source of blood meal for *G. m. centralis* in Kafue, Zambia, and Hurungwe, Zimbabwe. Recently, a human case of trypanosomiasis was reported in Kafue area (an old trypanosomiasis focus) indicating that *T. b. rhodesiense* is still in circulation in the area (Squarre *et al.*, 2016). A major finding from Kafue was the detection of domestic animals (dog and cattle) as blood meal sources with some of the flies infected with trypanosomes. This suggests that domestic animals in the area are at risk of animal trypanosomiasis. This is in agreement with a study that found canine African trypanosomiasis particularly in indigenous dogs (Lisulo *et al.*, 2014). Moreover, this finding could help narrow down on possible reservoirs of the trypanosomes for control measures. It was interesting to identify bats and rats as blood meal sources because such

mammals have not been given focus on their importance in the ecology of the tsetse fly or distribution of trypanosomes. The finding that rats and bats were possible blood meal sources of tsetse flies poses the necessities to investigate more on the trypanosome species in bats since they are known to hosts different species of trypanosomes (Lima *et al.*, 2013; Schwan *et al.*, 2016).

Since tsetse flies are obligate blood feeders for their nutrients, it still remains unknown how bacteria from extraneous sources affects the tsetse fly. A prominent limitation of 16S ribosomal RNA sequence analysis is that the gene exists in a single copy in some bacteria, but in multiple (sometimes more than 10) copies in other species so that this number is not truly a per-genome estimate thus interpretation of abundance analysis is difficult. Nevertheless, in this study, it was observed that abundance of *Wolbachia* was significantly higher in female flies. This result is in line with the finding that *Wolbachia* is known to be maternally transmitted into tsetse's intrauterine larva through mother's milk secretions (Attardo *et al.*, 2008). Although it has been reported that *Sodalis glossinidius* plays a role in tsetse fly refractoriness to trypanosome infection and/or the establishment, no significant association with *Sodalis* with trypanosome infection was observed which may be due to the detection of trypanosome in the midgut that could result from engorging infected blood and not successful establishment of trypanosome infections. The prevalence of *Sodalis* in the flies caught in Kafue was higher (27%) than that reported in a previous study (15.9%) on *G. m centralis* in western Zambia (Mbewe *et al.*, 2015). However, this could be due to the high sensitivity of high-throughput sequencing compared to conventional PCR. The prevalence of *Wigglesworthia* was 100% as expected since they are obligate symbionts of tsetse flies representing high abundance in individual flies compared to other genera. Results on the comparison of microbial composition in different groups show that there was no particular bacteria associated with blood meal source, sex or status of trypanosome infection agreeing with a previous study (Jacob *et al.*, 2017). Results on microbial evenness among groups showed that *Wolbachia*

was more in female flies compared to male flies. This study had some limitations in such as the low number of trypanosome infected flies as well as low and uneven numbers in groups of flies with mixed, single or no blood meal detected. However, the bacteria species that were detected in this study have been reported in previous studies (Aksoy *et al.*, 2014; Malele *et al.*, 2018). Results from this study suggest that blood meals from different hosts have little effect on tsetse fly bacterial microbiome. However, the high abundance of endosymbiotic bacteria makes it difficult to accurately analyze other bacteria present. Studies on other arthropod vectors have shown that bacterial microbiome is affected by blood meal from different hosts e.g. a recent study on ticks reported host blood meal-driven reduction of tick microbiome diversity where a blood meal from the Western fence lizards (a host that is refractory to the Lyme disease pathogen) significantly reduced microbiome diversity in ticks (Swei and Kwan, 2017). This has also been demonstrated recently in mosquitos (*Aedes aegypti*) (Muturi *et al.*, 2019). Most of the bacterial species identified in this study have been reported in other studies with similar findings that limited bacterial diversity in wild-caught tsetse flies (Aksoy *et al.*, 2014; Griffith *et al.*, 2018). Future studies need to focus on the accurate detection and analysis of non-symbiotic bacteria in tsetse flies to shed light on the relationship between bacterial microbiome, trypanosome infection, blood meal source and other factors and their importance in tsetse fly vector competency, immunity, and reproduction.

4 General conclusions

African trypanosomiasis is still considered a neglected disease that affects low-income populations in Sub-Saharan Africa. In order to control the spread of the disease, application of different strategies are needed. West African HAT which represents 98% of the reported cases can be effectively controlled by treatment of infected humans and tsetse fly vector control because humans are the major players in the maintenance of *T. b. gambiense*, the sub-species that caused the disease (WHO, 2018). On the other hand, East African HAT is considered a zoonosis and although few cases are reported, it is a major risk to human populations in endemic areas because the *Trypanosoma brucei rhodesiense* (responsible for this disease) is maintained in wild and domestic animals and therefore considered impossible to completely eliminate. The importance of identification of animal reservoirs necessitates the need to develop new methods for trypanosome species detection that are cost-effective, highly sensitive and field friendly. Due to the difficulty of surveillance of infections in wild animals, the studies focusing on detection of tsetse fly vertebrate hosts can help identify the possible animal reservoirs of trypanosomes in endemic areas. Epidemiology studies seek to generate data on the prevalence of trypanosomiasis in focus areas leading to customized control strategies thus there is a growing need to develop better tools that are sensitive and field friendly. Such tools will help in filling knowledge gaps in the evolution of trypanosomes and understanding the contribution of different trypanosome species and genotypes to human and livestock trypanosomiasis.

