Blood meal analysis of tsetse flies (*Glossina pallidipes*: Glossinidae) reveals a reduction in host fidelity when feeding on domestic compared to wild hosts

Manun Channumsin[1, 2*], Marc Ciosi[3, 4], Dan Masiga[4], Harriet Auty[5], C. Michael Turner[6,7] and Barbara K. Mable[1]

*Correspondence: manun_wo@rmutto.ac.th

[1] Institute of Biodiversity, Animal Health and Comparative Medicine (BAHCM), Graham Kerr Building, University of Glasgow, University Place, Glasgow G12 8QQ, UK

[2] Faculty of Veterinary Medicine, Rajamangala University of Technology Tawan-Ok, Chonburi, 20110, Thailand

[3] Institute of Molecular, Cell and Systems Biology (MCSB), University of Glasgow, University Place, Glasgow G12 8QQ, UK

[4] International Centre of Insect Physiology and Ecology (ICIPE), P.O. Box 30772, 00100 Nairobi, Kenya

[5] Future Farming Systems, Epidemiology Research Unit, Scotland's Rural College, Inverness, IV2 5NA, UK

[6] Institute of Infection, Immunity and Inflammation, Sir Graeme Davis Building, University of Glasgow, University Place, Glasgow G12 0PT, UK

[7] The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK

Abstract

Background

Both male and female tsetse flies, haematophagous insects, transmit trypanosomes between hosts and are the cyclical vectors for Human African Trypanosomiasis (HAT) and Animal African Trypanosomiasis (AAT). Trypanosomes responsible for AAT can be transmitted by tsetse between wild animals and livestock. However, the degree of connectivity between the sylvatic and domestic cycles is unknown. The objectives of this study were to investigate patterns of host feeding in relationship to trypanosome prevalence among Kenyan populations of *G. pallidipes* at the livestock-wildlife interface.

Methodology/Principal Findings

Sources of blood meals of *Glossina pallidipes* were identified by polymerase chain reaction amplification and sequencing of the mitochondrial cytochrome b gene and compared with previous characterization of trypanosome prevalence from the same flies. In the Nguruman region in southern Kenya, the majority (98%) of the 148 flies for which dominant hosts could be resolved fed on single host species and only a single fly had fed on a domestic host; intriguingly this was the only fly confirmed to have fed on cattle. In contrast, in the Shimba Hills region (South Coast), multiple host feeding was more common: 42% inside a fenced wildlife protected area, where 35% of dominant hosts were domestic animals or humans, compared with 62% from traps set along the border to an adjacent village, which was dominated by domestic hosts (77%). Across sites, 44% of flies tested positive for trypanosomes (28/50 domestic hosts; 78/193 wild hosts). Multiple Correspondence Analysis revealed strong correlations between feeding pattern, host type and site but these were resolved along a different dimension than trypanosome status.

Conclusions/Significance

Our results suggest that host fidelity when feeding on wild hosts in *G. pallidipes* could reduce risk of transmission of trypanosomes to domestic hosts in interface areas and emphasise the importance of considering vector behaviour when designing management interventions.

Author Summary

Tsetse flies are the main vectors transmitting trypanosomes, which cause disease in both humans and animals. Since tsetse flies feed on a wide range of vertebrate hosts, there is the potential for transmission between wild and domestic animals in regions where their ranges overlap. In this study, we used molecular methods to determine the hosts fed on by tsetse flies sampled from three sites in Kenya at the wildlife-livestock interface. In areas where wildlife dominated, tsetse tended to feed on single host species, whereas in areas with more domesticated animals, they tended to feed on multiple hosts. These results suggest either that tsetse flies get interrupted more while feeding on domestic hosts or that they prefer to feed on wildlife and so switch hosts more often when feeding on less desirable hosts. Using data from a previous study on the same samples, we found that trypanosome prevalence was not correlated with the type or number of hosts fed on. These results have important implications for understanding the risk of transmission between wildlife and livestock in regions bordering protected areas but the high host fidelity for wild hosts suggests that tsetse feeding preferences could reduce risks of disease transmission to livestock.

Introduction

In sub-Saharan Africa, changes in land use increase encroachment of domestic livestock into areas that are primarily managed to conserve wildlife. This increases risks that livestock will be exposed to a wider range of parasites, with potentially important consequences for disease burden and control. Wildlife can represent 'reservoir communities' (Haydon *et al.* 2002; Viana *et al.* 2014) for multi-host pathogens that could spill-over into domesticated animals, who could show more severe clinical disease, given limited opportunity for host-pathogen coevolution in novel hosts. This could be particularly true for vector-mediated transmission, where, even if fences are used to reduce contact between domestic and wild hosts, movement of the vectors could still facilitate parasite sharing across interface areas. Analysis of blood meals provides a powerful tool for assessing whether haematophagous insects cross-feed on different types of hosts (Nyingilili *et al.* 2016; Omondi *et al.* 2015; Snow *et al.* 1988). Although this approach has typically focused on vectors of particular pathogens, it has also been proposed that blood-sucking flies that do not themselves transmit the focal parasites could be used as "flying syringes" to investigate blood-borne pathogen diversity (Bitome-Essono *et al.* 2017; Hoffmann *et al.* 2016). Rather surprisingly, blood meals have not been used to investigate the intraspecific genetic diversity of the hosts, which could be used to infer patterns of connectivity to feed into epidemiological models.

One particularly complex system where this could be important to understand is trypanosome-mediated diseases transmitted by tsetse flies in East Africa. Although there are multiple species of tsetse flies that can transmit multiple species of trypanosomes, *Glossina pallidipes* is the most economically and medically important

species in Kenya (England & Baldry 1972b), because it transmits both the animal (Animal African Tryanosomiasis, AAT) and human (Human African Trypanosomiaasis, HAT) disease-causing species. Wild animals have been reported as reservoir hosts both for AAT (Anderson *et al.* 2011b; OIE 2013) and HAT (Geigy *et al.* 1973a; Geigy *et al.* 1971; Welburn *et al.* 2001; Welburn *et al.* 2005) but the extent of risk posed to domesticated animals has been hard to assess, partly due to the difficulty of directly screening wildlife for the presence of trypanosomes (but see (Geigy & Kauffman.M 1973; Geigy *et al.* 1973b; Heisch *et al.* 1958)).

Tsetse flies are generalist feeders on a wide variety of vertebrate host species (Weitz 1963), including: dogs (Eloy & Lucheis 2009; Lisulo *et al.* 2014); cats (Clausen *et al.* 1998; Maudlin *et al.* 2004); horses (Maudlin *et al.* 2004); chickens (Clausen *et al.* 1998); ruminants (Clausen *et al.* 1998; Maudlin *et al.* 2004; Nyawira 2009; Turner 1987); antelopes (Turner 1987); elephants (Muturi *et al.* 2011b; Nyawira 2009); warthogs (Nyawira 2009); and reptiles (Turner 1987). There is thus high potential for vector-mediated connection between parasite sylvatic and domestic cycles in wildlife-livestock interface areas. However, the frequency with which tsetse vectors cross-feed on different types of hosts where they occur sympatrically has not been clearly established and so the relative risks of increased trypanosome infections in livestock living near wildlife remains a critical gap in knowledge (Auty *et al.* 2016b). Although three trypanosome species are traditionally associated with the disease in livestock (*T. brucei*, *T. congolense*, and *T. vivax*), a higher diversity has been identified in wildlife (Auty *et al.* 2012), which could potentially increase risks of disease if transmission from wildlife to domesticated animals and humans is common.

Complex interactions among trypanosomes, vertebrate hosts and tsetse flies likely increase the difficulty of controlling and eradicating trypanosomes but few studies

have attempted to combine investigation of host-feeding patterns in relation to trypanosome infection with intrinsic factors of tsetse flies distributed in different regions. Identification of hosts through blood meal analyses is a highly useful tool that has been used to predict host preferences and feeding behaviours across a wide range of vectors (Gariepy *et al.* 2012; Haouas *et al.* 2007; Kent 2009). A commonly used approach has been to use polymerase chain reaction (PCR)-based techniques to amplify and sequence host DNA from blood meal contents in the guts of fed flies. This has largely been based on mitochondrial genes due to their high copy number and the extensive databases available due to their use as universal markers for DNA barcoding (Hebert *et al.* 2003; Lah *et al.* 2012; Pradhan *et al.* 2015). For example, Muturi *et al*. (Muturi *et al.* 2011b) surveyed blood meals of *G. swynertoni* from Tanzania and *G. pallidipes* from Kenya and Uganda using primers targeting the mitochondrial cytochrome b (cytB) and cytochrome oxidase I (COI) genes. African elephants, African buffalo, warthogs, lizards, giraffes, spotted hyenas baboons and Nile Monitor lizards were identified as sources of blood contained in tsetse abdomens. However, this was not compared to the relative availability of hosts to determine whether there were biases in host preferences. In the Serengeti ecosystem in Tanzania, which holds a high number and wide range of wild animals, a thorough investigation of blood meal composition in tsetse flies based on sequencing of the cytB gene compared to surveys of host density revealed strong preferences for particular hosts, which varied by species of fly (G*. swynnertoni* vs *G. pallidipes* (Auty *et al.* 2016a). This clearly demonstrated the value of relating feeding patterns to host densities but also the usefulness of using DNA sequencing to identify the range of diversity of hosts present. However, trypanosome prevalence was not quantified in that study and no domestic hosts were identified in the blood

meals; so, relative host preferences for wildlife compared to livestock could not be determined.

