January 2014

Analysis Of The Gut-Specific Microbiome Of Field-Captured Tsetse Flies, And Its Potential Relevance To Host Trypanosome Vector Competence

Bridget Catherine Hamilton Griffith
Yale University, bridgetcgriffith@gmail.com

Analysis of the gut-specific microbiome of field-captured tsetse flies, and its potential relevance to host trypanosome vector competence

Bridget C. Griffith

Presented in candidacy for the degree of

Masters of Public Health

Epidemiology of Microbial Diseases

Global Health

Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven,

CT. 06520

May 2014

Readers: Dr. Serap Aksoy, Dr. Andrew Goodman

Table of Contents:

Acknowledgements:
I would like to thank my advisors, Dr. Serap Aksoy and Dr. Andrew Goodman for their guidance and support. I would also like to thank all the members of the Aksoy lab, especially Dr. Brian Weiss and Emre Aksoy, and the Kenya Agricultural Research Institute Trypanosomiasis Research Centre staff, especially Dr.Grace Murilla and Joann Auma. Funding provided by the Yale Global Health Initiative. *Asante sana.*

Abstract:
The tsetse fly (*Glossina* sp.) gut is colonized by maternally transmitted and environmentally acquired bacteria. Maternally transmitted symbionts are well characterized, but little is known about the origin and function of environmentally acquired bacteria inhabiting the tsetse midgut. To address this shortcoming, culture dependent and independent methods were used to characterize and quantify bacterial communities that reside within the midgut of tsetse flies collected at five geographically isolated locations in Kenya and Uganda. Bacteria were isolated from 83.33% of flies using culture dependent methods. In total, 38 strains were isolated. To increase the depth of this analysis, Illumina-based deep-sequencing of the V4 hypervariable region of the bacterial 16S ribosomal RNA (rRNA) gene was used to determine the relationship between the type and number of bacteria present in *Glossina pallidipes* midguts and geographic location. The midgut was dominated by the obligate endosymbiont, *Wigglesworthia*, but a diversity of other microbiota in individuals that were unique to each location was also observed. This was consistent between trypanosome infected and uninfected samples. The results of this study will increase the understanding of the breadth and depth of tsetse's midgut bacterial communities as they relate to the environmental acquisition of the fly's microbiome and its correlation to trypanosome vectorial capacity. Future studies can then be performed to determine the specific molecular mechanisms that underlie bacteria-mediated trypanocidal immunity in tsetse's gut.

Introduction:

Tsetse flies are medically and agriculturally important insect vectors that transmit African trypanosomes, the causative agents of sleeping sickness in humans and nagana in domesticated animals [1]. Approximately 70 million people over an area of 1.55 million km^2 are at risk for contracting HAT, which is fatal if left untreated [2]. Additionally, nagana is estimated to cost African agriculture US$4.5 billion per year [3, 4]. By influencing food production, natural resource utilization, and human settlement patterns, trypanosomes are seen by the African Union as one of the greatest constraints to Africa's socio-economic development [5]. Studies that address the development of novel disease control strategies are urgently needed [2, 6, 7]. Current drug therapy for HAT is unsatisfactory, as medications are toxic and difficult to administer to humans. Furthermore, drug resistance is increasing [8]. Vaccines are not available and are unlikely to be developed due to the antigenic variation exhibited by the trypanosome in the mammalian host [9]. Both the World Health Organization (WHO) and African countries (through the Pan African Tsetse and Trypanosomiasis Eradication Campaign; PATTEC) have recognized the significant role that vector control should play in ridding the African sub-continent of trypanosomal diseases [10].

While only a small percentage of tsetse are infected with trypanosomes, all individuals house maternally-transmitted and environmentally-acquired bacteria in their gut. Two of these microbes, obligate *Wigglesworthia* and commensal *Sodalis*, are vertically transmitted via milk gland secretions to developing intrauterine larvae during tsetse's unique viviparous mode of reproduction. Obligate *Wigglesworthia* is found in all tsetse as part of an ancient host/symbiont relationship. Two distinct populations of this bacterium exist within tsetse; one intracellularly within a 'bacteriome' organ immediately adjacent to the fly's anterior midgut, and the second

extracellularly in milk gland secretions [11]. As a result of *Wigglesworthia's* long co-evolution with tsetse, the bacterium presents a highly reduced genome (700 kb). However, *Wigglesworthia* retains the capacity to synthesize a complement of vitamins and cofactors that likely supplement nutrients missing from tsetse's vertebrate blood-specific diet [12]. As such, flies raised experimentally in absence of *Wigglesworthia* are reproductively sterile. Additionally, these flies are immune-compromised in that they display highly compromised cellular and humoral responses following microbial challenge [13, 14]. Commensal *Sodalis'* genome is similar to that of many free living microbes, suggesting a recent association with its tsetse host [15]. This bacterium exhibits a distinctly different relationship with tsetse in that it resides intra- and extracellularly in numerous tsetse tissues, including the fly's gut, salivary glands, hemolymph, and milk gland. Unlike *Wigglesworthia*, *Sodalis* can be cultured outside of the tsetse fly [16, 17], and this characteristic makes it amenable as a potential candidate for paratransgenic control of trypanosome transmission [18]. *Sodalis'* functional role is currently unknown, although specific elimination of the bacterium from tsetse results in a reduction in fly longevity [19]. Finally, some populations of tsetse house parasitic *Wolbachia*, which is maternally transmitted through the germ line and can cause reproductive abnormalities such as cytoplasmic incompatibility [20].

In addition to the above-mentioned endosymbionts, tsetse also hosts a variety of environmentally-acquired bacteria in its gut. Previous studies demonstrated that these bacterial populations vary considerably between individual flies within a population [21-23]. Genera identified included *Acinetobacter*, *Enterobacter*, *Enterococcus* [21], *Chryseobacterium*, *Sphingobacterium*, *Providencia*, *Lactococcus*, *Pseudomonas* and *Staphylococcus* [23]. While

these studies provided a useful initial overview of tsetse's gut-associated microbiota, they were limited in their depth of coverage. Specifically, OTUs were identified by culture dependent methods and via the generation of 16s rRNA clone libraries. These techniques lacked the stringency required to identify microbes present at low densities. The origin and functions of environmentally-acquired gut microbes, as they relate to the physiology of their tsetse host, remains to be elucidated [24]. Research using other vector model systems (i.e., mosquitoes) suggests that environmental microbes present in the insect gut modulate their host's vector competence [13]. I propose that a similar situation occurs in tsetse.

In this study I used culture dependent and independent methods to examine the microbiota present in the midgut of field-captured tsetse flies. Culture dependent analysis was done on three species of tsetse fly from three geographically distinct areas. Bacteria were cultured and isolated from the midgut of twenty individuals and classified using PCR amplification and sequencing of the 16s rRNA of each isolate. Additionally, I used next generation sequencing of the 16s rRNA V4 hypervariable region to analyze the diversity of the midgut microbiota in three geographically distinct populations of *G. pallidipes*. The results of this analysis are compared to the culture-dependent results. Findings are discussed in terms of the relationship between culture dependent and independent isolates, and the differences in the microbial composition of the *G. pallidipes* midgut in different geographical areas, and the differences in α-diversity (species richness) of infected individuals.

Methods

Fly collections

Flies were trapped from May 2013 to August 2013 in four geographically distinct tsetse endemic regions of Kenya and Uganda (Figure 1, and see information on specific collection sites below). Tsetse were captured using Epsilon F3 *Ngu* cloth traps and biconical cloth traps baited with acetone and phenol, which are established tsetse olfactory attractants [25, 26]. *Glossina* were identified to species based on morphological criteria.

Site-specific collection details

Shimba Hills National Reserve, Kenya (S 4.18°, E 39.45°)

Flies were collected between June 11-14 2013 and transported to a lab facility in Kwale, Kenya. All dissections were done at a lab facility at Kwale District Hospital. Midguts were separated from fly carcasses and probed microscopically for trypanosome infection. For the culture dependent pathway, serial dilutions of homogenized fly midguts were plated on LB in ambient oxygen concentrations and incubated overnight. For the culture independent pathway, midguts were pooled by infection status into groups of 5 individuals and DNA was extracted using Qiagen DNeasy kit at laboratory facility in Kwale.

Trans Mara (S0° 12.819', E034°53.864')

Flies were collected between July 2-3 2013 and transported to a lab facility in Lolgorien, Kenya. All dissections were done at lab facility. For culture dependent methods, serial dilutions of fly midguts were plated on both LB and BHI media in microaerophillic conditions and on LB and

BHI in ambient oxygen concentrations. DNA extractions were done in KARI-TRC labs using Qiagen DNeasy kit.

Western Kenya (on the shore of Lake Victoria; S00°36.541', E034°5.500')

Flies were collected between July 15-17 2013 and transported to a lab facility in the Suba County District Hospital in Sindo, Kenya. All dissections were done at lab facility. Guts were plated on LB and BHI in microaerophillic and ambient oxygen concentrations. Midgut DNA was extracted at KARI-TRC using Qiagen DNeasy kit.

Nguruman escarpment (S01°53.327', E036°4.550')

Flies were collected between August 2-4 2013 and stored in 100% ethanol for transport to Yale. The fly midguts were dissected from the stored flies, and DNA extractions were done using Qiagen DNeasy kits.

Murchison Falls, Uganda (N 2.25°, E 31.8°)

Flies were collected in 2012 and stored in 100% ethanol for transport to Yale. The fly midguts were dissected from the stored flies, and DNA extractions were done using the MasterPure Complete DNA Purification kit (EpiCentre, Madison, WI, USA).

Culture dependent identification of bacteria

The culture dependent isolation of bacterial samples from the tsetse midgut (*G. pallidipes, G. brevipalpis, G. fuscipes, G. fuscipleuris*) was conducted according to the method outlined by Lindh and Lehane [22]. Flies were surface sterilized via submersion for 5 minutes in 10% bleach, 5 minutes in 70% ethanol, and 5 minutes in sterile water. All tools and dissection surfaces were sterilized after each dissection using 100% ethanol. The midgut of each fly was extracted and

homogenized with a pestle in 20ul sterile phosphate buffer saline (PBS). A sample of each midgut was probed microscopically for the presence of trypanosomes. Homogenized midguts were serially diluted up to a 10^{-10} dilution in sterile PBS, and a volume of each dilution was plated.