The most effective way to control the spread of HAT has been identified as tsetse fly control (WHO, 1998). However, this control method is expensive, requires a lot of human resources and implementation in large scale requiring inter-government cooperation. Besides, previously successful programs succeeded in eliminating tsetse flies in focus areas but re-infestation occurred

once the programs were stopped running out of funds (Fèvre *et al.*, 2005). The lack of effective and affordable vaccines and resistance of tsetse flies to insecticides severely limits trypanosomiasis control. New control strategies that inhibit trypanosome maturation within tsetse fly vectors are needed. Tsetse flies' enteric microbiota is increasingly being studied for application in this context, and one such novel strategy could use ' probiotic ' bacteria, the presence of which would alter the physiology of the intestine of the vector to make the environment inhospitable for pathogens (Weiss *et al.*, 2013). African Animal trypanosomiasis is responsible for major economic losses in low-income populations that depend on livestock (Bukachi, Wandibba and Nyamongo, 2017).

The present study was designed to provide simple, sensitive and efficient ways for trypanosome detection that can be applied on a large scale fashion, determine the blood meal hosts from tsetse flies and investigate the dynamics of the bacterial microbiome of tsetse flies using next-generation sequencing. The results of the studies can be summarized as follows;

1. Development of a method that combines PCR and next-generation sequencing for the detection of different trypanosome species and subspecies. Similar to the widely used bacterial metagenomic analysis protocol, this method uses a modular, two-step PCR process followed by sequencing of all amplicons in a single run, making the sequencing of amplicons more efficient and cost-effective when dealing with large sample sizes. This method is more accurate than traditional gel-based analyses which are prone to misidentification of species, is able to discriminate between subspecies of *T. congolense*, *T. simiae*, *T. vivax*, and *T. godfreyi* sub-species and thus has the potential to provide new insights into the epidemiology of different trypanosome genotypes and the discovery of new ones.

2. Novel internal transcribed spacer 1 primers that are optimized for short read sequencing and have slightly better sensitivity than conventional primers. They are more sensitive in the detection of *T. brucei* which includes the human infective parasites.
3. This is the first report of bat DNA detection in wild tsetse flies. Results of this study indicate that humans could be important blood meal sources in Kafue, Zambia. However, more studies are needed to support this. Domestic animals in this area are at high risk of infection with trypanosomes. Additionally wild animals (warthog, kudu waterbuck, and bush pig, buffalo, and African hunting dog), domestic animals (dog and cattle) and small mammals (rat and bat) could be reservoirs of animal-infective trypanosomes in both Kafue, Zambia and Hurungwe, Zimbabwe.
4. Results of this study suggests that different sources of blood meal do not significantly affect the diversity of tsetse flies' bacterial microbiome. However, future studies on the bacterial microbiome in tsetse flies need to factor in the predominance of symbiotic bacteria that make diversity analysis difficult.

The findings in this dissertation will contribute in the characterization of trypanosome species-specific prevalence, understanding trypanosome evolution, identification of possible animal reservoirs of trypanosomes in an area and deducing important variables and their effects on tsetse fly biology and ecology.

5 References

Adams, E. R., Hamilton, P. B. and Gibson, W. C. (2010) 'African trypanosomes: Celebrating diversity', *Trends in Parasitology*. Elsevier Ltd, 26(7), pp. 324–328.

Adams, E. R., Hamilton, P. B., Malele, I. I. and Gibson, W. C. (2008) 'The identification, diversity and prevalence of trypanosomes in field caught tsetse in Tanzania using ITS-1 primers and fluorescent fragment length barcoding', *Infection, Genetics and Evolution*. School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK. emily.adams@bristol.ac.uk, 8(4), pp. 439–444.

Aksoy, E., Telleria, E. L., Echodu, R., Wu, Y., Okedi, L. M., Weiss, B. L., Aksoy, S. and Caccone, A. (2014) 'Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota', *Applied and Environmental Microbiology*, 80(14), pp. 4301–4312.

Alam, U., Medlock, J., Brelsfoard, C., Pais, R., Lohs, C., Balmand, S., Carnogursky, J., Heddi, A., Takac, P., Galvani, A. and Aksoy, S. (2011) '*Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans*', *PLoS Pathogens*, 7(12). e1002415.

Attardo, G. M., Lohs, C., Heddi, A., Alam, U. H., Yildirim, S. and Aksoy, S. (2008) 'Analysis of milk gland structure and function in *Glossina morsitans*: Milk protein production, symbiont populations and fecundity', *Journal of Insect Physiology*, 54(8), pp. 1236–1242.

Auty, H., Anderson, N. E., Picozzi, K., Lembo, T., Mubanga, J., Hoare, R., Fyumagwa, R. D., Mable, B., Hamill, L., Cleaveland, S. and Welburn, S. C. (2012) 'Trypanosome diversity in wildlife species from the Serengeti and Luangwa valley ecosystems', *PLoS Neglected Tropical Diseases*, 6(10). e1828.

Bakhoum, M. T., Fall, M., Seck, M. T., Gardès, L., Fall, A. G., Diop, M., Mall, I., Balenghien, T., Baldet, T., Gimonneau, G., Garros, C. and Bouyer, J. (2016) 'Foraging range of arthropods with veterinary interest: New insights for Afrotropical Culicoides biting midges (Diptera: *Ceratopogonidae*) using the ring method', *Acta Tropica*, 157, pp. 59–67.

Barbosa, A. D., Gofton, A. W., Paparini, A., Codello, A., Greay, T., Gillett, A., Warren, K., Irwin, P. and Ryan, U. (2017) 'Increased genetic diversity and prevalence of co-infection with *Trypanosoma* spp. in koalas (*Phascolarctos cinereus*) and their ticks identified using next-generation sequencing (NGS)', *PLoS ONE*. Edited by V. Yurchenko. 12(7). e0181279.

Barry, J. D. and McCulloch, R. (2001) 'Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite.', *Advances in parasitology*, 49, pp. 1–70.

Berrang-Ford, L., Waltner-Toews, D., Charron, D., Odiit, M., McDermott, J. and Smit, B. (2005) 'Sleeping sickness in southeastern Uganda: A systems approach', *EcoHealth*, 2(3), pp. 183–194.