The relative risks of transmission between host types could depend not only on differences in the feeding behaviours of flies but also on their relative susceptibility to different species of trypanosomes. Since both sexes of tsetse flies feed on blood meals, they are both at risk of exposure to trypanosomes from infected vertebrate hosts after they emerge from their puparium (Jackson 1946). Relative prevalence and transmissibility could also vary by sex, age or species of fly. For example, higher prevalence of trypanosomes in females could be due to their longer lifespan compared to males (Nash 1936), which could increase both the number of hosts that they feed on and the length of time that they harbour trypanosomes. However, the consequences for transmission could also vary by species of trypanosomes. (e.g Peacock *et al*. (2012)).. Moreover, host feeding activity could be influenced by tsetse-specific factors. For example, old females are thought to be less active than younger ones, especially those in their first three weeks of their life (Jackson 1946). However, differences between sexes and age groups can also differ by species of tsetse (Moloo *et al.* 1992) (Wamwiri *et al.* 2013) so it is important to assess feeding behaviours separately for each species.

Furthermore, the availability of preferred and nonpreferred hosts could alter feeding patterns in a way that could affect the risk of transmission, even within species of tsetse. For example, the feeding-interval time for *G. pallidipes* is approximately two days but that of males has been found to be longer than that for females (Turner 1987) and times can extend to three days in mature tsetse flies seeking their preferred hosts (Bouyer *et al.* 2007). More frequent feeding might occur if flies are disrupted while feeding or if they abandon a host that they perceive to be unsuitable.

The dominance of nonpreferred hosts in a particular geographic area could thus result in more frequent host switching and so increased rates of multiple feeding and potentially higher exposure to a diverse range of parasites. In East Africa, *G. pallidipes* is widespread and has been demonstrated to feed on a wide range of hosts, including bovines (Clausen *et al.* 1998; England & Baldry 1972a; Muturi *et al.* 2011b; Okoth *et al.* 2007; Turner 1987), suids (Bett *et al.* 2008; Okoth *et al.* 2007), elephants (Muturi *et al.* 2011b; Nyawira 2009), antelopes (Allsopp *et al.* 1972b) and cattle (Muturi *et al.* 2011b). Warthogs, bushbuck and African buffalo have been suggested as the preferred hosts (Boakye *et al.* 1999; Clausen *et al.* 1998; Langridge *et al.* 1963; Leak 1998; Muturi *et al.* 2011b; Nyawira 2009; Okoth *et al.* 2007) but this varies by geographic region (Auty *et al.* 2016a; Farikou *et al.* 2010; Muturi *et al.* 2011b; Snow *et al.* 1988; Spath 2000) and relative preference for domestic and wild hosts has not been specifically assessed. Disease control measures might need to be specifically targeted to the particular communities of tsetse flies, vertebrate hosts and trypanosome species present in a particular region (Auty *et al.* 2016b). However, what are lacking are comparisons between trypanosome prevalence in geographic regions that differ in community composition of potential hosts, particularly at wildlife-livestock interface areas.

In a previous study, we established that the prevalence of trypanosomes in two regions of Kenya (Nguruman and the Shimba Hills) showed complex relationships with geographic location, tsetse specific factors (age, sex and fly species), species of trypanosome and the presence of an endosymbiont, *Sodalis glossinidius* (Channumsin *et al.* 2018). Here we focus on one of the four species of tsetse found in these regions (G*. pallidipes*) to investigate how blood feeding behavior of flies and host availability affect the potential for cross-transmission of parasites between wild and domestic hosts. Specifically, we aimed to determine whether: 1) tsetse flies

sampled from wildlife-livestock interface areas feed on both domestic and wild hosts; 2) the number of different host species fed on by individual flies varies by site or host type; 3) trypanosome prevalence varies by host type or species. We hypothesise that the type of hosts recently fed on would be associated with trypanosome status: specifically, that tsetse fly populations where multiple-host feeding is more frequent would show a higher prevalence of trypanosomes. We also used the blood meal data to examine mtDNA haplotype diversity for the most commonly fed on hosts, in the context of previous blood meal analyses from Kenya and the bordering Serengeti ecosystem in Tanzania (Auty *et al.* 2016a; Muturi *et al.* 2011a).

Methods

Sampling and tsetse fly characterisation

The *G. pallidipes* samples used in this study for host blood meal analysis were the same as those described in Channumsin et al. (2018), where details of the sampling strategy are provided (but see Supplementary Table 1 for locations of the traps). Tsetse flies were collected from three sites that differ in anticipated levels of abundance of livestock and wildlife. Two sites were sampled in the Shimba Hills National Reserve (Kwale County, in the coastal region of Kenya), which is a relatively small (250 km^2) protected area separated from surrounding agricultural areas by a wildlife fence. There is extensive habitat for tsetse flies, including on the park boundaries. Buffalo Ridge and Zungu Luka (approximately 20 km apart), have different types of vegetation and levels of human activity. Buffalo Ridge is in a fenced wildlife protected area in the middle of a thicket forest, where many tourists visit all year, while Zungu Luka has a woodland type of vegetation, and is located on the border of the park close to an inhabited rural area. In contrast, the Nguruman region contains lowland woodland patches surrounded by open savannah; habitats

which have been found to host a large number of *G. pallidipes* and *G. longipennis* (Brightwell *et al.* 1997). The sampling site (Mukinyo) is at the border of the Olkiramatian group ranch, which is a wildlife conservancy where the distribution of domestic and wild tsetse host overlap because there are no fences.

Characteristics of the flies and presence of trypanosomes were taken from Channumsin et al. (2018). Whole flies were preserved in 95% ethanol and stored at -80°C. Sex and species of flies were determined based on morphological characters, and age was based on a wing fray score (Jackson 1946). Presence of trypanosomes was determined using general primers targeting the ITS-1 region of the rDNA array (CF: 5' CCGGAAGTTACCGATATTG 3' and BR: 5' TTGCTGCGTTCTTCAACGAA 3'; (Njiru *et al.* 2005)) that allow identification of trypanosome species based on size of amplicons. For blood meal analyses in this study, we used DNA extracted from abdomen samples from the *G. pallidipes* (N = 577) samples described in Channumsin et al. (2018).

Identification of type and number of hosts from *G. pallidipes* blood meals

Although two sets of primers were tested for efficiency and reliability of amplification of vertebrate hosts from tsetse fly flood meals we selected primers developed by Kocher *et al*., (1989) targeting a 359 bp fragment of the mitochondrial gene cytochrome B (cytb) gene in mammals (Cb1: 5' CCATCCAACATCTCAGCATGATGAAA 3' and Cb2: 5' GCCCCTCAGAATGATATTTGTCCTCA 3'), for comparison with Auty et al. (2016a) and because they showed more reliable amplification in a pilot study (Wongserepipatana 2016) than primers targeting the mitochondrial cytochrome C oxygenase 1 (CO1) gene (VF1d-t1 and VR1d-t1) (Ivanova *et al.* 2007). PCR cycling

was carried out in 25 µl reaction mixtures containing: 1X Invitrogen PCR buffer; 0.2 mM dNTP mixture; 1.5 mM $MgCl_2$ (Thermo Scientific); 0.5 µM of each primer; 1 unit of *Taq* DNA polymerase (Invitrogen Inc, Carlsbad, CA., U.S.A.); and 2 µl tsetse abdomen DNA template. Samples were pre-heated at 94°C for 5 min, denatured at 94°C for 30 sec, annealed at 55°C for 45 sec, then extended at 72°C for 30 sec, with 35 cycles of the amplification and a final extension at 72°C for 10 min (Muturi *et al.* 2011b). PCR products were visualised using 1.5% UltraPure™ Agarose gels (Invitrogen, Paisley) with 2% Ethidium Bromide (Invitrogen, Paisley) in 1X TBE buffer (108 g of Tris Base, 55 g of Boric acid and 40 ml of 0.5 M EDTA). Results were visualised and analysed by a gel documentation system (UVIpro Platinum, UVITEC, Cambridge, UK or GeneDoc, BioRad Inc, UK).