Bacterial culture and isolation processes were performed under different combinations of growth medium and oxygen concentration. Depending on the resources available at each collection site, bacteria were cultured under a combination of ambient or microaerophillic (using GasPak EZ Campy Container System; BD Bioscience) oxygen concentrations in three types of media: Brain-Heart Infusion media with 10% blood (BHIB) and LB (Miller's modification). Stringent procedures were employed in an effort to process samples under sterile conditions. Sterile technique was confirmed using negative control plates maintained under identical conditions to each culture environment.

After incubation, all bacterial colonies expressing a unique morphology were inoculated into 900ul of respective media and incubated while shaking. Liquid cultures were preserved with 10% glycerol, flash frozen, and stored at -80 degrees C.

16s rRNA-based phylogenetic analysis of culturable bacteria present in tsetse's gut

DNA was extracted from bacterial cells using a Qiagen DNAEasy blood and tissue kit. DNA isolated from bacterial clones was PCR amplified using primers (SI table 1) that specifically target the bacterial 16s rRNA gene. 38 bacterial clones were taxonomically characterized using this procedure [27].

PCR products were sequenced at the DNA analysis facility on Science Hill at Yale University. 16s rRNA sequence data was compared to catalogued sequences using the Basic Local Alignment Search Tool (BLASTn) and CLC Bio software.

Culture independent identification of bacteria

Culture independent identification of bacterial samples from the gut of field-captured tsetse was conducted according to the method outlined by Lindh and Lehane [22]. Flies were stored in 100% ethanol and transported to laboratory facilities at Yale University. DNA was extracted using either a Qiagen DNeasy Blood & Tissue Kit (Qiagen) or an Epicentre MasterPure Complete DNA and RNA Purification kit. DNA integrity was assessed by PCR amplification (primers sequences listed in SI table 1) of a 500 bp fragment of tsetse β-tubulin gene. Trypanosome infection status was confirmed by PCR (primers sequences listed in SI table 1) using trypanosome Intertranscribed Spacer(ITS1) primers [28].

PCR amplification of the bacterial 16s rRNA gene and Illumina library preparation

Genomic DNA was extracted from the guts of 163 field captured tsetse (Gpd: n=108, Gmm: n=23, Gff: n=1, Gfp: n=9, Gb: n=9, Ga: n=13) using either a DNEasy Blood and Tissue Kit (Qiagen) or a Masterpure-Complete DNA and RNA Purification Kit (EpiCentre). Negative control extractions were performed on reagents from the Qiagen and Epicentre DNA extraction kits, as well as the New England Biolabs (NEB) Phusion PCR kit (see below). Samples were then PCR amplified using barcoded Illumina fusion primers (generously donated by Dr. Howard Ochman) that specifically target a 300 bp region of the V4 hypervariable region of the bacterial 16s gene

(Figure 2) [29]. Primers used to generate 16s rRNA products can be found at (ftp://ftp.metagenomics.anl.gov/data/misc/EMP/SupplementaryFile1_barcoded_primers_515F_806R.txt). Each sample was assigned a unique 12 base pair Golay barcode located on the 806R primer. Each PCR reaction was carried out in 30ul of volume containing 1ul of DNA, 0.2ul of Phusion Taq (NEB), 6ul of 5x reaction buffer, 0.6ul of 10mM dNTPs, 0.75ul of 10uM forward and reverse primers (Table S2) and 20.7uls of dH_20. Cycling conditions were 1 minute of initial denaturation at 98°C followed by 35 cycles at 98°C for 10 seconds, 54°C for 15 seconds, 72°C for 15 seconds and a final elongation step at 72°C for 2 minutes. PCR reactions were performed in triplicate, pooled together and analyzed on a 2% agarose gel. The final PCR product, which was 384 bp in length, included 5' and 3' Illumina barcodes that flanked the paired 300 bp target region (Figure 2).

The PCR fragments were cleaned using Agencourt AMpure XP beads (Catalog # A63880, Beckman Coulter), as per the manufacturer's protocol. PCR products were quantified using the Qubit dsDNA High Sensitivity Assay (Catalog # Q32851, Life Technologies, Guilford, CT), and positive reactions were pooled at an equal molar concentration. In total, 16s rRNA sequences were pooled from 163 fly midguts. The pooled sample was sent to Yale Center of Genome Analysis for analysis and sequenced on an Illumina MiSeq sequencing system.

A previous characterization tsetse's gut microbiota revealed that the population is dominated (> 99% of the total population) by obligate *Wigglesworthia* [30]. This phenomenon significantly reduced, or may have entirely excluded, detection and accurate quantification of other less well represented microbes. In an effort to reduce *Wigglesworthia* bias, a restriction digest of each PCR fragment was performed using the *Eco*RV restriction endonuclease (New

England BioLabs). This enzyme preferentially cleaves within the V4 hypervariable region of *Wigglesworthia's* 16s gene, thus reducing the abundance of this specific product relative to those remaining in each sample. It has been demonstrated that EcoRV is *Wigglesworthia* 16s rDNA specific, and it does not have a cleavage site for tsetse secondary symbionts, including *Sodalis* [31].

Next generation sequencing data analyses

Reads were quality checked using FASTqc. The 16s rRNA sequence dataset was demultiplexed and forward and reverse reads were paired using SeqPrep. In the event of a mismatched read, quality scores associated with each base were used to determine the best pairing. In order to improve sequencing accuracy of low diversity samples, a phiX DNA control was added. To remove the phiX reads from the data set, paired reads were mapped to the phiX genome using Bowtie2 [32]. A list of reads not matching the phiX genome was generated using SamTools [33], and the resulting reads were separated from the phiX genome using QIIME software filter_fasta.py script [34]. Sequences were entered into the QIIME pipeline using the split_libraries_fastq.py command. Sequences were clustered via Uclust using the pick_otus.py command at 97% sequence similarity against the Greengenes Ribosomal database (greengenes.lbl.gov). The reference file was customized to include *Wigglesworthia* and *Sodalis* V4 region sequence, which is absent from the Greengenes database. In order to remove possible chimeras and other PCR errors, all Operational Taxonomy Units (OTUs) that did not align to the Greengenes database were excluded from the analysis. To remove OTUs with low read number, OTUs tables were filtered at 0.005% of the total number of reads. Additionally,

Halomonas OTUs, a known contaminant in the PCR kits used, were removed from the analysis. The *G. pallidipes* samples were separately analyzed to determine the relationship between gut microbiota composition and tsetse trypanosome infection status, geographical location within one fly species. The relative abundance of each OTU was measured using the summarize_taxa.py command.

For α-diversity (species richness) calculations, dataset were rarefied to a depth of 15,001 sequences per sample. α-diversity was calculated using the "observed species" metric with 10 iterations at each sequencing depth. The number of OTUs at each sampling depth was averaged to make the rarefaction curves. To compare the α-diversity between *G. pallidipes* populations in different locations, at two-sample t-test with 1000 Monte Carlo permutations was used. Beta diversity (β-diversity) was calculated for comparisons between geographically distinct populations of *Glossina pallidipes* for geographically distinct origins. Jackknifed Principal Coordinate Analysis (PCoA) and the unweighted UniFrac distance metric was used to visualize the difference between microbial communities form each population using 10 Jackknife replicates.

Results

Glossina collections

For this study, 183 tsetse flies representing 6 species (*G. pallidipes*, *n*=125; *G. morsitans*, *n*=17; *G. fuscipes*, *n*=4, *G. fusciplures*, *n*=12; *G. brevipalpis*, *n*=12; *G. austeni*, *n*=13), were collected from 5 sites (Table 1). Of these 183 flies, 20 individuals representing 4 species (*G. pallidipes*, *n*=11; *G. fuscipes*, *n*=3, *G. fusciplures*, *n*=3; *G. brevipalpis*, *n*=3) were sacrificed in an effort to culture bacteria from their guts (Table 2). For culture independent characterization of tsetse's gut, analysis was limited to 114 trypanosome infected and uninfected *G. pallidipes* individuals that were captured at three distinct geographic locations (Table 3).

Bacteria cultured from the tsetse fly gut

Twenty-four fly midguts were subjected to identical bacterial isolation processes, and 20 [83.33%; 11 *G. pallidipes* (55.0%), 3 *G. brevipalpis* (15.0%), 3 *G. fuscipleuris* (15.0%) and 3 *G. fuscipes* (15.0%)] yielded culturable bacterial clones (Tables 4 and 5). Between 1 and 4 bacterial OTUs were identified from each fly, and 34 bacterial strains were isolated in total. The 16s rRNA gene from all 34 isolates was amplified and sequenced. Following the sequence alignment, members of 14 different genera were identified, of which 7 (50.0%) were Gram-positive and 7 (50.0%) were Gram-negative (Table 3). The most common isolate, *Bacillus sp.,* was identified from 10 (50.0%) of individuals. Both *Bacillus* and *Staphylococcus* were isolated from individuals in all three collection sites. *G. fuscipleuris* collected in Lolgorian, Kenya had the greatest number of genera represented (7), and only one genera was isolated from Shimba Hills *G. brevipalpis*. Of the 20 flies examined, 13 housed one culturable OTU while 7 housed two or more.

Out of the 34 isolates, 21 belonged to the phylum Firmicutes, 6 to Proterobacteria and 7 to Actinobacter. With the exception of *G. brevipalpis*, in which only one bacterial genus was found, representative isolates of the above three phyla were found in each tsetse species.

Tsetse gut-associated microbes identified via culture independent methods

The presence of *Wigglesworthia* and *Sodalis* in tsetse's midgut is well documented [13-15, 35-38]. To acquire a more global view of bacterial taxa present in this environment, I deep sequenced tsetse guts using the Illumina MiSeq platform. A total of 21,728,153 reads were obtained from the 209 barcoded individuals. After quality filtering, a total of 16,635,470 sequences were written (average 79, 727 sequences per sample) and obtained for alignment and taxonomic assignment using the QIIME software pipeline.