Bonnet, J., Boudot, C. and Courtioux, B. (2015) 'Overview of the diagnostic methods used in the field for Human African Trypanosomiasis: What could change in the next years?', *BioMed Research International*, 2015, pp. 1–10.

Bouyer, J., Bouyer, F., Donadeu, M., Rowan, T. and Napier, G. (2013) 'Community - and farmer-based management of animal African trypanosomosis in cattle', *Trends in Parasitology*,

pp. 519–522.

Broderick, N. A., Raffa, K. F., Goodman, R. M. and Handelsman, J. (2004) 'Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods', *Applied and Environmental Microbiology*, 70(1), pp. 293–300.

Brun, R. and Blum, J. (2012) 'Human African trypanosomiasis', *Infect.Dis.Clin.North Am.*, 26(2), pp. 261–273.

Bukachi, S. A., Wandibba, S. and Nyamongo, I. K. (2017) 'The socio-economic burden of human African trypanosomiasis and the coping strategies of households in the South Western Kenya foci', *PLoS Neglected Tropical Diseases*. Edited by O. B. Akogun. 11(10). e0006002.

Büscher, P. (2005) 'Diagnosis of Human and Animal African Trypanosomiasis', in *The African Trypanosomes*, pp. 51–63.

Büscher, P., Cecchi, G., Jamonneau, V. and Priotto, G. (2017) 'Human African trypanosomiasis', *The Lancet*. Elsevier, 390(10110), pp. 2397–2409.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*, 10, pp. 1–9.

Chandler, J. A., Lang, J., Bhatnagar, S., Eisen, J. A. and Kopp, A. (2011) 'Bacterial communities of diverse *Drosophila* species: Ecological context of a host-microbe model system', *PLoS Genetics*, 7(9). e15537404.

Cheng, Q. and Aksoy, S. (1999) 'Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies', *Insect Molecular Biology*, 8(1), pp. 125–132.

Clausen, P. H., Adeyemi, I., Bauer, B., Breloeer, M., Salchow, F. and Staak, C. (1998) 'Host preferences of tsetse (Diptera: *Glossinidae*) based on bloodmeal identifications', *Medical and Veterinary Entomology*, 12(2), pp. 169–180.

Cox, A., Tilley, A., McOdimba, F., Fyfe, J., Eisler, M., Hide, G. and Welburn, S. (2005) 'A PCR based assay for detection and differentiation of African trypanosome species in blood', *Experimental Parasitology*, 111(1), pp. 24–29.

Cuypers, B., Van Den Broeck, F., Van Reet, N., Meehan, C. J., Cauchard, J., Wilkes, J. M., Claes, F., Goddeeris, B., Birhanu, H., Dujardin, J. C., Laukens, K., Büscher, P. and Deborggraeve, S. (2017) 'Genome-wide SNP analysis reveals distinct origins of *Trypanosoma evansi* and *Trypanosoma equiperdum*', *Genome Biology and Evolution*, 9(8), pp. 1990–1997.

Dekkera, J. P. (2018) 'Metagenomics for clinical infectious disease diagnostics steps closer to reality', *Journal of Clinical Microbiology,* 56 (9) pp. 1–70.

Desquesnes, M. and M.R. Dávila, A. (2002) 'Applications of PCR-based tools for detection and identification of animal trypanosomes: A review and perspectives', *Veterinary Parasitology*, 109(3–4), pp. 213–231.

Desquesnes, M., Holzmuller, P., Lai, D. H., Dargantes, A., Lun, Z. R. and Jittaplapong, S. (2013) '*Trypanosoma evansi* and surra: A review and perspectives on origin, history, distribution, taxonomy, morphology, hosts, and pathogenic effects', *BioMed Research International,* 2013, pp. 176-194.

Desquesnes, M., McLaughlin, G., Zoungrana, A., Dávila, A. M. R. and Davila, A. M. (2001) 'Detection and identification of *Trypanosoma* of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA', *International Journal for Parasitology*, 31(5–6), pp. 610–614.

Dukes, P., Gibson, W. C., Gashumba, J. K., Hudson, K. M., Bromidge, T. J., Kaukus, A., Asonganyi, T. and Magnus, E. (1992) 'Absence of the LiTat 1.3 (CATT antigen) gene in *Trypanosoma brucei gambiense* stocks from Cameroon', *Acta Tropica*, 51(2), pp. 123–134.

Dukes, P., Kaukas, A., Hudson, K. M., Asonganyi, T. and Gashumba, J. K. (1989) 'A new method for isolating *Trypanosoma brucei gambiense* from sleeping sickness patients', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 83(5), pp. 636–639.

Farikou, O., Njiokou, F., Simo, G., Asonganyi, T., Cuny, G. and Geiger, A. (2010) 'Tsetse fly blood meal modification and trypanosome identification in two sleeping sickness foci in the forest of southern Cameroon', *Acta Tropica*, 116(1), pp. 81–88.

Fèvre, E. M., Picozzi, K., Fyfe, J., Waiswa, C., Odiit, M., Coleman, P. G. and Welburn, S. C. (2005) 'A burgeoning epidemic of sleeping sickness in Uganda', *Lancet*, 366(9487), pp. 745–747.

Food and Agriculture Organization (FAO) (1982) 'Training manual for tsetse control personnel; Ecology and behaviour of tsetse', *FAO*, Volume 2.

Ford, J. (1970) 'The geographical distribution of *Glossina*', in *In: Mulligan, H.W. (ed.) The African Trypanosomiases*. London: George Allen and Unwin, pp. 274–297.

Frampton, M. and Houlston, R. (2012) 'Generation of artificial FASTQ files to evaluate the performance of next-generation sequencing pipelines', *PLoS ONE*, 7(11), pp. 193-208.