The feeding status of tsetse flies was classified as follows: flies showing cytb amplicons of the expected size (359 bp) were classified as “fed” and negative amplifications as “unfed”, with sequencing of the cytb fragment used to identify host species fed on. Since “unfed” would include false negatives for PCR amplification, analyses were only conducted based on the “fed” products. PCR products at ≥ 20 ng/15 µl were cleaned using ExoSAP-IT PCR Clean-up Kits (GE Healthcare). In cases where the yield of PCR products was lower than this threshold, multiple PCR products were concentrated and QIAquick Gel Extraction Kits (Qiagen Inc, Paisley, UK) were applied to extract the PCR products from agarose gels. All purified samples were sent for Sanger sequencing in both forward and reverse directions, using the Sequencing Service at the University of Dundee.

Chromatogram sequences were corrected and aligned using the Sequencher software program, version 5.3 (Gene Codes Corporation, Ann Arbor, MI USA). The Basic Local Alignment Tool for somewhat similar sequences (BLASTn) (Altschul *et*

al. 1990), was used to identify the closest matching sequences in the GenBank database to determine the host identity of each consensus sequence. Chromatographs with only single peaks based on direct sequences were classified as “single host feeding”. Sequences that showed more than one clear peak at multiple positions were classified as “multiple host feeding”. While the difficulty of resolving the phase of mutations precluded identification of all hosts from direct sequencing of multiple-peak products, the dominant host was determined based on BLASTn analysis of the most prominent peaks. Multiple-peak sequences that clearly showed the presence of more than one host species but for which a dominant sequence was not resolved, were considered as “multiple feeding” but were excluded from statistical analyses.

A subset of samples was rechecked to confirm host-feeding patterns by cloning using TOPO®-TA Cloning Kits (Invitrogen, UK), with at least six plasmids of each sample sent for sequencing, after purifying using QIAprep Spin Miniprep Kits (Qiagen Inc, Paisley, UK). Ten samples whose chromatographs showed double or triple peaks at single positions in the direct sequences were cloned to confirm that multiple peaks were due to feeding on multiple host species rather than poor quality sequences (five from Buffalo Ridge; three from Zungu Luka; two from Mukinyo). An additional seven samples that appeared to have fed on single hosts but with ambiguous peaks were also cloned (five from Buffalo Ridge and two from Zungu Luka).

Dominant hosts resolved were classified as “wild” or “domestic” (the latter including humans, livestock and companion animals). For all unambiguous direct and cloned sequences, separate alignments were made for each host species identified through BLASTn, using Sequencher version 5.3 (Gene Codes Corporation, Ann Arbor, MI

USA). Sequences were exported to Se-Al version 2.0 (Rambaut 1996) to manually align and prune sequences to the same length and DNAsp version 5.0 (Librado & Rozas 2009) was used to resolve variants into unique haplotypes within host species. For sequences identified as human, samples were compared to the Asian variant sequenced from the main investigator; any sequences matching those were considered as possible contaminants and so were excluded from further analyses.

Statistical Analyses

Generalised linear models (as implemented in R version 3.3.2 (Team 2016) were used to test whether feeding behaviour of the flies (single vs multiple hosts) was influenced by site, type of host (domestic vs wild), tsetse-specific traits (sex, age) or the presence of trypanosomes. The full model included interactions between type of host with site, sex and age of the flies, as well as interactions between site with tsetse specific factors and trypanosome presence. Feeding behaviour was modelled as a binary response variable, using the logit link function. Model selection was performed using likelihood ratio tests to find the minimum model that best explained the data. The same approach was used to test whether the presence of trypanosomes was explained by feeding behaviour, host type, site, or tsetse-specific factors, including all pairwise interactions and a three-way interaction between sex, age and site, based on conclusions from the previous study (Channumsin *et al.* 2018). Model fit was assessed using pseudo R^2 (as implemented in the R package, BaylorEdPsych (Beaujean 2012)). Given the wide range of hosts that the flies feed on, the influence of particular host species on trypanosome prevalence was considered only qualitatively.

Given the complexities of possible associations between variables (Channumsin *et al.* 2018), multivariate analyses were applied to visualise graphic correlations of trypanosome status with tsetse intrinsic factors and feeding patterns for the *G. pallidipes* samples for which feeding patterns and dominant host species could be assigned. Multiple Correspondence Analysis (MCA), as implemented in the FactoMineR package (version 1.30 (Le *et al.* 2008)) was used, considering the categorical variables: site, presence or absence of *Trypanosoma spp*, sex, age, feeding pattern (single vs multiple) and dominant host type (domestic vs wild). Variation along pairs of principal component axes was visualised using “ggplot2()” (Ginestet 2011) in R (Team 2016).

Host species diversity

To assess how intraspecific variation among the hosts identified in our study related to that found in previous cytb-based tsetse fly blood meal analysis surveys in eastern Africa we downloaded and aligned popset sequences from: 1) a study comparing wild host usage in the Nguruman region of Kenya (where Mukinyo is located) and the Serengeti National Park in Tanzania (Muturi *et al.* 2011b); and 2) a survey conducted in the Serengeti that had also quantified wild host usage in relation to host density (Auty *et al.* 2016a). We selected these studies because they had sequences available for all flies screened using the same cytb primer set that we used, rather than just unique host haplotypes found; although Muturi et al. (2011) had also surveyed a site In Uganda, only domestic cattle were identified as hosts at that site. We used DNAsp version 5.0 (Librado & Rozas 2009) to collapse the alignment into unique haplotypes and to compute the relative frequency of each in the various geographic locations represented in the three studies. Minimum

spanning networks were then plotted using PopArt (Leigh & Bryant 2015), to indicate relative frequencies across the sampling sites.

Results

Identification of type and number of hosts from *G. pallidipes* blood meals

From 573 *G. pallidipes*, 128 flies showed no evidence of a recent blood meal based on lack of amplification products following screening with the Cb1 and Cb2 primers; although these were classified as "unfed", the possibility that some represented failure of the PCR meant that they were excluded from analyses (Table 1). The remaining 445 flies showed amplified products of the expected size, which were sequenced and used to classify feeding status (Table 1; Supplementary Table 2). Among these "fed" flies, chromatograms of 198 samples were not of sufficient quality to reliably determine the source of the blood meals (these were classified as "unidentified"), but the proportion of samples for which hosts could be reliably identified varied by site (Table 1). For 197 of the 247 samples for which dominant hosts could be resolved to species, a single amplification product was apparent in the chromatographs; these were classified as having recently fed on a single host. Cloning of seven of these samples confirmed amplification of DNA from only a single host species (Supplementary Table 2). The chromatographs for the other 50 samples clearly showed multiple peaks that could be confidently attributed to feeding on multiple hosts rather than poor sequence quality. The dominant host could be identified to species for these 50 samples. Cloning of 10 of these cytb PCR products confirmed the amplification of DNA from more than one host, with up to four different host species identified in single flies (Supplementary Table 3).

There was no apparent detection bias in relation to age of the flies in being able to resolve blood meals, with the relative proportions of flies with unidentified hosts, or having fed on single or multiple hosts similar across age categories (Supplementary Figure 1) and roughly in proportion to the number flies sampled in a particular age category (Figure 1; Supplementary Table 4).

Most sequences obtained from human hosts among the samples screened were similar to cytb sequences identified from tsetse blood meals in the Serengeti, Tanzania (type 2) or Zambia (type 3) (Supplementary Table 6). However, three samples with evidence of only a single host matched an Asian haplotype from Taiwan (type 1: one from each of the three sites), which is the ethnic origin of the primary researcher; these were excluded from analyses because they were suspected contaminants.