Midgut microbial diversity across multiple G. pallidipes populations

The overwhelming abundance of *Wigglesworthia* in tsetse's midgut makes it difficult to visualize other microbes in this tissue [30]. In an effort to ameliorate this impediment, sample DNA was treated with *Eco*RV, which cuts specifically within the V4 region of *Wigglesworthia's* 16s rRNA gene. While this endonuclease treatment did eliminate a significant proportion of *Wigglesworthia* sequence, this endosymbiont still represented the majority of OTUs by % abundance in all samples (NGU=91.99%, SH=76.67%, MF=90.26%). Nonetheless, this procedure did allow for a relatively more in-depth visualization of other microbes present in each sample. *Sodalis*, a commensal tsetse endosymbiont previously linked to trypanosome infection prevalence [19, 35, 36, 39, 40], was the second most abundant microbe (NGU=2.37%, SH=

23.05%, UG=8.09%), although the density of this bacterium varied greatly by geographic location and individual (Figure 6). While some flies harbored high *Sodalis* densities, the opposite was true of many individuals (e.g., %OTU <0.1%). Figures 3-6 display the % OTUs by abundance for each trapping sites. In each figure % abundance of *Wigglesworthia* and *Sodalis* is shown on the bottom graph, while the top graph shows the next 6 most abundant bacteria (averaged across samples and excluding *Wigglesworthia* and *Sodalis*). Results were presented in this manner in an effort to better visualize exogenous microbiota.

In addition to *Wigglesworthia* and *Sodalis*, the next most common taxa varied by location. At Nguruman, the next 6 most abundant taxa represented on average 2.71% of the total OTUs present per sample. These OTUs included members of the family *Enterobacteriaceae* as well as individuals from the genera *Serratia*, *Enterobacter*, *Staphyloccus*, *Pseudoxanthomonas* and *Cloacibacterium* (Figure 3). In *G. pallidipes* captured at Shimba Hills, the next six most abundant gut-associated taxa represented on average 0.25% of the total OTUs per sample, and included members of the family *Enterobacteriaceae* and individuals from the genera *Serratia, Acinetobacter*, *Staphylococcus* and *Pseudomonas* (Figure 4). Finally, the 6 most abundant taxa (minus *Wigglesworthia* and *Sodalis*) comprised 0.62% of all bacteria found in guts of *G. pallidipes* collected in Murchison Falls, Uganda. These microbes included members of the order *Streptophyta* and family *Enterobacteriaceae* as well as individuals from the genera *Acientobacter, Streptococucus, Staphylococcus* and *Serratia* (Figure 5). Figure 6 displays the averages of each sampling site by % abundance. Of the three sites, the Shimba Hills samples are the most highly dominated by *Sodalis* and *Wigglesworthia*, while over 5% of the total microbial abundance is explained by other organisms in Nguruman. These averages illustrate the clear

overall difference in the microbial composition of the microbiome of tsetse from different geographical areas.

Diversity measures

I used the "observed species" metric to measure α-diversity (species richness) of bacteria found in guts of *G. pallidipes* collected from all three sites. The rarefaction curve leveled off at 6,000 sequences per sample, indicating that an adequate sequencing depth and OTU discovery was achieved (Figure 7). At a 95% confidence level (α=0.05), I observed significant differences in α-diversity between two of the three study sites (Table 5). Shimba Hills (SH) had the lowest species richness (12.38±2.45), which was significantly different from Nguruman (NGU; 25.05±17.11; p=0.024). Shimba Hills was not significantly different from Murchison Falls, Uganda (MF; 25.5±28.73; p=0.261), and Murchison Falls and Nguruman were not significantly different from each other (p=1; Figure 8). The species richness of infected samples from all collection sites was calculated using the observed species metric. This data was displayed using a phylogenic tree to illustrate the relative α-diversity relatedness of the samples. When compared to the % abundance of OTUs of each sample, the relatedness of the samples is clearly higher among those with similar *Sodalis* densities (Figure 9b). This was also illustrated on the principal coordinates plot (Figure 9a), in which the samples cluster into two distinct groups when the distance between each sample as calculated by β-diversity of each sample (weighted UniFrac).

I next used the unweighted UniFrac metric, and the Bray Curtis method, to measure β-diversity of bacteria within and between *G. pallidipes* individuals collected at the three

geographic regions. Figure 10 displays the average UniFrac distance within each collection site (a) and between each collection site (b). The principal coordinate plots derived from both tests suggest that the samples do cluster by geographic location (Figures 11 and 12). Using a nonparametric paired t test, significant differences in the β-diversity among the *G. pallidipes* flies trapped in Shimba Hills site and Murchison Falls ($p=0.028$) were observed. The β-diversity of the other two sites was not significantly different at a 95% confidence level (Table 6).

Comparison of culture dependent and independent results

The % OTUs of all individuals within a geographic location was averaged, and the proportion of OTUs cultured was calculated using weighted and unweighted methods. Although the culture dependent data for *G. pallidipes* flies only came from Shimba Hills, the results were compared to all three *G. pallidipes* populations. A small proportion (3.3% from Nguruman, 12.5% from Shimba Hills and 6.0% from Murchison Falls) of the total genera identified by culture independent methods are represented in the population of bacteria successfully cultured from tsetse guts. Furthermore, a small proportion (0.31% from Nguruman, 0.01% from Shimba Hills and 0.12% from Murchison Falls) of the total genera by % OTU were cultured. This result is heavily influenced by the large % OTUs in each sample that are representative of *Wigglesworthia* and *Sodalis*. *Wigglesworthia* cannot be cultured, and *Sodalis* is an obligate anaerobe that is difficult to culture in the presence of other bacteria [16].

Discussion

In the present study, I used culture dependent and independent methods to characterize the tsetse fly midgut microbiota. Culture dependent analysis demonstrated that specific bacterial taxa resident within tsetse's midgut can be grown *in vitro* using previously described methods [22]. Further, more rigorous analysis using next-generation sequencing technology revealed that bacterial taxa present in the midgut one tsetse species (*G. pallidipes*) varied depending on the site at which flies were collected. This habitat-based variation in microbiota composition was reflected in the abundance of both maternally transmitted endosymbionts and environmentally acquired commensals. Additionally, clear differences were present in the microbiota composition between trypanosome infected and uninfected flies of the same *Glossina* species coming from the same location. Overall, these results suggest that 1) tsetse's environment plays an important role in the composition of the fly's midgut microbiome and 2) there may be patterns in the midgut microbial composition that relate to fly trypanosome infection status. More research should be performed to determine the role of environmentally acquired bacteria as they relate to the physiology of their tsetse host.

In this study, 34 bacterial isolates were successfully cultured, of which 21 belonged to the phylum Firmicutes, 6 to Proteobacteria and 7 to Actinobacter. With the exception of *G. brevipalpis*, from which only one bacterial genus was cultured, representative isolates of the above three phyla were found in each tsetse species. Both *Bacillus* and *Staphylococcus* were isolated from individuals in all three collection sites. In this study, Firmicutes and *Bacillus* were the dominant phylum and genus, respectively. Two previous studies [21, 22, 23] have employed culture dependent methods to investigate the microbial composition of tsetse midguts from fly

populations geographically distinct from the ones used here in. Bacterial strains from the Firmicutes, Proteobacteria and Actinobacter were identified in each case, although the dominant phyla observed in each study was variable. This variability could be due to the fact that different culture conditions were utilized in each case. In this study three different types of media were tested, and all samples were grown at room temperature in anaerobic and ambient oxygen concentrations. Thus, it appears as though the diversity and relative abundance of microbes that can be cultured from the tsetse gut depends upon several factors, including the environment in which the fly lives as well as the conditions in which the isolates were grown. Further studies that engage different culture conditions should be performed in an effort to identify additional culturable bacteria taxa from the tsetse fly gut. These bacteria may be of particular importance because they can potentially be genetically modified and/or trans-located between tsetse species or between field-captured flies and insectary-reared individuals that harbor different microbiota. Thus, these taxa may be useful for performing functional experiments that will provide insight into how tsetse's microbiota modifies their host's physiology, including trypanosome vector competence. Additionally, culturable isolates may serve as candidates in a control strategy in which tsetse are colonized with recombinant bacteria that express anti-trypanosomal effector molecules. These 'paratransgenic' flies could exhibit a reduced ability to successfully transmit parasites between mammalian hosts[18, 40].

Previous experiments performed to characterize tsetse's midgut microbiome have lacked the illustrative power of comparing culture dependent and independent analyses in the same study population. This experimental shortcoming is important for two reasons. First, it fails to address what proportion of the total bacterial population is represented by culturable

organisms. Secondly, it does not allow for the identification of taxa that cannot be cultured under the specific experimental conditions used in the study. As such, the vast majority of bacteria present in the niche are left unidentified. Recently, deep sequencing technology, based on the Illumina MiSeq platform, was used in an attempt to acquire a more in-depth overview of the microbiota found in guts of tsetse collected in Uganda [30].This study successfully revealed that tsetse does house a more taxonomically complex gut microbiota than that identified via culture dependent methods and 16s rRNA clone libraries. However, determining a comprehensive picture of the population structure of environmentally acquired microbes present in samples from the Ugandan study was likely obfuscated by the fact that obligate *Wigglesworthia* represented greater than 99% of the cumulative OTUs observed. This impediment was partially circumvented in this study by digesting *Wigglesworthia*-specific V4 PCR products with *Eco*RV endonuclease prior to library sequencing. This treatment succeeded in eliminating a significant proportion of *Wigglesworthia* reads (14.0% of total OTUs) thus allowing for a more comprehensive view of other microbial taxa resident in tsetse midguts analyzed in this study.

Results from this study indicate that adult tsetse flies house a taxonomically complex population of bacteria in their gut. The biological mechanisms that underlie colonization of tsetse's midgut by environmental bacteria requires further investigation. However, the dynamics of this process are presumably different from that which occurs in other well-studied insect models. Larval fruit flies and mosquitoes essentially consume their resident environment, which is rotting organic matter and fetid water, respectively. As such, immature stages of these insects house a robust gut microbiota [41, 42]. Conversely, immature tsetse undergo their

entire larval development within their mother's uterus. This environment is devoid of environmental microbes, and during this time larvae are exposed exclusively to low densities of maternally-transmitted endosymbiotic *Wigglesworthia, Sodalis and Wolbachia* [13, 43]. This temporal discrepancy in exposure to environmental bacteria by *Drosophila* and *Anopheles* in comparison to tsetse may shape immune function in subsequent adult stages. In support of this theory, adult *Drosophila* present potent basal immunity in their gut while tsetse does not [14, 44]. This phenomenon may account for the fact that the microbiome of adult tsetse is represented by approximately 3x more phyla than that found in *Drosophila* (12 in tsetse versus 4 in *Drosophila*) [this study, and 42]. Overall, more research must be done to understand how this delayed acquisition of environmentally acquired flora impacts tsetse physiology.