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C. and Hansen, A. J. (2017) 'Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates', *Nature Communications*. Springer US, 8(1), pp. 17-28.

Garcia, H. A., Rodrigues, C. M. F., Rodrigues, A. C., Pereira, D. L., Pereira, C. L., Camargo, E. P., Hamilton, P. B. and Teixeira, M. M. G. (2018) 'Remarkable richness of trypanosomes in tsetse flies (*Glossina morsitans morsitans* and *Glossina pallidipes*) from the Gorongosa National Park and Niassa National Reserve of Mozambique revealed by fluorescent fragment length barcoding (FFLB)', *Infection, Genetics and Evolution*, 63, pp. 370–379.

Gardiner, P. R. and Mahmoud, M. M. (2012) 'Salivarian trypanosomes causing disease in livestock outside sub-Saharan Africa', in *Parasitic Protozoa: Second Edition*, pp. 277–314.

Gardner, S. N. and Slezak, T. (2014) 'Simulate_PCR for amplicon prediction and annotation from multiplex, degenerate primers and probes', *BMC Bioinformatics*, 15(1), pp. 2–7.

Garside, L. H. and Gibson, W. C. (1995) 'Molecular characterization of trypanosome species and subgroups within subgenus Nannomonas', *Parasitology*. 111(Pt 3), pp. 301–12.

Geiger, A., Fardeau, M. L., Grebaut, P., Vatunga, G., Josénando, T., Herder, S., Cuny, G., Truc, P. and Ollivier, B. (2009) 'First isolation of *Enterobacter*, *Enterococcus*, and *Acinetobacter* spp. as inhabitants of the tsetse fly (*Glossina palpalis palpalis*) midgut', *Infection, Genetics and Evolution*, 9(6), pp. 1364–1370.

Geiger, A., Ponton, F. and Simo, G. (2015) 'Adult blood-feeding tsetse flies, trypanosomes, microbiota and the fluctuating environment in sub-Saharan Africa', *ISME Journal*, 9(7), pp. 1496–1507.

Geiger, A., Ravel, S., Mateille, T., Janelle, J., Patrel, D., Cuny, G. and Frutos, R. (2007) 'Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*', *Molecular Biology and Evolution*, 24(1), pp. 102–109.

Gibson, W. (2003) 'Species concepts for trypanosomes : from morphological to molecular', *Kinetoplastid Biology and Disease*, 6, pp. 1–6.

Gibson, W. (2007) Resolution of the species problem in African trypanosomes, *International Journal for Parasitology*, 37, pp. 829–838

Gibson, W. C., Stevens, J. R., Mwendia, C. M. T., Makumi, J. N., Ngotho, J. M. and Ndung'u, J. M. (2001) 'Unravelling the phylogenetic relationships of African trypanosomes of suids', *Parasitology*, 122(6), pp. 625–631.

Grébaut, P., Melachio, T., Nyangmang, S., Eyenga, V. E. o., Njitchouang, G. R., Ofon, E., Njiokou, F. and Simo, G. (2016) 'Xenomonitoring of sleeping sickness transmission in Campo (Cameroon)', *Parasites and Vectors*, 9(1), pp. 1–9.

Griffing S. M., Mixson-Hayden T., Sridaran S., Alam M. T., McCollum A. M., Cabezas C., Marquiño Quezada W., Barnwell J. W., Macedo De Oliveira A., Lucas C., Arrospide N., Escalante A. A., Bacon D. J. and Udhayakumar V. (2011) 'South American Plasmodium falciparum after the Malaria Eradication Era: Clonal Population Expansion and Survival of the Fittest Hybrids', PLOS ONE. Public Library of Science, 6(9), e23486.

Griffith, B. C., Weiss, B. L., Aksoy, E., Mireji, P. O., Auma, J. E., Wamwiri, F. N., Echodu, R., Murilla, G. and Aksoy, S. (2018) 'Analysis of the gut-specific microbiome from field-captured tsetse flies, and its potential relevance to host trypanosome vector competence', *BMC Microbiology*, 18(1), pp. 146.

Haag, J., O'hUigin, C. and Overath, P. (1998) 'The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria', *Molecular and Biochemical Parasitology*, 91. pp 37-49.

Hamilton, P. B., Adams, E. R., Malele, I. I. and Gibson, W. C. (2008) 'A novel, high-throughput technique for species identification reveals a new species of tsetse-transmitted trypanosome related to the *Trypanosoma brucei* subgenus, *Trypanozoon*', *Infection, Genetics and Evolution*, 8(1), pp. 26–33.

Illumina (2013) '16S Metagenomic Sequencing Library Preparation', *Illumina.com*, (B), pp. 1–28.

Isaac, C., Ciosi, M., Hamilton, A., Scullion, K. M., Dede, P., Igbinosa, I. B., Nmorsi, O. P. G., Masiga, D. and Turner, C. M. R. (2016) 'Molecular identification of different trypanosome species and subspecies in tsetse flies of northern Nigeria', *Parasites & Vectors*, 9(1), pp. 301.

Jacob, F., Melachio, T. T., Njitchouang, G. R., Gimonneau, G., Njiokou, F., Abate, L., Christen, R., Reveillaud, J. and Geiger, A. (2017) 'Intestinal bacterial communities of trypanosome-infected and uninfected *Glossina palpalis palpalis* from three human African trypanomiasis foci in Cameroon', *Frontiers in Microbiology*, 8(1464).156-172.

Katoh, K. and Standley, D. M. (2013) 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', *Molecular Biology and Evolution*, 30(4), pp. 772–780.

Kennedy, P. G. E. (2004) 'Human African trypanosomiasis of the CNS: Current issues and challenges', *Journal of Clinical Investigation*, pp. 496–504.