There was extensive variation among sites in both the degree of multiple feeding and host types (Table 1; Figure 1). The two sites from the Shimba Hills (Buffalo Ridge and Zungu Luka) showed a higher proportion of flies that appear to have fed on multiple hosts than the site from Nguruman (Mukinyo): 41.8% from Buffalo Ridge; 61.5% from Zungu Luka, compared with 2.0% from Mukinyo. However, this appeared to be influenced by the type of host. At Mukinyo, only a single fly fed recently on a domestic host (identified as single feeding on cattle) and only 3/149 flies with dominant wild hosts identified had fed on more than one host species (confirmed by cloning for two of the individuals; Supplementary Table 3). In contrast, 77% of dominant hosts were identified as domestic for Zungu Luka and 23/30 of these had fed on multiple host species, compared with only 1/9 of the flies for which dominant sequences were identified as wild hosts. All three multiple feeding flies cloned from this site had fed on humans, with one also

having fed on both domestic (goat, mouse) and wild (antelope) hosts, one having also fed on a single wild host (bushbuck) and another on a domestic host (chicken) (Supplementary Table 3). The site in the middle of the fenced wildlife protected area close to this site (Buffalo Ridge) showed a much higher proportion of flies that had fed on wild hosts (65.5%) but most individuals with a dominant domestic host had fed on multiple species (18/19 for domestic hosts, compared with 5/36 for wild hosts). Cloning revealed that all five of the individuals classified as multiple feeding screened had fed on humans and at least one other domestic animal; four of the individuals had also fed on a wild host (Supplementary Table 3).

Using the type of feeding behaviour (single vs multiple hosts) as a binary response variable, the final model selected by maximum likelihood included a significant interaction between host type (domestic vs wild) and age of the flies (LRT: χ_2= 8.9, df = 8, p = 0.012), as well as a significant effect of site (LRT: χ_2= 13.7, df = 8, p = 0.001). There was no effect of sex of the flies (LRT: χ_2= 0.01, df = 9, p = 0.995), or presence of trypanosomes (LRT: χ_2= 3.56, df = 9, p = 0.172). Pseudo R^2 indicated that the model explained a substantial amount of the variation (McFadden's = 0.63; count = 0.93). Although the number of individuals was small, at all three sites, only juvenile flies fed multiple times when they had access to wild hosts, whereas young flies fed multiple times more often than the other age classes when exposed to domestic hosts (Figure 1). There were fewer old flies than the other age classes at all sites but they were at highest frequency at Mukinyo, where they fed singly on wild hosts; only a single juvenile had fed recently on a domestic host.

Host composition varied across sites (Table 2), with buffalo dominating in the two wildlife protected areas (Buffalo Ridge and Mukinyo) and humans predominating in

the site bordering the SHNR (Zungu Luka), where no buffalo were detected (Supplementary Figure 2). Mukinyo had a wider range of wild hosts than Buffalo Ridge but elephants, antelope and warthog were found at both sites. Buffalo Ridge also shared most of the same domestic host species as Zungu Luka, demonstrating that flies moved across the fenced interface to feed. Only a single fly (from Mukinyo) was confirmed to have fed on domestic cattle.

Prevalence of *Trypanosoma spp*. in relation to *G. pallidipes* feeding patterns

Across sites, 44% (n = 107) of the flies for which hosts could be identified to species tested positive for any trypanosome species, with 56% associated with dominant domestic hosts and 40% with wild (Table 1). Of the three most abundant host species, flies that had fed on humans and elephants had the highest prevalence of trypanosomes (51.8% and 51.4%, respectively), followed by buffalo (38.5%). Of flies feeding on multiple hosts, 58% tested positive for trypanosomes, compared to 40% that had fed on single hosts, but this was influenced by the higher rate of infection in Zungu Luka (61.5%), where single feeding was rare, compared to Buffalo Ridge (36%) and Mukinyo (42%). Coinfections between more than one species of trypanosomes occurred in 7.2% of flies (n = 18), including four with domestic hosts dominant but all of which had fed on multiple host species (two from Buffalo Ridge and two from Zungu Luka) (Supplementary Table 5). The remainder of flies with coinfections showed patterns of single feeding on buffalo (three from Buffalo Ridge and seven from Mukinyo) or elephants (four from Mukinyo

As found in our previous study (Channumsin *et al.* 2018), generalised linear models using trypanosome presence as a response variable were not very informative, possibly due to lack of orthogonality between explanatory variables related to tsetse specific factors and site of study. Initial model selection using LRTs retained interactions between fly sex and age, fly age and site, as well as an individual fixed effect of host feeding pattern (single vs multiple). However, final LRTs to test the significance of the variables included in the final model resulted in a significant effect only of host feeding pattern (LRT: χ_2= 5.08, df = 2, p = 0.02), with flies that had recently fed on more than one host being more likely to test positive for trypanosomes than those that had fed on only a single host. However, testing the fit of the two models using pseudo-R^2 indicated that only a small amount of the variation in trypanosome presence was explained by either model (more complex model: McFadden's = 0.08; count = 0.65; minimal model: McFadden's = 0.11; count = 0.65).

For this reason, multivariate ordination analyses were used to visualise associations between variables. Based on MCA analyses, strong correlations among site of *G. pallidipes* collection, host feeding pattern and type of host were apparent in dimension 1 (Table 3; Figure 2). Trypanosome status was resolved primarily along dimension 2, showing a positive association with juvenile male flies, while trypanosome negative flies tended to be found in young female flies.

Trypanosoma vivax was the predominant species identified in all populations sampled (59 individuals as a single infection and 16 in coinfections with other species of trypanosomes) and occurred in association with all hosts, except for the two hyaena samples and the single cattle sample (Supplementary Table 5). *Trypanosoma congolense* was also found at all sites (18 as a single infection and 13

in coinfections) and in many different host species. *Trypanosoma brucei* spp. was present in seven single and six coinfections and was only found in flies that recently fed on buffalo, elephants and humans.

Host species diversity

In addition to identifying just the species of host from the blood meals, intraspecific variation in mtDNA haplotypes within wild host species was found in our samples (Figure 3; Supplementary Table 6). Comparing intraspecific variation to other studies (Supplementary Table 7), diversity across the three Kenyan sample sites in our study was lower than that in the Serengeti samples obtained by Auty et al. (2016) and Muturi et al. (2011), except for elephants, for which all three haplotypes found amongst our samples were found in other populations (Figure 3a). Buffalo (Figure 3b) had 16 haplotypes (7 of which were also found in our study), but most were differentiated by single nucleotide substitutions. Giraffes had five haplotypes but only one of the two most frequent in the Serengeti was found at the Kenyan sites (Figure 3c). The BLAST analysis revealed that there is not a very comprehensive database of antelope species available in GenBank and so not all of the haplotypes we identified could be identified to species (Figure 3d): three haplotypes were at least 98% similar to bushbuck (*Tragelephus scriptus*), and one to Lesser Kudu (99% to *Tragelaphus imberbis*) but the other two divergent haplotypes were not similar to any named species in the database (Supplementary Table 6). Three haplotypes were at least 98% similar to bushbuck; one was 99% similar to lesser Kudu and two were highly divergent and showed <95% similarity to bushbuck. However, the latter two were found in single clones, in flies that had fed on at least one other host. Warthogs showed particularly high haplotype diversity in the Serengeti, but only 3 haplotypes were found at our sites (Figure 3e). Excluding sequences that were

identical to the sequence obtained from the main investigator (of Asian origin), three other human haplotypes were found, one of which was also found in the Serengeti (Figure 3f). A single haplotype was found for hyaenas (*Crocuta crocuta*); although this species was identified as the dominant host only for two flies from Mukinyo, the same haplotype was found in the Serengeti.

Single haplotypes were found for all of the domestic hosts except humans: mouse (*Mus musculus*), chickens (*Gallus gallus*), goat (*Capra hircus*) and cattle (*Bos taurus*). For cattle, Muturi et al. (2011) found the same haplotype as in our sample from Mukinyo, along with two that differed by one bp substitution each in samples from Uganda. Auty et al. (2016) did not find any domestic hosts in their Serengeti sampling.

Discussion

Our results suggest that blood meal analysis of tsetse flies provides a useful tool for assessing the potential for cross-species transmission of parasites between domestic and wild hosts in interface areas between human settlements and wildlife protected areas. Although host preferences were not specifically assessed in our study, rates of host switching appeared to be higher on flies that had fed on domestic hosts compared to wild hosts. While feeding on multiple hosts could increase transmission of parasites between host species, the high host fidelity of flies feeding on wild hosts could reduce risk of transmission from wildlife to livestock or humans. Site-specific differences and age distribution of the flies also may have influenced the feeding patterns in different types of host. As in our previous study (Channumsin *et al.* 2018), MCA suggested that prevalence of trypanosomes was associated more with geographic location than with intrinsic tsetse factors but

feeding patterns and type of host also explained some of the variance. Together, these results suggest that differences in host communities in different regions could influence the risk of transmission between vectors and hosts in complex ways. Haplotype networks comparing cytb variation within host species that tsetse had fed on from Kenya and Tanzania indicated the potential to use intraspecific variation in mammals to infer patterns of host movement and connectivity, if more variable genetic markers were used. The value of blood-feeding arthropods as "flying syringes" (Bitome-Essono *et al.* 2017) could thus have added value for noninvasive sampling of wild animals for conservation genetics purposes and to infer connectivity among hosts that could inform epidemiological models, rather than just identifying hosts and pathogens to species.