To date, no experimental evidence exists to suggest that environmentally acquired bacteria mediate trypanosome infection outcomes in tsetse. However, studies performed using other insect vector models systems indicate bacteria of this nature do modulate host vector competence. For example, *Anopheles gambiae*, which is the principle vector of human malaria (*Plasmodium* sp.), harbors a taxonomically diverse assemblage of gut-associated bacteria [45]. Among this population, a commensal from the genus *Enterobacter* (designated '*Enterobacter* sp. Z, or *Esp*_Z) exhibits direct anti-*Plasmodium* activity via the production of toxic reactive oxygen intermediates [58]. *Plasmodium*-susceptible laboratory lines of *A. gambiae* were rendered highly resistant to parasite infection with when they had been inoculated with *Esp*-Z prior to exposure to an infectious blood meal [58]. Additionally, *A. gambiae* can harbor *Serratia marcescens* in its gut, and this bacterium is also associated (through a currently unknown mechanism) with a *Plasmodium*-refractory host phenotype [46, 47]. Both *Enterobacter* and

Serratia strains were identified within the gut of field-captured tsetse (this study, and [48]). Additionally, *Serratia* was found to reside at low densities within the gut of colonized flies. The functional relationship between these bacteria (as well as other environmentally-acquired commensals) and tsetse's competence as a vector of African trypanosomes remains to be elucidated.

Tsetse's more well-studied maternally-transmitted endosymbionts have been shown to mediate their host's trypanosome vector competence through indirect mechanisms. Specifically, tsetse populations that harbor relatively high *Sodalis* densities are more refractory to trypanosome infection than are individuals that house less of this bacterium(this study and [30, 49]. Although the mechanism that underlies this phenomenon is currently not well understood, it may involve a *Sodalis*-hosted prophage that induces potent antimicrobial tsetse immune responses[50]. Additionally, tsetse's obligate endosymbiont, *Wigglesworthia*, also indirectly modulates trypanosome infection outcomes in tsetse. When flies undergo intrauterine larval development in the absence of this bacterium, they exhibit an unusually high susceptibility to trypanosomes[37, 51]. *Wigglesworthia*-free tsetse fail to produce a peritrophic matrix (PM), which is a sleeve-like barrier that lines the fly midgut and separates immuno-reactive epithelial cells from the parasite-containing blood bolus. As such, the midgut of PM compromised flies respond immunologically to the presence of parasites earlier in the infection process than in wild-type individuals that house an intact PM. This irregular immune response presents limited trypanocidal activity, thus resulting in the parasite-susceptible phenotype exhibited by these flies [51].

Significance

The Bacteria that make up the tsetse microbiome play an important role in many biological processes within the host, including vector competence. The study provided a more complete picture of the tsetse gut microbiota and a more specific analysis of the diversity of the non-maternally inherited, environmentally acquired gut bacteria. Examining the gut microbiome by fly location can aid in the determination of how and where in the environment tsetse acquire gut symbionts beyond those that are maternally transmitted [21-24]. I hypothesize that these microorganisms contribute to the mediation of their host's competency as a vector of pathogenic trypanosomes. The findings from this study will contribute to the understanding of the dynamics of trypanosome infection establishment and transmission in wild tsetse by providing more information on the variability of the tsetse gut microbiome as it relates to the observed highly variable vector competence and fly species [10]. Additionally, this study will contribute our repertoire of culturable insect gut bacteria that may potentially find application in microbe-driven modulation of vector competence in tsetse and related flies. These findings can find application in the design of tsetse vector control strategies using paratransgenic microbes to halt the transmission of trypanosomes within the tsetse fly. Future studies will aim at further investigating the relationship between host vector competence and the presence of environmentally acquired microbes.

Figures

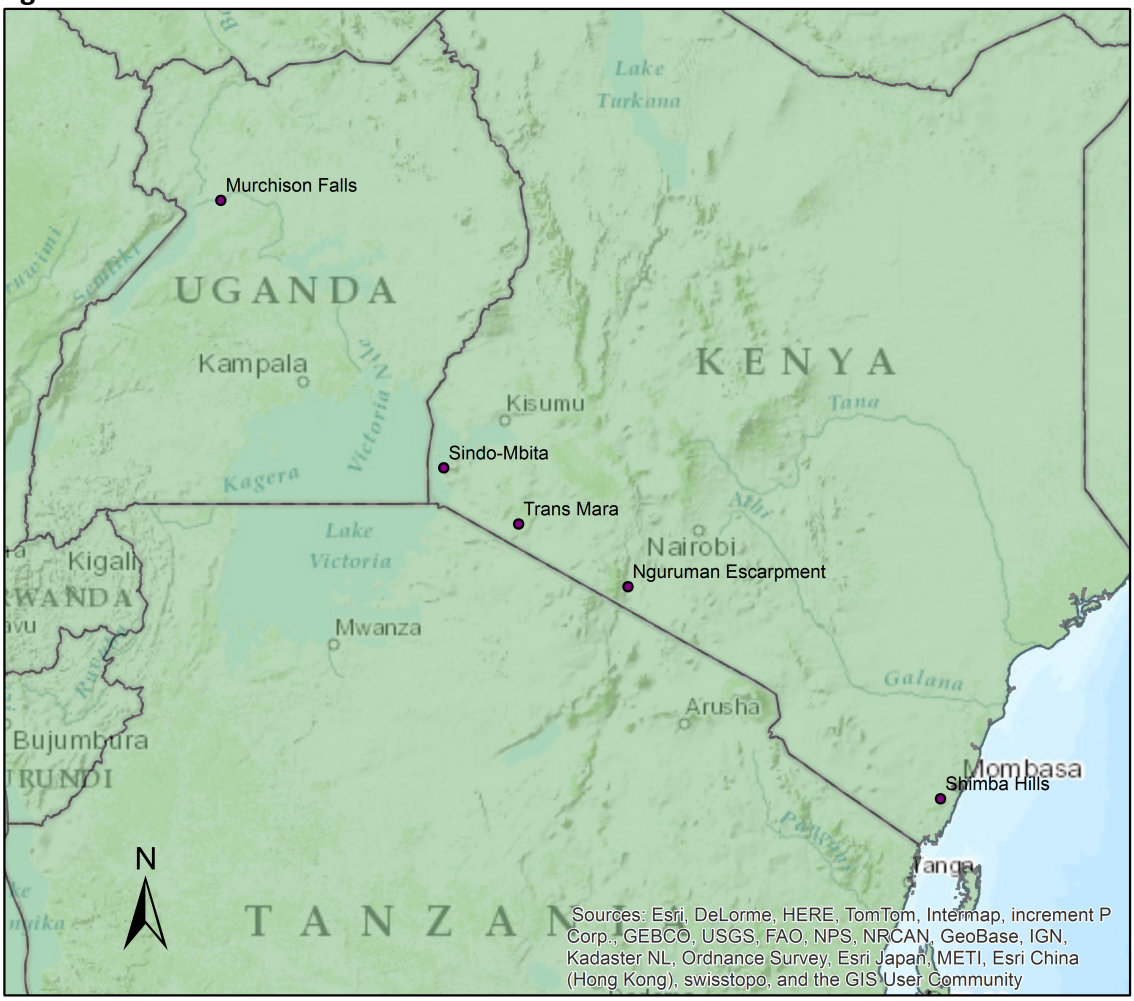

Figure 1: Map displaying 5 sites at which tsetse flies were collected for this study.

Table 1: Number and Geographic origin of total study collection							
Location	*G. palledipies*	*G. brevipalpis*	*G. austeni*	*G. fusciplueris*	*G. fuscipies*	*G. morsitans*	Total
Shimba Hills	26	12	13	0	0	0	51
Trans Mara	0	0	0	12	0	0	12
Ungoye(Western)	0	0	0	0	4	0	4
Uganda	47	0	0	0	0	17	64
Nguruman	52	0	0	0	0	0	52
Infected	9	3	7	3	0	2	24
Total	125	12	13	12	4	17	183

Table 2: Number and geographic origin of the four Glossina species analyzed by culture dependent methods					
Glossina species	*G. pallidipes*	*G. brevipalpis*	*G. fuscipleuris*	*G. f. fuscipes*	Total
Location					
Shimba Hills (Kwale)	11	3	0	0	14
Trans Mara (Lolgorian)	0	0	3	0	3
Western Kenya (Ungoye)	0	0	0	3	3
Total	11	3	3	3	20

Table 3 Number and Geographic Origin of *Glossinia pallidipes* Species Analyzed and (Rate of Infection)				
	Nguruman Escarpment	Shimba Hills (Kwale)	Murchison Falls (Uganda)	Total
Total(% infected)	52(7.7)	15(20)	47(4.3)	114

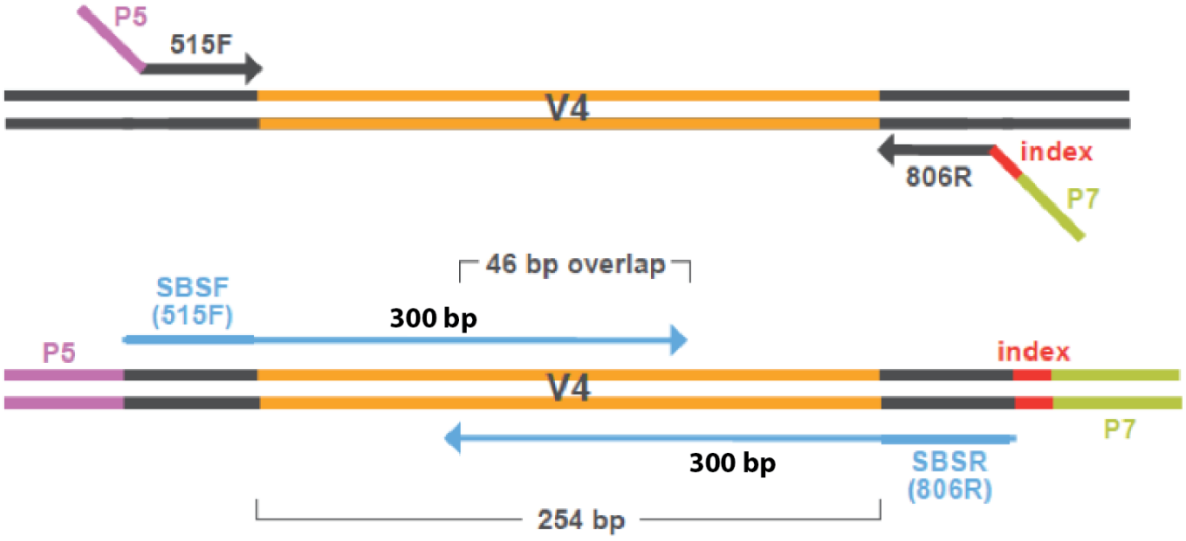

Figure 2: V4 region was amplified for each sample using 515F and 806R primers, tailed with P5 and P7 respectively. Paired read 300 sequencing gives a full-length 354 bp fragment of V4 with a 346 bp overlap.