Kennedy, P. G. E. (2008) 'Diagnosing central nervous system trypanosomiasis: two stage or not to stage?', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(4), pp. 306–307.

Kent, R. J. (2009) 'Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies', *Molecular Ecology Resources*, pp. 4–18.

Lai, D.-H., Hashimi, H., Lun, Z.-R., Ayala, F. J. and Lukes, J. (2008) 'Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*', *Proceedings of the National Academy of Sciences*, 105(6), pp. 1999–2004.

Lehane, M. J., Msangi, A. R., Whitaker, C. J. and Lehane, S. M. (2000) 'Grouping of trypanosome species in mixed infections in *Glossina pallidipes*', *Parasitology*, 120(6), pp. 583–592.

Letunic, I. and Bork, P. (2011) 'Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy', *Nucleic Acids Research*, 39(SUPPL. 2), pp. 475–478.

Li, F. J., Gasser, R. B., Lai, D. H., Claes, F., Zhu, X. Q. and Lun, Z. R. (2007) 'PCR approach for the detection of *Trypanosoma brucei* and *T. equiperdum* and their differentiation from *T. evansi* based on maxicircle kinetoplast DNA', *Molecular and Cellular Probes*, 21(1), pp. 1–7.

Lima, L., Espinosa-Álvarez, O., Hamilton, P. B., Neves, L., Takata, C. S., Campaner, M., Attias, M., De Souza, W., Camargo, E. P. and Teixeira, M. M. (2013) '*Trypanosoma livingstonei*: A new species from African bats supports the bat seeding hypothesis for the *Trypanosoma cruzi* clade', *Parasites and Vectors*, 6(1), pp. 1–17.

Lindh, J. M. and Lehane, M. J. (2011) 'The tsetse fly *Glossina fuscipes fuscipes* (Diptera: *Glossina*) harbours a surprising diversity of bacteria other than symbionts', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 99(3), pp. 711–720.

Lisulo, M., Sugimoto, C., Kajino, K., Hayashida, K., Mudenda, M., Moonga, L., Ndebe, J., Nzala, S. and Namangala, B. (2014) 'Determination of the prevalence of African trypanosome species in indigenous dogs of Mambwe district, eastern Zambia, by loop-mediated isothermal amplification', *Parasites and Vectors*, 7(1), pp. 7-19.

Lun, Z. R. and Desser, S. S. (1995) 'Is the broad range of hosts and geographical distribution of *Trypanosoma evansi* attributable to the loss of maxicircle kinetoplast DNA?', *Parasitology Today*, 11(4), pp. 131–133.

Majiwa, P. A. O., Maina, M., Waitumbi, J. N., Mihok, S. and Zweygarth, E. (1993) '*Trypanosoma* (*Nannomonas*) *congolense*: Molecular characterization of a new genotype from Tsavo, Kenya', *Parasitology*, 106(2), pp. 151–162.

Malele, I., Nyingilili, H., Lyaruu, E., Tauzin, M., Bernard Ollivier, B., Cayol, J.-L., Fardeau, M.-L. and Geiger, A. (2018) 'Bacterial diversity obtained by culturable approaches in the gut of *Glossina pallidipes* population from a non sleeping sickness focus in Tanzania: preliminary results', *BMC Microbiology*, 18(S1), pp. 164.

Masiga, D. K., Smyth, A. J., Hayes, P., Bromidge, T. J. and Gibson, W. C. (1992) 'Sensitive detection of trypanosomes in tsetse flies by DNA amplification', *International Journal for Parasitology*, 22(7), pp. 909–918.

Matthee, C. A. and Davis, S. K. (2001) 'Molecular insights into the evolution of the family Bovidae: A nuclear DNA perspective', *Molecular Biology and Evolution*, 18(7), pp. 1220–1230.

Maudlin, I. and and Welbum, S. C. (1994) 'Maturation of trypanosome infections in tsetse, a review', *Exp. Parasitol.*, 79(0031–1820 (Print)), pp. 202–205.

Maudlin, I., Kabayo, J. P., Flood, M. E. T. and Evans, D. A. (1984) 'Serum factors and the maturation of *Trypanosoma congolense* infections in *Glossina morsitans*', *Zeitschrift für Parasitenkunde Parasitology Research*, 70(1), pp. 11–19.

Maudlin, I., Welburn, S. C. and Milligan, P. (1990) 'Salivary gland infection: a sex-linked recessive character in tsetse?', *Acta Tropica*, 48(1), pp. 9–15.

Mbewe, N. J., Mweempwa, C., Guya, S. and Wamwiri, F. N. (2015) 'Microbiome frequency and their association with trypanosome infection in male *Glossina morsitans centralis* of Western Zambia', *Veterinary Parasitology*, 211(1–2), pp. 93–98.

Médecins, S. F. (2004) Sleeping sickness or Human African Trypanosomiasis Fact sheet. Available at: http://www.accessmed-msf.org/fileadmin/user_upload/diseases/other_diseases/sleepingsicknessfs.pdf (Accessed: 27 February 2018).

Mesu, V. K. B. K. et al. (2018) 'Oral fexinidazole for late-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial', *The Lancet*, 391(0116), pp. 144-154

Mihok, S., Olubayo, R. O., Darji, N. and Zweygarth, E. (1993) 'The influence of host blood on infection rates in *Glossina morsitans* spp. infected with *Trypanosoma congolense*, *T. brucei* and *T. simiae*', *Parasitology*, 107(1), pp. 41–48.

Moloo, S. K. (1993) 'The distribution of *Glossina* species in Africa and their natural hosts', *International Journal of Tropical Insect Science*, 14(4), pp. 511–527.

Moloo, S. K., Asonganyi, T. and Jenni, L. (1986) Cyclical development of *Trypanosoma brucei gambiense* from cattle and goats in *Glossina*, *Acta Tropica*, 43, pp. 407–408.