Feeding patterns of *G. pallidipes*: Type of hosts

We identified the dominant hosts for 46% of the *G. pallidipes* samples screened (56% of the samples that showed positive amplification products), which is comparable or higher than previous studies using the same primers (Auty *et al.* 2016a; Bitome-Essono *et al.* 2017). We found extensive variation among the species fed on in two different geographic areas. In the Shimba Hills, where a fenced wildlife protected area is located within a few km of human settlements, flies fed on both domestic and wild hosts, with blood meals from both host types detectable within individual flies. In contrast, in the Nguruman region, which is within 70 km of one of the highest density wildlife populations in the region (the Seregenti National Park in Tanzania), only a single fly was identified that had fed on a domestic host. This is consistent with Muturi et al. (2011a), who also did not identify domestic hosts in their

survey of the Nguruman region, despite finding predominantly cattle at a site surveyed in Uganda. Snow et al. (1988) suggested that, even though flies in areas dominated by cattle fed readily on these domestic hosts, a positive correlation between the number of wild herbivores and the abundance of *G. pallidipes* suggested that feeding success was poor on local livestock (based on a low density of flies where cattle were numerous). This could help to explain decreased density of tsetse traversing from wildlife-protected areas through to livestock dominated areas in interface areas (Lord *et al.* 2018).

Surprisingly, no domestic cattle were detected in our study from the Shimba Hills, despite the proximity to farms with mixed herds of cattle, sheep and goats (Njenga *et al.* 2011). However, domestic hosts were also not identified in the Shimba Hills region in a previous study based on host detection using haemagglutinin assays (Snow *et al.* 1988). This could indicate that flies avoid livestock when more favourable hosts are present. However, there also could be seasonal differences, as trypanosome prevalence in cattle was high (33.9%) in Kwale County, in a previous study that also found *G. pallidipes* at high abundance (Mbahin *et al.* 2013). At Mukinyo a single individual fed on domestic cattle but there was a very low proportion of flies that fed on multiple hosts (2%) and a predominance of buffalo (71%) among the samples where the dominant host could be identified. Although our blood meal analysis cannot predict host preferences, it is possible that buffalo are abundant hosts that are easy to feed on and so flies could learn to return to the same host species (Bouyer *et al.* 2007).

Our finding of African buffalo as the main hosts of *G. pallidipes* in Nguruman and Buffalo Ridge supports previous reports that ruminants are attractive to adult *G. pallidipes*, *G. fuscipes* and *G. brevipalpis* (Allsopp *et al.* 1972a; Harraca *et al.* 2009). However, host selection has been found to vary extensively by population

(Supplementary Table 8). Differences across studies could be due to differences in methodology but also could be due to microhabitat differences (Spath 2000), such as seasonal variation in host availability, the vegetation type or cover at particular sites, or particular environmental conditions in different years, which affects overlap of habitat and activities between tsetse flies and hosts (Allsopp *et al.* 1972a; Clausen *et al.* 1998). It is interesting that no buffalo were detected at Zungu Luka, despite its close proximity to Buffalo Ridge, where buffalo were the most frequent host. This could suggest that flies feeding in human settlements move into the park to feed on wildlife but once feeding on their preferred wildlife, they do not move out into the human-settled regions or that flies do not move very far. Our results cannot confirm host preferences because we do not have robust data on host availability (Auty *et al.* 2016a) but the directionality of movements of flies in interface areas would be interesting to assess. Preferences for different wild host species in tsetse flies do not necessarily reflect host densities in large wildlife protected areas (Allsopp *et al.* 1972a; Auty *et al.* 2016a), and variation in host preference of domestic hosts has been found in the absence of wildlife (England & Baldry 1972b). However, specific choice tests between domestic and wild hosts could reveal important information about preferences that could inform control interventions, as has been done for malaria-carrying mosquitos (Lyimo & Ferguson 2009). Nevertheless, the finding of flies collected in the same traps feeding on both wild and domestic hosts emphasizes the high potential for cross- feeding between these host types when they occur sympatrically.

It is more difficult to interpret patterns of feeding on humans because of the potential for biting during setting traps or sampling or contamination during DNA extractions or PCR amplifications (England & Baldry 1972b). Humans have been suggested as inappropriate hosts because they camouflage their odours, apply chemical

repellents, and react strongly to tsetse bites, which could result in unsuccessful feeding (Baylis 1996; Baylis & Nambiro 1993; Hargrove 1976) that would lead to host switching. Hargrove (1976) found that the presence of humans not only repelled tsetse flies but also inhibited the landing response to approach other potential hosts nearby. The relatively high proportion of blood meals containing human DNA was thus not expected. Most of the mtDNA haplotypes we identified from human samples were consistent with those expected regionally and one haplotype was shared with those found from the Serengeti (Figure 3); we also did not find evidence of feeding on humans in Mukinyo. However, we cannot completely rule out biases leading to an apparently high rate of human feeding.

Feeding patterns of *G. pallidipes*: Number of hosts

We found that the propensity for feeding on single compared to more than one host species was determined by an interaction between age of the flies and site, as well as the type of host fed on, with more multiple feeding involving domestic hosts, which could further reflect a preference for feeding on wildlife. Cloning and sequencing revealed that flies feeding on domestic hosts had fed on up to four different host species and confirmed that single feeding was more common in flies feeding on wildlife. Theoretically, the number of clones could be used to predict which host was last fed on, but this would also depend on the rate of feeding of the fly (e.g. if they were interrupted and switched hosts very rapidly, more than one blood meal might have a similar DNA concentration) and lack of bias in PCR amplifications. There also could be behavioural differences that could result in detection biases: 1) flies might feed more thoroughly on their preferred hosts,

increasing the blood meal volume from that host; 2) flies might feed multiple times on the same host species occurring at high local densities; or 3) hosts might differ in effective defense mechanisms, resulting in low blood meal volumes due to interrupted feeding (Torr *et al.* 2001). If feeding on an initial host is interrupted or too low quality ("unsuccessful"), flies might switch hosts. Unsuccessful feeding of tsetse flies on cattle have been attributed to host defense (69%) and competition with other biting insects (31%) (Schofield & Torr 2002). Wild animals might react less to tsetse flies feeding and/or be surrounded by less other biting insects than domesticated animals. Nevertheless, our results suggest higher host fidelity when feeding on wild, compared to domestic, hosts.

Prevalence of *Trypanosoma spp*.

There was not a clear association between prevalence of trypanosomes, type of host or host-feeding patterns in the tsetse flies. However, the lack of an association is consistent with the MCA, which suggested a stronger correlation among feeding pattern, host type and site than trypanosome status, sex and age of tsetse flies. We hypothesised that feeding on multiple hosts could increase risk of trypanosome infection in flies but that was not apparent in the MCA (Figure 2), possibly due to the much stronger effects of site on feeding patterns and type of hosts. Bouyer *et al*., (2007) suggested that repeated feeding on the same host species was likely to increase risk of trypanosome transmission within species, but to decrease risk between species. Leak (1998) suggested that trypanosome infection might influence feeding success and feeding behaviour of the flies. For example, high numbers of *T. congolense*, which attach to the cuticle of the proboscis, could interrupt feeding and result in more frequent probing. Alternatively, nutritional status of the flies could affect their relative susceptibility to trypanosome establishment (Kubi *et al.* 2006).

Thus, an association between the frequency of feeding and trypanosome infection status should be further studied in laboratory experiments to test whether trypanosome infection causes a feeding pattern change or differences in feeding patterns promote trypanosome infection.

There are also could be differences in the susceptibility of different host species to trypanosomes, regardless of the feeding rate by tsetse. Although African buffalo were the most frequent dominant hosts detected, a larger proportion of flies that had fed on African elephants were trypanosome positive. This could support previous suggestions that African buffalo might be resistant to trypanosome infection (Grootenhuis *et al.* 1990; Olubayo *et al.* 1996). Other studies have reported an association between trypanosome infections and particular species of wild animals (Anderson *et al.* 2011a; Snow *et al.* 1988). Nevertheless, our results do not suggest an increased prevalence of trypanosomes in communities where both domestic and wild hosts are fed on that would suggest increased risks in livestock interface regions.