Table 4: Bacterial isolates by the culture dependent method per individual			
Location	Species	Fly ID	Isolate
SHIMBA HILLS			
	G. pallidipes	1	*Staphylococcus sp.*
		2	*Xylella sp. , Staphylococcus sp.*
		3	*Staphylococcus sp.*
		4	*Agrococcus sp., Arthrobacter sp.*
		5	*Bacillus sp.*
		8	*Staphylococcus sp., Bacillus sp.*
		10	*Staphylococcus sp.*
		11	*Staphylococcus sp.*
		12	*Enterobacter sp., Bacillus sp.*
		13	*Exiguobacterium sp.*
		14	*Bacillus sp., Exiguobacterium sp.*
	G. brevipalpis	6	*Bacillus sp.*
		7	*Bacillus sp.*
		9	*Bacillus sp.*
LOLGORIEN			
	G. fuscipleuris	15	*Staphylococcus sp., Bacillus sp.*
		16	*Kocuria sp., Microbacterium sp.*
		17	*Exiguobacterium sp., Sinomonas sp.*
UNGOYE			
	G. fuscipes	18	*Bacillus sp., Oceanimonas sp., Microbacterium sp., Staphylococcus sp.*
		19	*Arthrobacter sp., Aeromonas sp., Providencia sp., Bacillus sp.*
		20	*Pantoea sp.*

Table 5: Bacterial Isolates by the culture dependent method				
Location/Fly Species	Phylogenetic affiliation	Closest Relative According to BLAST	No. of isolates	E-value
Shimba Hills				
G. palledipies	*Staphylococcus sp.*	*Staphylococcus saprophyticus* strain CD250Y	2	0
	Staphylococcus sp.	*Staphylococcus sp.* MOLA 313	1	0
	Staphylococcus sp.	*Staphylococcus haemolyticus* strain B-16	1	0
	Staphylococcus sp.	*Staphylococcus sciuri* strain CD97	1	0
	Staphylococcus sp.	*Staphylococcus saprophyticus* strain CD250Y	2	0
	Xylella sp.	Uncultured *Xylella sp.* clone I10-307	2	0
	Agrococcus sp.	*Agrococcus sp.* PLB053	1	0
	Arthrobacter sp.	*Arthrobacter sp.* OS-10C	1	0
	Enterobacter sp.	*Enterobacter cloacae* strain Y219	2	0
	Exiguobacterium sp.	*Exiguobacterium acetylicum* strain N5	1	0
	Exiguobacterium sp.	*Exiguobacterium indicum* strain B3	1	0
	Exiguobacterium sp.	*Exiguobacterium sp.* P11	1	0
	Bacillus sp.	*Bacillus aryabhattai* strain fwz21	2	0
	Bacillus sp.	*Bacillus sp.* NQ18	1	0
	Bacillus sp.	*Bacillus flexus* strain NM25	1	0
	Bacillus sp.	*Bacillus sp.* Q-5	1	0
	Bacillus sp.	*Bacillus thuringiensis* serovar kurstaki str. HD73	1	2.5E-143
	Bacillus sp.	*Bacillus subtilis* strain CICC10200	1	0
G. brevipalpis	*Bacillus sp.*	*Bacillus megaterium* strain IARI-BC-13	1	0
	Bacillus sp.	*Bacillus aryabhattai* strain fwzb8	2	0
Trans Mara				
G. fuscipes	*Staphylococcus sp.*	*Staphylococcus haemolyticus* strain B-16	1	0
	Bacillus sp.	*Bacillus horneckiae* type strain DSM 23495T	1	0
	Bacillus sp.	*Bacillus horneckiae* strain RCTy1	1	4.25283E-74

	Oceanimonas sp.	*Oceanimonas sp.* GK1	1	0.068421
	Microbacterium sp.	*Microbacterium foliorum* strain N1-12	1	0
	Arthrobacter sp.	*Arthrobacter sp.* R-10(2010)	1	9.6916E-140
	Aeromonas sp.	*Aeromonas sp.* DH-6	1	0
	Providencia sp.	*Providencia rettgeri* strain ALK058	1	0
	Pantoea sp.	*Pantoea dispersa* strain 1413	1	0
Ungoye (Western Kenya)				
G. fusciplueris	*Staphylococcus sp.*	*Staphylococcus hominis* subsp. novobiosepticus strain ALK519	1	0
	Bacillus sp.	*Bacillus sp.* FJAT-17442	1	0
	Kocuria sp.	*Kocuria palustris* strain Abk-6	1	0
	Microbacterium sp.	*Microbacterium sp.* U1370-101126-SW187	1	0
	Sinomonas sp.	*Sinomonas atrocyanea* strain JN170	1	0
	Exiguobacterium sp.	*Exiguobacterium sp.* AFB-18	1	0
	n/a	Uncultured bacterium clone bb1w5	2	0

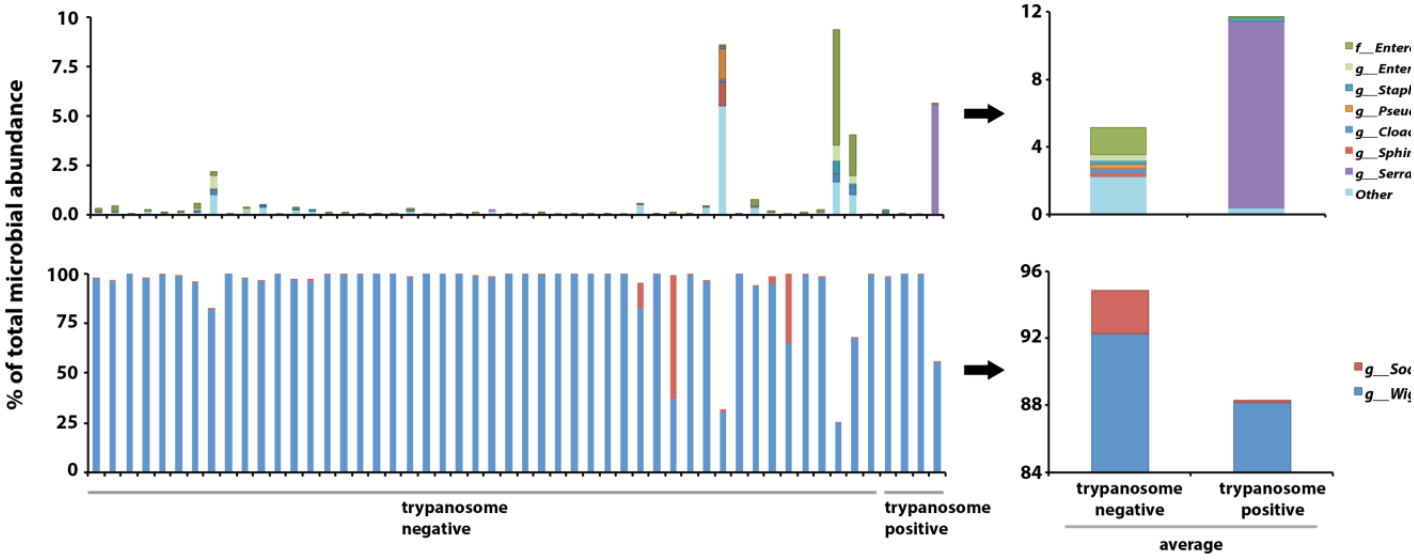

Figure 3: Microbial abundance by % OTUs in *G. pallidipes* flies caught in Nguruman, Kenya. Charts are divided to display abundance data for maternally inherited symbionts and environmentally acquired symbionts. The averages show that the composition of the midgut, both environmentally acquired and maternally inherited, differs between trypanosome infected and uninfected flies.

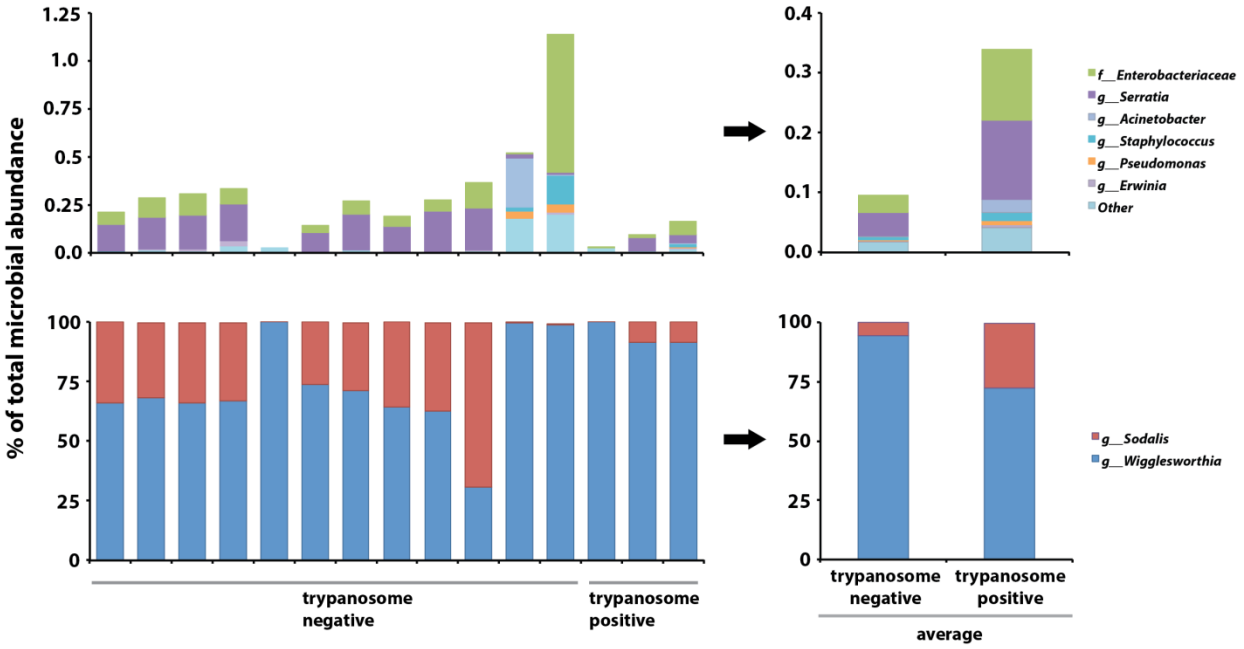

Figure 4: Microbial abundance by % OTUs in *G. pallidipes* flies caught in Shimba Hills, Kenya. Charts are divided to display abundance data for maternally inherited symbionts and environmentally acquired symbionts. There is variation in the composition of the midgut microbiome by individual. The averages show that the composition of the midgut, both environmentally acquired and maternally inherited, differs between trypanosome infected and uninfected flies.