Morrison, L. J., Vezza, L., Rowan, T. and Hope, J. C. (2016) 'Animal African Trypanosomiasis: Time to increase focus on clinically relevant parasite and host species', *Trends in Parasitology*, pp. 599–607.

Muhanguzi, D., Picozzi, K., Hattendorf, J., Thrusfield, M., Kabasa, J. D., Waiswa, C. and Welburn, S. C. (2014) 'The burden and spatial distribution of bovine African trypanosomes in small holder croplivestock production systems in Tororo District, south-eastern Uganda', *Parasites and Vectors*, 7(1), pp. 603.

Murray, C. J. L. (1994) 'Quantifying the burden of disease: The technical basis for disability-adjusted life years', *Bulletin of the World Health Organization*, pp. 429–445.

Muturi, C. N., Ouma, J. O., Malele, I. I., Ngure, R. M., Rutto, J. J., Mithöfer, K. M., Enyaru, J. and Masiga, D. K. (2011) 'Tracking the feeding patterns of tsetse flies (*glossina* genus) by analysis of bloodmeals using mitochondrial cytochromes genes', *PLoS ONE*, 6(2). e001547

Muturi, E. J., Dunlap, C., Ramirez, J. L., Rooney, A. P. and Kim, C. H. (2019) 'Host blood-meal source has a strong impact on gut microbiota of *Aedes aegypti*', *FEMS microbiology ecology*, 95(1).

Naessens, J. (2006) 'Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?', *International Journal for Parasitology*, 36(5), pp. 521–528.

Ngaira, J. M., Njagi, E. N. M., Ngeranwa, J. J. N. and Olembo, N. K. (2004) 'PCR amplification of RoTat 1.2 VSG gene in *Trypanosoma evansi* isolates in Kenya', *Veterinary Parasitology*, 120(1–2), pp. 23–33.

Njiru, Z. K., Constantine, C. C., Guya, S., Crowther, J., Kiragu, J. M., Thompson, R. C. A. and Dávila, A. M. R. (2005) 'The use of ITS1 rDNA PCR in detecting pathogenic African trypanosomes', *Parasitology Research*, 95(3), pp. 186–192.

Njiru, Z. K., Constantine, C. C., Masiga, D. K., Reid, S. A., Thompson, R. C. A. and Gibson, W. C. (2006) 'Characterization of *Trypanosoma evansi* type B', *Infection, Genetics and Evolution*, 6(4), pp. 292–300.

Odiit, M., Shaw, A., Welburn, S. C., Fèvre, E. M., Coleman, P. G. and McDermott, J. J. (2004) 'Assessing the patterns of health-seeking behaviour and awareness among sleeping-sickness patients in eastern Uganda', *Annals of Tropical Medicine & Parasitology*, 98(4), pp. 339–348.

Omolo, M. O., Hassanali, A., Mpiana, S., Esterhuizen, J., Lindh, J., Lehane, M. J., Solano, P., Rayaisse, J. B., Vale, G. A., Torr, S. J. and Tirados, I. (2009) 'Prospects for developing odour baits to control *Glossina fuscipes* spp., the major vector of human African trypanosomiasis', *PLoS Neglected Tropical Diseases*, 3(5), e184201.

Pakpour N, Akman-Anderson L, Vodovotz Y, Luckhart S. (2013) 'The effects of ingested mammalian blood factors on vector arthropod immunity and physiology'. *Microbes and Infection*,15, pp. 243–254.

Palmer, J. M., Jusino, M. A., Banik, M. T. and Lindner, D. L. (2018) 'Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data', *PeerJournal*, 6, e4925.

Paparini, A., Gofton, A., Yang, R., White, N., Bunce, M. and Ryan, U. M. (2015) 'Comparison of Sanger and next generation sequencing performance for genotyping *Cryptosporidium* isolates at the 18S rRNA and actin loci', *Experimental Parasitolog*, 151–152, pp. 21–27.

Peacock, L., Cook, S., Ferris, V., Bailey, M. and Gibson, W. (2012) 'The life cycle of *Trypanosoma* (*Nannomonas*) *congolense* in the tsetse fly', *Parasites and Vectors*, 5(1), pp. 109.

Peacock, L., Ferris, V., Bailey, M. and Gibson, W. (2007) 'Dynamics of infection and competition between two strains of *Trypanosoma brucei brucei* in the tsetse fly observed using fluorescent markers', *Kinetoplastid Biology and Disease*, 6(1), pp. 4.

Picozzi, K., Carrington, M. and Welburn, S. C. (2008) 'A multiplex PCR that discriminates between *Trypanosoma brucei brucei* and zoonotic *T. b. rhodesiense*', *Experimental Parasitology*, 118(1), pp. 41–46.

Rosli, M. K. A., Zamzuriada, A. S., Syed-Shabthar, S. M. F., Mahani, M. C., Abas-Mazni, O. and Md-Zain, B. M. (2011) 'Optimization of PCR conditions to amplify Cyt b, COI and 12S rRNA gene fragments of Malayan gaur (*Bos gaurus hubbacki*) mtDNA', *Genetics and Molecular Research*, 10(4), pp.

Sánchez, E., Perrone, T., Recchimuzzi, G., Cardozo, I., Biteau, N., Aso, P., Mijares, A., Baltz, T., Berthier, D., Balzano-Nogueira, L. and Gonzatti, M. (2015) 'Molecular characterization and classification of *Trypanosoma* spp. Venezuelan isolates based on microsatellite markers and kinetoplast maxicircle genes', *Parasites and Vectors*, 8(1).27-38.

Saridaki, A. and Bourtzis, K. (2010) '*Wolbachia*: more than just a bug in insects genitals', *Current Opinion in Microbiology*, pp. 67–72.