Diversity of host species

Comparison of mtDNA haplotype diversity in the most frequent hosts with published studies suggested higher diversity in African buffalo and warthogs than in elephants and giraffes. However, this comparison highlights some limitations of blood meal

analyses based on using only a single genetic marker. We found that a higher proportion of hosts could be resolved from Mukinyo than the other areas, possibly due to the dominance of wild hosts. However, there is also potential that biases in PCR amplification could have influenced this pattern. In the Shimba Hills, although goats were identified from flies sampled from both sites, there were also additional samples that matched goats in BLAST analyses that were not included in the analysis because it was difficult to determine whether the sequences represented multiple feeding or just poor sequence quality. Both in the sequences from Auty et al. (2016) and our sequences, warthog tended to include more ambiguous base calls than the other hosts, which might have inflated diversity estimates. Previous studies comparing the relative reliability of cytb and COI mtDNA (e.g. (Muturi *et al.* 2011a) have found that neither alone amplifies products from all potential host types present. Differences in amplification biases and sequence quality between host type sequences could thus bias the interpretation of single vs multiple feeding and the propensity for feeding on domestic vs wild hosts.

Moreover, blood meal analyses rely on completeness of reference databases. We found several cytb haplotypes that were closest to antelope in BLAST but the similarity was too low to resolve to species (93%); this lack of reference sequences could have lead to underestimates of host usage in previous blood meal analyses. We took a conservative approach to host identification by basing analyses only on the dominant host species identified in each fly, but this could represent either the most recent blood meal or the hosts on which most efficient feeding occurred. There has been a recent shift towards using next generation sequencing approaches for amplicon based host identification (Hoffmann *et al.* 2016; Reeves *et al.* 2018) that would allow consideration of relative read numbers for hosts in mixed samples; while

this could provide more powerful information on feeding behaviour of haemotophagous insects the same issues with availability of reference databases and amplification biases still need to be overcome. Non-PCR based assays such as high resolution melting point analysis (Farikou *et al.* 2010; Omondi *et al.* 2015) or NGS alternatives such as hybrid sequence capture (Barrow *et al.* 2019; Metsky *et al.* 2019) might thus provide a more comprehensive assessment of feeding patterns and range of hosts fed on.

Conclusions

Identification of the hosts that *G. pallidipes* fed on based on direct PCR sequencing revealed evidence for both use of a wide range of hosts and multiple feeding bouts by individual flies. However, in wildlife dominated areas, there was a much stronger tendency for flies to feed on single host species compared to a site on the edge of a conservation area, where domestic hosts were more common and flies fed on up to four different detectable host species. If this indicates that domestic animals are not preferred hosts, this could have important implications for understanding risk of transmission of trypanosomes between wildlife and livestock. Our results also highlight the potential to use blood meals not only to characterize pathogen diversity in relation to host usage but also to quantify genetic variation among hosts and patterns of population genetic structure that could inform epidemiological models of transmission risk between different geographic areas or types of hosts.

Data Availability

Sequences have been deposited to Genbank, with accession numbers MN148732-MN148768 (Supplementary Table 9).

Acknowledgments

This work was funded by the Wellcome Trust (grant number: 093692) to the University of Glasgow and a Phd scholarship from the Thailand government to MCh. We thank Elizabeth Kilbride for assistance with laboratory work. The authors are grateful to the Kenya Wildlife Service for permission to perform this study in protected areas. Esther Waweru and James Kabii provided technical and administrative assistance at *icipe*. James Kabii, Moses Tanju, Loya Suleyman, John Andoke and his team (*icipe*) provided invaluable expertise and help for the trapping of tsetse.

References

Allsopp R, Baldry DAT, Rodrigues C (1972a) Influence of Game Animals on Distribution and Feeding Habits of Glossina-Pallidipes in Lambwe Valley. *Bulletin of the World Health Organization* **47**, 795-809.

Allsopp R, Baldry DAT, Rodrigues C (1972b) The influence of game animals on the distribution and feeding habits of Glossina pallidipes in the Lambwe Valley. *Bulletin of the World Health Organization* **47**, 795-809.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

Anderson NE, Mubanga J, Fevre EM, *et al.* (2011a) Characterisation of the wildlife reservoir community for human and animal trypanosomiasis in the Luangwa Valley, Zambia. *PLoS neglected tropical diseases* **5**, e1211.

Anderson NE, Mubanga J, Fevre EM, *et al.* (2011b) Characterisation of the Wildlife Reservoir Community for Human and Animal Trypanosomiasis in the Luangwa Valley, Zambia. *PLoS Neglected Tropical Diseases* **5**.

Auty H, Anderson NE, Picozzi K, *et al.* (2012) Trypanosome diversity in wildlife species from the serengeti and Luangwa Valley ecosystems. *PLoS Negl Trop Dis* **6**, e1828.

Auty H, Cleaveland S, Malele I, *et al.* (2016a) Quantifying heterogeneity in host-vector contact: tsetse (*Glossina swynnertoni* and *G. pallidipes*) host choice in Serengeti National Park, Tanzania. *PloS one* **11**, e0161291.

Auty H, Morrison LJ, Torr SJ, Lord J (2016b) Transmission dynamics of Rhodesian sleeping sickness at the interface of wildlife and livestock areas. *Trends in Parasitology* **32**, 608-621.

Barrow LN, Allen JM, Huang X, Bensch S, Witt CC (2019) Genomic sequence capture of haemosporidian parasites: Methods and prospects for enhanced study of host-parasite evolution. *Molecular Ecology Resources* **19**, 400-410.

Baylis M (1996) Effect of defensive behaviour by cattle on the feeding success and nutritional state of the tsetse fly, *Glossina pallidipes* (Diptera: *Glossinidae*). *Bulletin of Entomological Research* **86**, 329-336.

Baylis M, Nambiro CO (1993) The effect of cattle infection by *Trypanosoma congolense* on the attraction, and feeding success, of the tsetse fly *Glossina pallidipes*. *Parasitology* **106 (Pt 4)**, 357-361.

764 Beaujean A (2012) BaylorEdPsych: R package for Baylor University educational
765 psychology quantitative courses. R package version 5.0.
766 Bett B, Irungu P, Nyamwaro SO, *et al.* (2008) Estimation of tsetse challenge and its
767 relationship with trypanosomosis incidence in cattle kept under pastoral
768 production systems in Kenya. *Veterinary Parasitology* **155**, 287-298.
769 Bitome-Essono PY, Ollomo B, Arnathau C, *et al.* (2017) Tracking zoonotic
770 pathogens using bloodsucking flies as 'flying syringes'. *Elife* **6**.
771 Boakye DA, Tang J, P. T, Merriweather A, Unnasch TR (1999) Identification of
772 bloodmeals in haematophagous Diptera by cytochrome B heterodupliex
773 analysis. *Medical and Veterinary Entomology* **13**, 282-287.
774 Bouyer J, Pruvot M, Bengaly Z, Guerin PM, Lancelot R (2007) Learning influences
775 host choice in tsetse. *Biology Letters* **3**, 113-116.
776 Brightwell R, Dransfield RD, Stevenson P, Williams B (1997) Changes over twelve
777 years in populations of *Glossina pallidipes* and *Glossina longipennis*
778 (Diptera: *Glossinidae*) subject to varying trapping pressure at Nguruman,
779 south-west Kenya. *Bulletin of Entomological Research* **87**, 349-370.
780 Channumsin M, Ciosi M, Masiga D, Turner CMR, Mable BK (2018) Sodalis glossinidius
781 presence in wild tsetse is only associated with presence of trypanosomes in
782 complex interactions with other tsetse-specific factors. *BMC Microbiology*
783 **18**, 163.
784 Clausen PH, Adeyemi I, Bauer B, *et al.* (1998) Host preferences of tsetse (Diptera:
785 *Glossinidae*) based on bloodmeal identifications. *Medical and Veterinary*
786 *Entomology* **12**, 169-180.
787 Eloy L, Lucheis S (2009) Canine trypanosomiasis: etiology of infection and
788 implications for public health. *Journal of Venomous Animals and Toxins*
789 *including Tropical Diseases* **15**, 589-611.
790 England EC, Baldry DAT (1972a) The hosts and trypanosome infection rates of
791 *Glossina pallidipes* in the Lambwe and Roo valleys. *Bulletin of the World*
792 *Health Organization* **47**, 785-788.
793 England EC, Baldry DAT (1972b) Observations on relative attractiveness to *Glossina*
794 *pallidipes* of different animal baits, a tsetse Trap, and a fly-round patrol.
795 *Bulletin of the World Health Organization* **47**, 789-793.
796 Farikou O, Njiokou F, Simo G, *et al.* (2010) Tsetse fly blood meal modification and
797 trypanosome identification in two sleeping sickness foci in the forest of
798 southern Cameroon. *Acta Tropica* **116**, 81-88.
799 Gariepy TD, Lindsay R, Ogden N, Gregory TR (2012) Identifying the last supper:
800 utility of the DNA barcode library for bloodmeal identification in ticks.
801 *Molecular Ecology Resources* **12**, 646-652.
802 Geigy R, Kauffman.M (1973) Sleeping sickness survey in the Serengeti area
803 (Tanzania) 1971: I. Examination of large mammals for trypanosomes. . *Acta*
804 *Tropica* **30**, 12-23.
805 Geigy R, Kauffman.M, Jenni L (1973a) Wild Mammals as Reservoirs for Rhodesian
806 Sleeping Sickness in Serengeti, 1970-71. *Transactions of the Royal Society of*
807 *Tropical Medicine and Hygiene* **67**, 284-286.
808 Geigy R, Kauffman.M, Mayende JSP, Mwambu PM, Onyango RJ (1973b) Isolation of
809 Trypanosoma (Trypanozoon) rhodesiense from game and domestic animals
810 in Musoma District, Tanzania. *Acta Tropica* **30**, 50-56.
811 Geigy R, Mwambu PM, Kauffman.M (1971) Sleeping sickness survey in Musoma
812 District, Tanzania. IV. Examination of wild mammals as a potential reservoir
813 for *T. rhodesiense*. *Acta Tropica* **28**, 211-220.
814 Ginestet C (2011) ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal*
815 *Statistical Society Series a-Statistics in Society* **174**, 245-245.