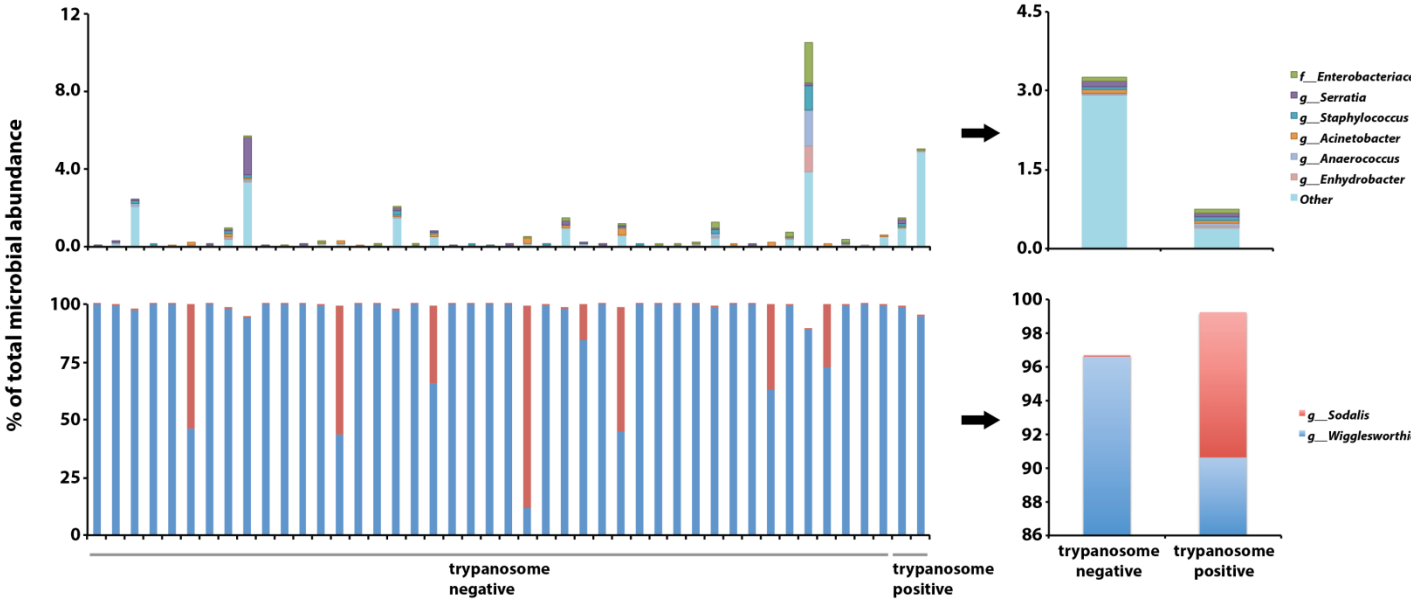

Figure 5: Microbial abundance by % OTUs in G. pallidipes flies caught in Murchison Falls, Uganda. Charts are divided to display abundance data for maternally inherited symbionts and environmentally acquired symbionts. There is variation in the composition of the midgut microbiome by individual, with some containing high densities of *Sodalis*. The averages show that the composition of the midgut, both environmentally acquired and maternally inherited, differs between trypanosome infected and uninfected flies.

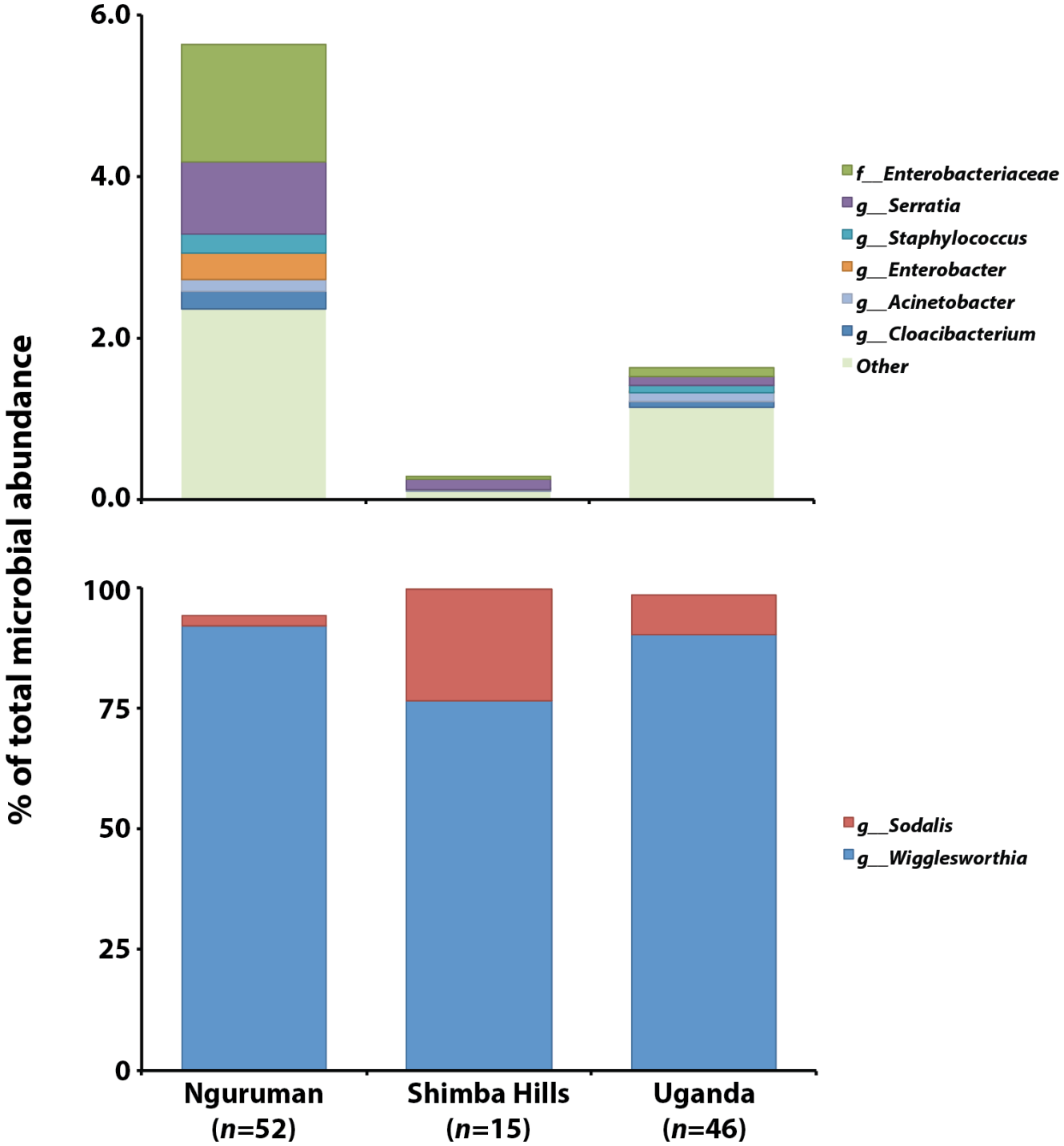

Figure 6: Average microbial abundance by % OTUs for *G. pallidipes* samples caught in three locations. Chart is divided to display abundance data for maternally inherited symbionts and environmentally acquired symbionts. The densities of both maternally inherited and environmentally acquired bacteria differ by trapping site.

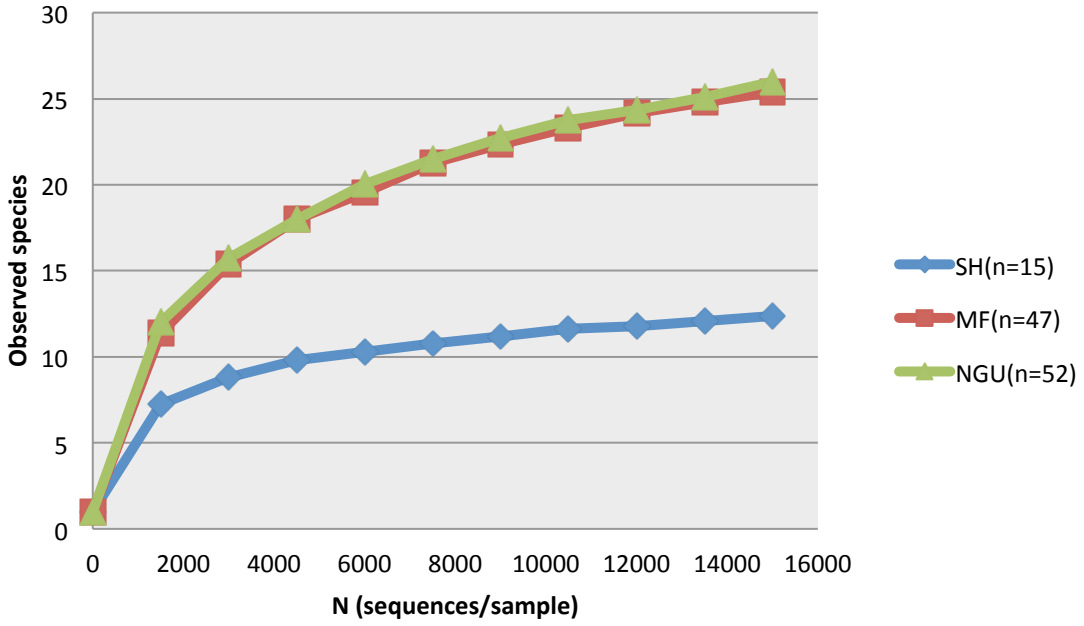

Figure 7: Rarefaction curve for *G. pallidipes* samples collected from three trapping locations. The rarefaction curve levels off at 6,000 sequences per sample, indicating that an adequate sequencing depth and OTU discovery is achieved at this value.