Schwan, T. G., Lopez, J. E., Safronetz, D., Anderson, J. M., Fischer, R. J., Maïga, O. and Sogoba, N. (2016) 'Fleas and trypanosomes of peridomestic small mammals in sub-Saharan Mali', *Parasites and Vectors*, 9(1), pp. 1–7.

Simarro, P. P., Cecchi, G., Franco, J. R., Paone, M., Diarra, A., Ruiz-Postigo, J. A., Fèvre, E. M., Mattioli, R. C. and Jannin, J. G. (2012a) 'Estimating and mapping the population at risk of sleeping sickness', *PLoS Neglected Tropical Diseases*. Edited by J. M. Ndung'u, 6(10), e1859.

Simarro, P. P., Jannin, J. and Cattand, P. (2008) 'Eliminating human African trypanosomiasis: Where do we stand and what comes next?', *PLoS Medicine*, 5(2), pp. 0174–0180.

Simo, G., Njitchouang, G. R., Njiokou, F., Cuny, G. and Asonganyi, T. (2011) '*Trypanosoma brucei* s.l.: Microsatellite markers revealed high level of multiple genotypes in the mid-guts of wild tsetse flies of the Fontem sleeping sickness focus of Cameroon', *Experimental Parasitology*, 128(3), pp. 272–278.

Squarre, D., Kabongo, I., Munyeme, M., Mumba, C., Mwasinga, W., Hachaambwa, L., Sugimoto, C. and Namangala, B. (2016) 'Human African Trypanosomiasis in the Kafue National Park, Zambia', *PLoS Neglected Tropical Diseases*, 10(5), e0004567.

Stephen G.A. Leak (1999) *Tsetse biology and ecology: Their role in the epidemiology and control of trypanosomiasis*. New York, NY: CABI.

Sternberg, J. M. (2004) 'Human African trypanosomiasis: Clinical presentation and immune response', *Parasite Immunology*, 26, pp. 469–476.

Stevens, J. R. and Brisse, S. (2004) 'Systematics of trypanosomes of medical and veterinary importance', in *The trypanosomiases*, edited by Maudlin I, Holmes P. H, and Miles M.A, pp. 1–23.

Stevens, J. R., Noyes, H. A., Schofield, C. J. and Gibson, W. (2001) 'The molecular evolution of *Trypanosomatidae*', *Advances in Parasitology*, 48, pp. 1–53.

Swei, A. and Kwan, J. Y. (2017) 'Tick microbiome and pathogen acquisition altered by host blood meal', *ISME Journal*. Nature Publishing Group, 11(3), pp. 813–816.

Tran, T., Napier, G., Rowan, T., Cordel, C., Labuschagne, M., Delespaux, V., Van Reet, N., Erasmus, H., Joubert, A. and Büscher, P. (2014) 'Development and evaluation of an ITS1 "Touchdown" PCR for assessment of drug efficacy against animal African trypanosomosis', *Veterinary Parasitology*. Elsevier, 202(3–4), pp. 164–170.

Urakawa, T. , Majiwa, P. , Hirumi, H. (1998) 'Comparative analyses of ribosomal RNA genes of African and related trypanosomes, including *Trypanosoma evansi*', *The journal of protozoology research*, 8(4), pp. 224–226.

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M. and Miya, M. (2017) 'Environmental DNA enables detection of terrestrial mammals from forest pond water', *Molecular Ecology Resources*, 17(6), e63–e75.

Van den Bossche, P. and Delespaux, V. (2011) 'Options for the control of tsetse-transmitted livestock trypanosomosis. An epidemiological perspective', *Veterinary Parasitology*, 181(1), pp. 37–42.

Van Den Bossche, P., Ky-Zerbo, A., Brandt, J., Marcotty, T., Geerts, S. and De Deken, R. (2005) 'Transmissibility of *Trypanosoma brucei* during its development in cattle', *Tropical Medicine and International Health*, 10(9), pp. 833–839.

Van den Bossche, P., Rocque, S. de La, Hendrickx, G. and Bouyer, J. (2010) 'A changing environment and the epidemiology of tsetse-transmitted livestock trypanosomiasis', *Trends in Parasitology*, pp. 236–243.

Vickerman, K. (1965) 'Polymorphism and mitochondrial activity in sleeping sickness trypanosomes', *Nature*, 208(5012), pp. 762–766.

Vickerman, K. (1985) 'Developmental cycles and biology of pathogenic trypanosomes.', *British medical bulletin*, 41(2), pp. 105–114.

Vun, V. F., Mahani, M. C., Lakim, M., Ampeng, A. and Md-Zain, B. M. (2011) 'Phylogenetic relationships of leaf monkeys (Presbytis; Colobinae) based on cytochrome b and 12S rRNA genes.', *Genetics and molecular research*,10(1), pp. 368–381.

Weiss, B. and Aksoy, S. (2011) 'Microbiome influences on insect host vector competence', *Trends in Parasitology*. doi: 10.1016/j.pt.2011.05.001.

Weiss, B. L., Wang, J., Maltz, M. A., Wu, Y. and Aksoy, S. (2013) 'Trypanosome Infection Establishment in the Tsetse Fly Gut Is Influenced by Microbiome-Regulated Host Immune Barriers', *PLoS Pathogens*, 9(4).251-268.

welburn, S. C., Arnold, K., Gooday, G. W. and Maudlin, I. (1993) 'Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies', *Parasitology*, 107(2), pp. 141–145.

WHO (1998) 'Control and surveillance of African trypanosomiasis', Worrld Health Organization Technical Report Series, pp. 1–113. Available at: http://apps.who.int/iris/bitstream/10665/42087/1/WHO_TRS_881.pdf (Accessed: 30 October 2017).

WHO (2013) 'Control and surveillance of human African trypanosomiasis.', World Health

Organization technical report series, (984), pp. 1–237.