816 Grootenhuis JG, Dwinger RH, Dolan RB, Moloo SK, Murray M (1990) Susceptibility of
817 African buffalo and Boran cattle to Trypanosoma congolense transmitted by
818 Glossina morsitans centralis. *Veterinary Parasitology* **35**, 219-231.
819 Haouas N, Pesson B, Boudabous R, *et al.* (2007) Development of a molecular tool
820 for the identification of leishmania reservoir hosts by blood meal analysis in
821 the insect vectors. *American Journal of Tropical Medicine and Hygiene* **77**,
822 1054-1059.
823 Hargrove JW (1976) The effect of human presence on the behaviour of tsetse
824 (*Glossina spp.*) (Diptera, *Glossinidae*) near a stationary ox. *Bulletin of*
825 *Entomological Research* **66**, 173-178.
826 Harraca V, Syed Z, Guerin PM (2009) Olfactory and behavioural responses of tsetse
827 fies, *Glossina spp.*, to rumen metabolites. *Journal of Comparative*
828 *Physiology A* **195**, 815-824.
829 Haydon DT, Cleaveland S, Taylor LH, Laurenson MK (2002) Identifying Reservoirs of
830 Infection: A Conceptual and Practical Challenge. *Emerging Infectious*
831 *Diseases* **8**, 1468-1473.
832 Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications
833 through DNA barcodes. *Proceedings of the Royal Society of London Series B-*
834 *Biological Sciences* **270**, 313-321.
835 Heisch RB, Mcmahon JP, Mansonbahr PEC (1958) The Isolation of Trypanosoma
836 Rhodesiense from a Bushbuck. *British Medical Journal* **2**, 1203-1204.
837 Hoffmann C, Stockhausen M, Merkel K, Calvignac-Spencer S, Leendertz FH (2016)
838 Assessing the feasibility of fly based surveillance of wildlife infectious
839 diseases. *Scientific Reports* **6**, 37952.
840 Ivanova NV, zemlak TS, Hanner rH, Hebert PDN (2007) Universal primer cocktails
841 for fish DNA barcoding. *Molecular Ecology Notes* **7**, 544–548.
842 Jackson CHN (1946) An artificially isolated generation of tsetse flies (*Diptera*).
843 *Bulletin of Entomological Research* **37**, 291-299.
844 Kent RJ (2009) Molecular methods for arthropod bloodmeal identification and
845 applications to ecological and vector-borne disease studies. *Molecular*
846 *Ecology Resources* **9**, 4-18.
847 Kocher TD, Thomas WK, Meyer A, *et al.* (1989) Dynamics of mitochondrial DNA
848 evolution in animals: amplification and sequencing with conserved primers.
849 *Proceedings of the National Academy of Sciences of the United States of*
850 *America* **86**, 6196-6200.
851 Kubi C, Van den Abbeele J, De Deken R, *et al.* (2006) The effect of starvation on
852 the susceptibility of teneral and non-teneral tsetse flies to trypanosome
853 infection. *Medical and Veterinary Entomology* **20**, 388-392.
854 Lah EFC, Ahamad M, Haron MS, Ming HT (2012) Establishment of a molecular tool
855 for blood meal identification in Malaysia. *Asian Pacific Journal of Tropical*
856 *Biomedicine* **2**, 223-227.
857 Langridge WP, Kernaghan RJ, Glover PE (1963) A review of recent knowledge of
858 the ecology of the main vectors of trypanosomiasis. *Bulletin of the World*
859 *Health Organization* **28**, 671-701.
860 Le S, Josse J, Husson F (2008) FactoMineR: An R package for multivariate analysis.
861 *Journal of Statistical Software* **25**, 1-18.
862 Leak SGA (1998) Tsetse biology and ecology. In: *The role in the epidemiology and*
863 *control of trypanosmiasis*. CABI Publishing, London.
864 Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network
865 construction. *Methods in Ecology and Evolution* **6**, 1110-1116.
866 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
867 polymorphism data. *Bioinformatics* **25**, 1451-1452.

868 Lisulo M, Sugimoto C, Kajino K, *et al.* (2014) Determination of the prevalence of
869 African trypanosome species in indigenous dogs of Mambwe district, eastern
870 Zambia, by loop-mediated isothermal amplification. *Parasites & vectors* **7**,
871 19.
872 Lord JS, Torr SJ, Auty HK, *et al.* (2018) Geostatistical models using remotely-
873 sensed data predict savanna tsetse decline across the interface between
874 protected and unprotected areas in Serengeti, Tanzania. *Journal of Applied*
875 *Ecology* **55**, 1997-2007.
876 Lyimo IN, Ferguson HM (2009) Ecological and evolutionary determinants of host
877 species choice in mosquito vectors. *Trends in Parasitology* **25**, 189-196.
878 Maudlin I, Holmes PH, Miles MA (2004) *The trypanosomiases* CABI Publishing,
879 Wallingford, UK.
880 Mbahin N, Affognon H, Andoke J, *et al.* (2013) Parasitological prevalence of bovine
881 trypanosomiasis in Kubo Division of Kwale Country of Coastal: baseline
882 survey *American Journal of Animal and Veterinary Sciences* **8**, 28-36.
883 Metsky HC, Siddle KJ, Gladden-Young A, *et al.* (2019) Capturing sequence diversity
884 in metagenomes with comprehensive and scalable probe design. *Nature*
885 *Biotechnology* **37**, 160-+.
886 Moloo SK, Sabwa CL, Kabata JM (1992) Vector competence of *Glossina pallidipes*
887 and *G. morsitans centralis* for *Trypanosoma vivax*, *T. congolense* and *T. b.*
888 *brucei*. *Acta Tropica* **51**, 271-280.
889 Muturi CN, Ouma JO, Malele, II, *et al.* (2011a) Tracking the feeding patterns of
890 tsetse flies (Glossina genus) by analysis of bloodmeals using mitochondrial
891 cytochromes genes. *PloS one* **6**, e17284.
892 Muturi CN, Ouma JO, Malele II, *et al.* (2011b) Tracking the feeding patterns of
893 tsetse flies (*Glossina* genus) by analysis of bloodmeals using mitochondrial
894 cytochromes genes. *PloS one* **6**, e17284.
895 Nash TAM (1936) The relationship between the maximum temperature and the
896 seasonal longevity of *Glossina submorsitans*, Newst., and *G. tachinoides*,
897 Westw., in Northern Nigeria. *Bulletin of Entomological Research* **27**, 273-
898 279.
899 Njenga SM, Mwandawiro CS, Muniu E, *et al.* (2011) Adult population as potential
900 reservoir of NTD infections in rural villages of Kwale district, Coastal Kenya:
901 implications for preventive chemotherapy interventions policy. *Parasites &*
902 *vectors* **4**, 1-6.
903 Njiru ZK, Constantine CC, Guya S, *et al.* (2005) The use of ITS1 rDNA PCR in
904 detecting pathogenic African trypanosomes. *Parasitology Research* **95**, 186-
905 192.
906 Nyawira NC (2009) *Analysis of tsetse fly (Diptera: Glossinidae) blood meals using*
907 *mitochondrial cytochrome genes for vertebrate host identification*, Ugerton
908 University.
909 Nyingilili HS, Malele I, Nkwengulila G (2016) Diversity of Blood Meal Hosts in
910 Glossina Pallidipes and Its Role in the Epidemiology of Trypanosomiasis at a
911 Localized Area in Serengeti National Park. *Imperial Journal of*
912 *Interdisciplinary Research* **2**, 1694-1698.
913 OIE (2013) Trypanosomosis (tsetse-transmitted).
914 Okoth SO, Kokwaro ED, Kirugu JM, Murila GA (2007) *Glossina pallidipes* and host
915 interactions: implications of host preference on tranmission Rick of
916 Rhodesian Sleeping Sickness in Kenya. *Trends in Applied Sciences Research*
917 **2**, 386-394.
918 Olubayo RO, Moloo SK, Naessens J (1996) Comparative parasite development in
919 African buffalo and N'Dama cattle infected with either *Trypanosoma*