Table 6. Student's t test to measure species richness between study sites					
Group 1	Group 2	Group 1 mean	Group 2 mean	t statistic	p-value
Shimba_Hills	Uganda	12.38±2.45	25.5±28.73	-1.6785332	0.261
Shimba_Hills	**Nguruman**	**12.38±2.45**	**25.05±17.11**	**-2.710701748**	**0.024**
Uganda	Nguruman	25.5±28.73	25.05±17.11	0.087681828	1

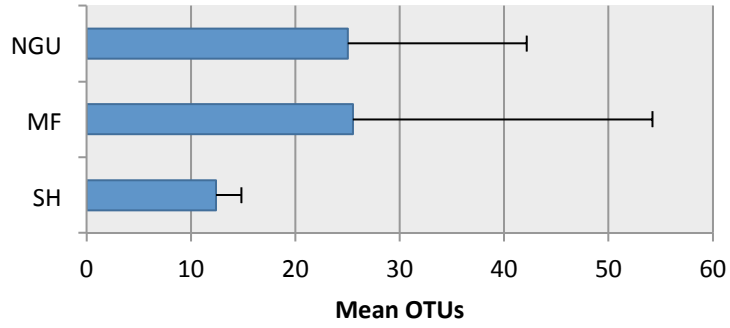

Figure 8: Mean OTUs per *G. pallidipes* per geographic location. This represents the average number of bacterial genera assigned from samples collected at each site. There was a high amount of variation in the number of OTUs per sample. Overall, the Murchison Falls site in Uganda had the highest mean OTUs.

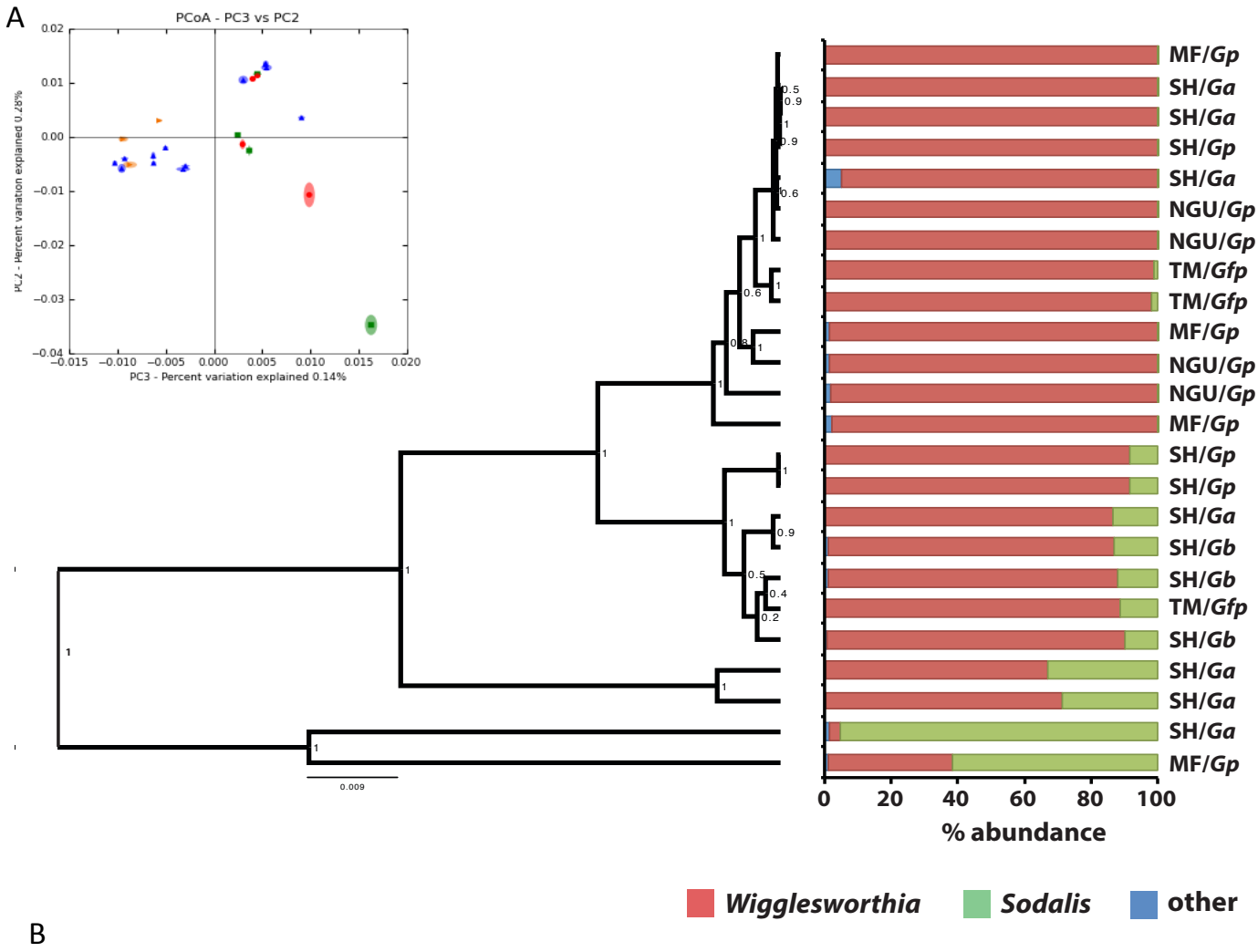

Figure 9: Phylogenetic analysis of all trypanosome infected samples. Tree was established by measuring alpha diversity of each sample using the observed species metric, which is based on a count of the unique OTUs in a sample. Each sample is represented by % abundance of the total microbial composition of the sample.

Table 7. Student's t test to measure β-diversity between study sites		
Group	t statistic	Nonparametric p-value (Bonferroni-corrected)
Shimba Hills vs. Uganda	**4.093919463**	**0.028**
Shimba Hills vs. Nguruman	-2.880723694	0.112
Uganda vs. Nguruman	-0.709294559	1

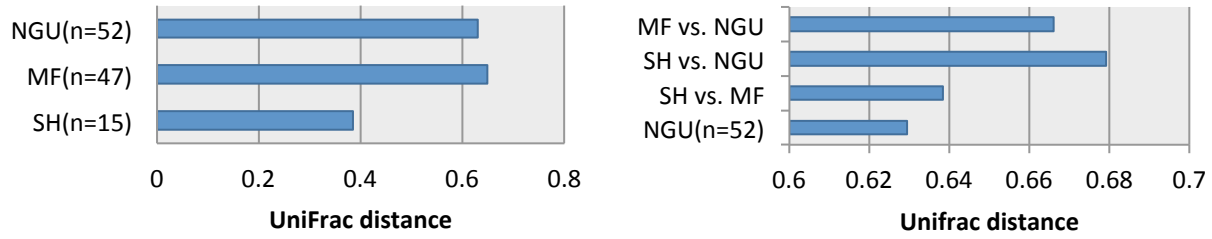

Figure 10 a & b: Unifrac distance within locations and between locations. Unifrac distance is a measure of dissimilarity between samples using β-diversity.

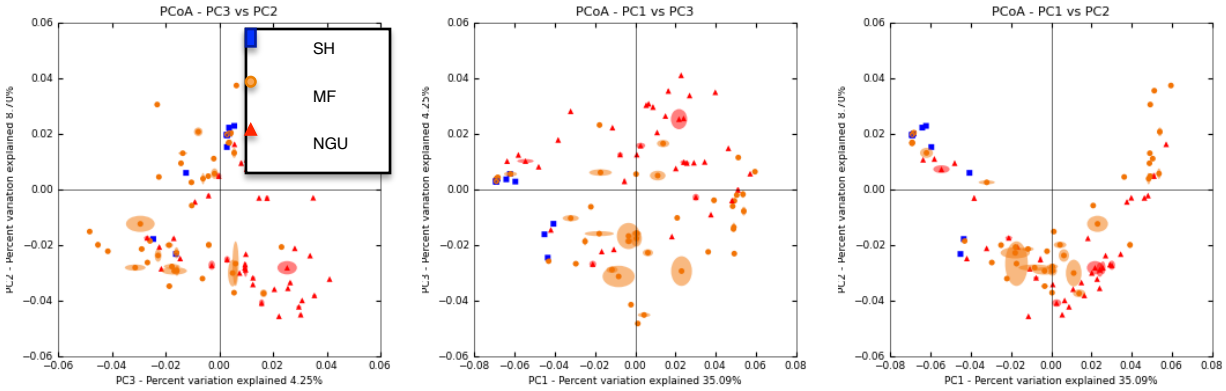

Figure 11: Principal coordinate plots depicting the bacterial composition differences between locations measured by the unweighted UniFrac distance metric, which is based on phylogenetic tree branch length. Samples cluster by geographic location.

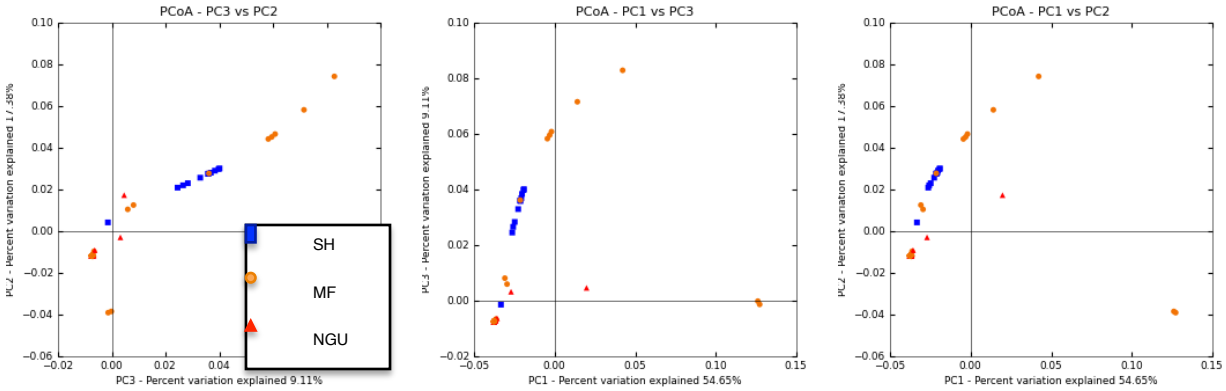

Figure 12: Principal coordinate plots depicting the bacterial composition differences of samples from different locations measured by the Bray-Curtis (unweighted) method. Samples cluster by geographic location.