WHO (2015) ‘Investing to overcome the global impact of neglected tropical diseases: third World Health Organization report on neglected diseases,’ pp. 191.

WHO (2017) ‘World Health Organization Cardiovascular diseases Fact sheet’, Available at: http://www.who.int/mediacentre/factsheets/fs317/en/.

WHO (2018) ‘Trypanosomiasis, human African (sleeping sickness), World Health Organization-Fact sheets’. Available at: http://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness) (Accessed: 16 February 2018).

Wuyts, N., Chokesajjawatee, N. and Panyim, S. (1994) ‘A simplified and highly sensitive detection of *Trypanosoma evansi* by DNA amplification.’, *The Southeast Asian journal of tropical medicine and public health*, 25(2), pp. 266–271.

6 Acknowledgements

I owe my cordial appreciation to my supervisors, Prof. Chihiro Sugimoto and Prof. Junya Yamagishi, who have guided my research and allowed me the freedom and flexibility to explore my research ideas. I would also like to thank them for providing resources to conduct my research and for availing and providing various platforms and opportunities for meeting researchers in my field. I would like to thank Dr. Kyoko Hayashida for her support, advice, and guidance throughout the course of my wet lab. experiments. My sincere gratitude to Dr. Axel Martinelli for his advice. I am grateful for Ms. Kawai Naoko for her support in the wet lab. activities and her kind assistance with Japanese documents during my study period. I would like to acknowledge the Leading Program in the Graduate School of Veterinary Medicine for providing grant for internship and epidemiology work in abroad. My sincere gratitude to Prof. Aksoy Serap, Yale School of Public Health, Yale University, USA, for hosting me as an intern in her laboratory. I would like to remember and thank to my all lab mates for their assistance and nice company during the course of my studies. I would like to thank my parents and siblings for their support in my quest to aquire my doctorate. Finally, I deeply appreciate the experimental animals that supported my research activities.

7 Summary in Japanese

Human African trypanosomiasis（HAT、あるいはアフリカ睡眠病）および Animal African trypanosomiasis（AAT）、あるいはナガナ病）は共にサブサハラアフリカの風土病である。その病原体はトリパノソーマ属の原虫であり、ツェツェバエにより媒介される。なかでも、HAT は、*Trypanosoma brucei rhodesiense* および、*T. b. gambiense* によって引き起こされることが知られている。一方、AAT は *T. b. brucei*、*T. congolense*、*T. vivax*、*T. evansi* など多様な病原体が引き起こす。これらの原虫を検出するにあたり、Internal Transcribed Spacer 1 (ITS1)を標的とした PCR が広く用いられており、増幅産物のサイズにより種の同定が可能である。さらに詳細な種内系統に関する情報を得るためには、増幅産物の塩基配列を解析する必要があるが、解析費用や多重感染が問題となるため一般的に行われているとは言い難い。

第 1 章では、上記の増幅産物をプールし次世代シーケンサーを用いて一括解析する系（Amplicon-seq）の確立を行った。付随して、感度に優れるプライマーの再設計・評価も行い、それらを組み合わせることで、低コスト、かつ多重感染サンプルの解析を可能とし、さらに、従来法と比較して感度と特異性においても優れるトリパノソーマ原虫の検出方法の開発に成功した。本方法を用いて、ザンビアおよびジンバブエで捕集したツェツェバエ（Rufunsa, *n*=200, 種未同定；Kafue, *n*=85, *Glossina morsitans centralis*；Hurungwe, *n*=188, 種未同定）におけるトリパノソーマ原虫の保有率に関する調査を実施した結果、約 30%のツェツェバエから、5 種、378 系統のトリパノソーマ原虫由来配列の検出に成功した。

第 2 章では、ツェツェバエの吸血源動物種を特定し、トリパノソーマ原虫の reservoir を推定することを目的に、哺乳動物 12S rRNA 配列の増幅と次世代シーケンサーによる Amplicon-seq 解析を実施した。鋳型には、ITS1 解析に供試したものと同じツェツェバエ由来 DNA （Kafue, *n*=85）を用いた。その結果、ヒト由来配列を 43 検体から、black rat（*Rattus rattus*）由来配列を 27 検体から検出した。加えて、African buffalo（*Syncerus caffer*）、common warthog（*Phacochoerus africanus*）、greater kudu（*Tragelaphus strepsiceros*）、straw-colored fruit bat（*Eidolon helvum*）などのアフリカに

生息する各種野生動物、cattle（*Bos taurus*）、dog（*Canis lupus familiaris*）などの家畜に由来する配列も検出された。これまで black rat や straw-colored fruit bat は吸血源として重要視されていなかったことから、本研究により Kafue 地域のツェツェバエが想定以上に多様な動物を吸血源とすることが示唆された。

第 3 章では、同じく Kafue のツェツェバエから精製した DNA を鋳型に、16S rRNA V3-V4 領域の多様性に基づく細菌叢解析を行った。その結果、ツェツェバエの共生細菌として知られている *Sodalis*、*Wolbachia*、*Wigglesworthia* の 3 属が過去の知見と同様に優占的に検出された。続いて多次元尺度構成法を用いて細菌叢を 2 次元にプロットし、吸血源、ハエの性別、トリパノソーマ感染の有無との相関を解析したが、現状では明らかな違いは認められなかった。

以上、本研究ではツェツェバエから抽出したDNAから、トリパノソーマ原虫の感染の有無、およびその系統、吸血源、細菌叢を大規模に解析する系の確立を行った。本系を用いることで同一個体から上記情報を得ることができる。従って、今後、検体数を増やし各情報間の相関を取ることで、reservoirに関する手がかり、あるいは、新たな原虫 - 細菌叢の相互作用等が見出される可能性がある。