Table 8: Comparison of Culture dependent and independent data					
	No. Genera Cultured	No. total genera	Unweighted	Weighted	Genera (% OTUs)
NGU	2	64	3.125	0.31	*Staphylococcus*(0.24), *Bacillus*(0.07)
SH	1	8	12.5	0.01	*Staphylococcus*(0.01)
MF	3	50	6.0	0.12	*Staphylococcus*(0.10), *Bacillus*(0.01), *Actinomyces*(0.01)
Average	2	53	3.773584906	0.15	*Staphylococcus*(0.12), *Bacillus*(0.03)

Supplemental information:

List of PCR primers used:

SI table 1. Primers used in this study.

gene	primer sets	Tm (°C)
tsetse β-tubulin	F - 5'-ACGTATTCATTTCCCTTTGG-3'	54
	R - 5'-AATGGCTGTGGTGTTGGACAAC-3'	55
eubacterial 16s rRI	F - 5'-AGAGTTTGATCCTGGCTCAG-3'	54
	R - 5'-GGTTACCTTGTTACGACTT-3'	55
kinetoplastid ITS1	F - 5'-GCGTTCAAAGATTGGGCAAT-3'	54
	R - 5'-CGCCCGAAAGTTCACC-3'	55

Lat/long coordinates of trapping sites:

Name	Longitude	Latitude
Trans Mara	34.89773333	-1.21365
Nguruman Escarpment	36.07583333	-1.888783333
Sindo-Mbita	34.09166667	-0.609016667
Murchison Falls	31.68622	2.27747
Shimba Hills	39.444631	-4.177326

Master data sheet: KARI-TRC Sample Working Log.xlsx

References

1. Louis, F.J., P.P. Simarro, and J. Jannin, *[Status of sleeping sickness in 2003].* Med Trop (Mars), 2003. **63**(3): p. 228-30.
2. Simarro, P.P., et al., *Estimating and Mapping the Population at Risk of Sleeping Sickness.* PLoS Negl Trop Dis, 2012. **6**(10): p. e1859.
3. Reinhardt, E., *Travailler ensemble: la mouche tsé-tsé et la pauvreté rurale.* Chronique ONU, ONU Editor, September, 2002.
4. Organization, W.H., *Human African trypanosomiasis (sleeping sickness): epidemiological update.* Wkly Epidemiol Rec, 2006. **81**(8): p. 71-80.
5. Kabayo, J.P., *Aiming to eliminate tsetse from Africa.* Trends in Parasitology, 2002. **18**(11): p. 473-475.
6. Barrett, M.P., *The rise and fall of sleeping sickness.* Lancet, 2006. **367**(9520): p. 1377-8.
7. Matovu, E., et al., *Drug resistance in Trypanosoma brucei spp., the causative agents of sleeping sickness in man and nagana in cattle.* Microbes Infect, 2001. **3**(9): p. 763-70.
8. Delespaux, V. and H.P. de Koning, *Drugs and drug resistance in African trypanosomiasis.* Drug Resistance Updates, 2007. **10**(1–2): p. 30-50.
9. Abd-Alla, A.M., et al., *Improving Sterile Insect Technique (SIT) for tsetse flies through research on their symbionts and pathogens.* J Invertebr Pathol, 2012.
10. Solano, P., S. Ravel, and T. de Meeûs, *How can tsetse population genetics contribute to African trypanosomiasis control?* Trends in Parasitology, 2010. **26**(5): p. 255-263.
11. Attardo, G.M., et al., *Analysis of milk gland structure and function in Glossina morsitans: Milk protein production, symbiont populations and fecundity.* Journal of Insect Physiology, 2008. **54**(8): p. 1236-1242.
12. Chen, X., S. Li, and S. Aksoy, *Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia.* J Mol Evol, 1999. **48**(1): p. 49-58.
13. Weiss, B.L., J. Wang, and S. Aksoy, *Tsetse immune system maturation requires the presence of obligate symbionts in larvae.* PLoS Biol, 2011. **9**(5): p. e1000619.
14. Weiss, B.L., M. Maltz, and S. Aksoy, *Obligate symbionts activate immune system development in the tsetse fly.* J Immunol, 2012. **188**(7): p. 3395-403.
15. Aksoy, S., X. Chen, and V. Hypsa, *Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae).* Insect Mol Biol, 1997. **6**(2): p. 183-90.
16. Matthew, C.Z., et al., *The rapid isolation and growth dynamics of the tsetse symbiont Sodalis glossinidius.* FEMS Microbiology Letters, 2005. **248**(1): p. 69-74.
17. Dale, C. and I. Maudlin, *Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans.* International Journal of Systematic Bacteriology, 1999. **49**(1): p. 267-275.
18. Medlock, J., et al., *Evaluating paratransgenesis as a potential control strategy for african trypanosomiasis.* PLoS Negl Trop Dis, 2013. **7**(8): p. e2374.
19. Hamidou Soumana, I., et al., *Population dynamics of Glossina palpalis gambiensis symbionts, Sodalis glossinidius, and Wigglesworthia glossinidia, throughout host-fly development.* Infect Genet Evol, 2013. **13**: p. 41-8.

20. Alam, U., et al., *Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly Glossina morsitans.* PLoS pathogens, 2011. **7**(12): p. e1002415.
21. Geiger, A., et al., *First isolation of Enterobacter, Enterococcus, and Acinetobacter spp. as inhabitants of the tsetse fly (Glossina palpalis palpalis) midgut.* Infect Genet Evol, 2009. **9**(6): p. 1364-70.
22. Lindh, J.M. and M.J. Lehane, *The tsetse fly Glossina fuscipes fuscipes (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts.* Antonie Van Leeuwenhoek, 2011. **99**(3): p. 711-20.
23. Geiger, A., et al., *Bacterial diversity associated with populations of Glossina spp. from Cameroon and distribution within the Campo sleeping sickness focus.* Microb Ecol, 2011. **62**(3): p. 632-43.
24. Hamidou Soumana, I., et al., *The bacterial flora of tsetse fly midgut and its effect on trypanosome transmission.* J Invertebr Pathol, 2012.
25. Dransfield, R., et al., *The use of odour attractants for sampling Glossina pattidipes Austen (Diptera: Glossinidae) at Nguruman, Kenya.* Bull. ent. Res. **76**: p. 607-619.
26. Torr, S.J., T.N. Mangwiro, and D.R. Hall, *The effects of host physiology on the attraction of tsetse (Diptera: Glossinidae) and Stomoxys (Diptera: Muscidae) to cattle.* Bull Entomol Res, 2006. **96**(1): p. 71-84.
27. Fukatsu, T., et al., *Bacterial endosymbiont of the slender pigeon louse, Columbicola columbae, allied to endosymbionts of grain weevils and tsetse flies.* Appl Environ Microbiol, 2007. **73**(20): p. 6660-8.
28. Desquesnes, M., et al., *Detection and identification of< i> Trypanosoma</i> of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA.* International journal for parasitology, 2001. **31**(5): p. 610-614.
29. Caporaso, J.G., et al., *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4516-22.
30. Aksoy, E., et al., *Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota.* Applied and Environmental Microbiology, 2014.
31. Snyder, A.K. and R.V. Rio, *Interwoven Biology of the Tsetse Holobiont.* Journal of bacteriology, 2013. **195**(19): p. 4322-4330.
32. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
33. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
34. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
35. Wang, J., B.L. Weiss, and S. Aksoy, *Tsetse fly microbiota: form and function.* Front Cell Infect Microbiol, 2013. **3**: p. 69.
36. Wamwiri, F.N., et al., *Wolbachia, Sodalis and trypanosome co-infections in natural populations of Glossina austeni and Glossina pallidipes.* Parasit Vectors, 2013. **6**(1): p. 232.

37. Pais, R., et al., *The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly.* Applied and environmental microbiology, 2008. **74**(19): p. 5965-5974.
38. Hamidou Soumana, I., et al., *The bacterial flora of tsetse fly midgut and its effect on trypanosome transmission.* J Invertebr Pathol, 2013. **112 Suppl**: p. S89-93.
39. Maltz, M.A., et al., *OmpA-mediated biofilm formation is essential for the commensal bacterium Sodalis glossinidius to colonize the tsetse fly gut.* Appl Environ Microbiol, 2012. **78**(21): p. 7760-8.
40. Hu, Y. and S. Aksoy, *An antimicrobial peptide with trypanocidal activity characterized from Glossina morsitans morsitans.* Insect Biochem Mol Biol, 2005. **35**(2): p. 105-15.
41. Wang, Y., et al., *Dynamic gut microbiome across life history of the malaria mosquito Anopheles gambiae in Kenya.* PLoS One, 2011. **6**(9): p. e24767.
42. Wong, C.N., P. Ng, and A.E. Douglas, *Low-diversity bacterial community in the gut of the fruitfly Drosophila melanogaster.* Environ Microbiol, 2011. **13**(7): p. 1889-900.
43. Rio, R.V., et al., *Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora.* Proc Biol Sci, 2006. **273**(1588): p. 805-14.
44. Bonnay, F., et al., *big bang gene modulates gut immune tolerance in Drosophila.* Proc Natl Acad Sci U S A, 2013. **110**(8): p. 2957-62.
45. Boissiere, A., et al., *Midgut microbiota of the malaria mosquito vector Anopheles gambiae and interactions with Plasmodium falciparum infection.* PLoS Pathog, 2012. **8**(5): p. e1002742.
46. Bando, H., et al., *Intra-specific diversity of Serratia marcescens in Anopheles mosquito midgut defines Plasmodium transmission capacity.* Sci Rep, 2013. **3**: p. 1641.
47. Bahia, A.C., et al., *Exploring Anopheles gut bacteria for Plasmodium blocking activity.* Environmental microbiology, 2014.
48. Weiss, B. and S. Aksoy, *Microbiome influences on insect host vector competence.* Trends in Parasitology, 2011. **27**(11): p. 514-522.
49. Geiger, A., et al., *Sodalis glossinidius (Enterobacteriaceae) and vectorial competence of Glossina palpalis gambiensis and Glossina morsitans morsitans for Trypanosoma congolense savannah type.* Current microbiology, 2005. **51**(1): p. 35-40.
50. Hamidou Soumana, I., et al., *The transcriptional signatures of Sodalis glossinidius in the Glossina palpalis gambiensis flies negative for Trypanosoma brucei gambiense contrast with those of this symbiont in tsetse flies positive for the parasite: Possible involvement of a Sodalis-hosted prophage in fly Trypanosoma refractoriness?* Infect Genet Evol, 2014. **24C**: p. 41-56.
51. Weiss, B.L., et al., *Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers.* PLoS Pathog, 2013. **9**(4): p. e1003318.