920 *congolense* or *T. vivax*. *Bulletin of Animal Health and Production in Africa*
921 **44**, 23-32.
922 Omondi D, Masiga DK, Ajamma YU, *et al.* (2015) Unraveling host-vector-arbovirus
923 Interactions by two-gene high resolution melting mosquito bloodmeal
924 analysis in a Kenyan wildlife-livestock Interface. *PloS one* **10**, e0134375-
925 e0134375.
926 Peacock L, Ferris V, Bailey M, Gibson W (2012) The influence of sex and fly species
927 on the development of trypanosomes in tsetse Flies. *PLoS neglected tropical*
928 *diseases* **6**, e1515.
929 Pradhan V, Kamble Y, Ladniya V, Mogul M (2015) A overview of species
930 identification by DNA barcoding. *International Journal of Current*
931 *Microbiology and Applied Sciences* **4**, 127-140.
932 Rambaut A (1996) Se-Al: Sequence Alignment Editor. Department of Zoology,
933 University of Oxford, Oxford, UK.
934 Reeves L, Gillett-Kaufman J, Kawahara A, Kaufman P (2018) Barcoding blood
935 meals: New vertebrate-specific primer sets for assigning taxonomic
936 identities to host DNA from mosquito blood meals. *PLoS Neglected Tropical*
937 *Disease* **12**, e0006767.
938 Schofield S, Torr SJ (2002) A comparison of the feeding behaviour of tsetse and
939 stable flies. *Medical and Veterinary Entomology* **16**, 177-185.
940 Snow WF, Tarimo SA, Staak C, Butler L (1988) The feeding-habits of the Tsetse,
941 *Glossina pallidipes* Austen on the South Kenya Coast, in the context of its
942 host range and trypanosome infection-rates in other parts of East-Africa.
943 *Acta Tropica* **45**, 339-349.
944 Spath J (2000) Feeding patterns of three sympatric tsetse species (Glossina spp,)
945 (Diptera : Glossinidae) in the preforest zone of Cote d'Ivoire. *Acta Tropica*
946 **75**, 109-118.
947 Team RC (2016) R: A language and environment for statistical computing. . R
948 Foundation for Statistical Computing, Vienna, Austria.
949 Torr SJ, Wilson PJ, Schofield S, *et al.* (2001) Application of DNA markers to
950 identify the individual-specific hosts of tsetse feeding on cattle. *Medical Vet*
951 *Entomology* **15**, 78 - 86.
952 Turner DA (1987) The population ecology of *Glossina pallidipes* Austen (Diptera:
953 *Glossinidae*) in the Lambwe Valley, Kenya. I. Feeding behaviour and activity
954 patterns. *Bulletin of Entomological Research* **77**, 317-333.
955 Viana M, Mancy R, Biek R, *et al.* (2014) Assembling evidence for identifying
956 reservoirs of infection. *Trends in Ecology & Evolution* **29**, 270-279.
957 Wamwiri FN, Alam U, Thande PC, *et al.* (2013) *Wolbachia*, *Sodalis* and
958 trypanosome co-infections in natural populations of *Glossina austeni* and
959 *Glossina pallidipes*. *Parasites & vectors* **6**, 232.
960 Weitz B (1963) The feeding habits of *Glossina*. *Bulletin of World Health*
961 *Organization* **28**, 711-729.
962 Welburn SC, Picozzi K, Fevre EM, *et al.* (2001) Identification of human-infective
963 trypanosomes in animal reservoir of sleeping sickness in Uganda by means of
964 serum-resistance-associated (SRA) gene. *Lancet* **358**, 2017-2019.
965 Welburn SC, Picozzi K, Kaare M, *et al.* (2005) Control options for human sleeping
966 sickness in relation to the animal reservoir of disease. *Conservation and*
967 *Development Interventions at the Wildlife/Livestock Interface*, 55-61.
968 Wongserepipatana M (2016) *Prevalence and associations of Trypanosoma spp. and*
969 *Sodalis glossinidius with intrinsic factors of tsetse flies*, University of
970 Glasgow.

Table 1. Summary of blood meal analysis results based on direct sequencing. Cytb negative samples were classified as “unfed flies” but were not considered in the analyses since they could represent lack of amplification rather than lack of feeding. Single host feeding refers to cases where the cytb sequence had only single chromatograph peaks. Multiple host feeding were samples for which cytb was amplified but the sequences showed multiple peaks but the dominant sequence could be identified to species. Flies showing strongly amplified cytb PCR products but for which the number or type of host species could not be confirmed due to poor sequencing quality are labelled as “not identified”. The number of flies that tested positive for the presence of trypanosomes is indicated in parentheses. The final column shows the percent of samples for which dominant hosts could be identified for each population, with the % of those testing positive for trypanosomes in parentheses. The human samples that were potential contaminants were excluded but none tested positive for trypanosomes.

Site	**Single Feeding**		**Multiple Feeding**		**Not identified**	**Total Fed**	**“Unfed”**	**Total Screened**	**% Identified**
	Domestic	**Wild**	**Domestic**	**Wild**					
Zungu Luka	7 (3)	8 (5)	23 (15)	1 (1)	60 (31)	99 (55)	29 (22)	128 (77)	31 (62)
Buffalo Ridge	1 (1)	31 (8)	18 (8)	5 (3)	62 (23)	118 (43)	33 (10)	150 (53)	37 (36)
Mukinyo	1(1)	146 (60)	0	3 (2)	76 (22)	226 (85)	66 (20)	292 (105)	51 (42)

Table 2. Dominant host species resolved from blood meal analysis *G. pallidipes* sampled from the Shimba Hills (Buffalo Ridge and Zungu Luka) and the Nguruman region of Kenya, based on direct sequencing of cytb. Homozygous amplicons were classified as "single" feeding whereas sequences with multiple peaks were classified as having fed on "multiple" hosts. The dominant host was identified based on BLASTn. The relative abundance of the various host species is expressed as the total % of sequences for which the dominant host could be identified within that site. Species are ordered by relative abundance of wild and domestic hosts.

Site	**Buffalo Ridge**			**Zungu Luka**			**Mukinyo**		
Host	**Single**	**Multiple**	**Total %**	**Single**	**Multiple**	**Total %**	**Single**	**Multiple**	**Total %**
Buffalo	25	3	50.9	0	0	0.0	105	2	71.3
Elephant	2	1	5.5	1	1	5.1	29	1	20.0
Antelope	2	1	5.5	6	0	15.4	1	0	0.7
Warthog	2	0	3.6	1	0	2.6	5	0	3.3
Giraffe	0	0	0.0	0	0	0.0	4	0	2.7
Hyaena	0	0	0.0	0	0	0.0	2	0	1.3
Human[a]	1	9	18.2	6	11	43.6	0	0	0.0
Goat	0	7	12.7	1	7	20.5	0	0	0.0
Mouse	0	2	3.6	0	4	10.3	0	0	0.0
Chicken	0	0	0.0	0	1	2.6	0	0	0.0
Cattle	0	0	0.0	0	0	0.0	1	0	0.7
Total	32	23		14	25		147	3	

[a] Excluding potential contaminants

Table 3. Adjusted eta squared of the combination of variables in dimensions 1-3 in the MCA analysis, showing the variation explained for trypanosome status (positive or negative), Site (Mukinyo, Buffalo Ridge or Zungu Luka), Feeding pattern (single or multiple), Sex (male or female), Age (young, juvenile or old) and Host type (wildlife or domestic).

Variables	Dimension 1	Dimension 2	Dimension 3
Trypanosome status	0.059	0.203	0.206
Site	0.712	0.031	0.564
Feeding pattern	0.776	0.004	0.004
Sex	0.002	0.466	0.161
Age	0.024	0.565	0.200
Host Type	0.853	0.010	0.000

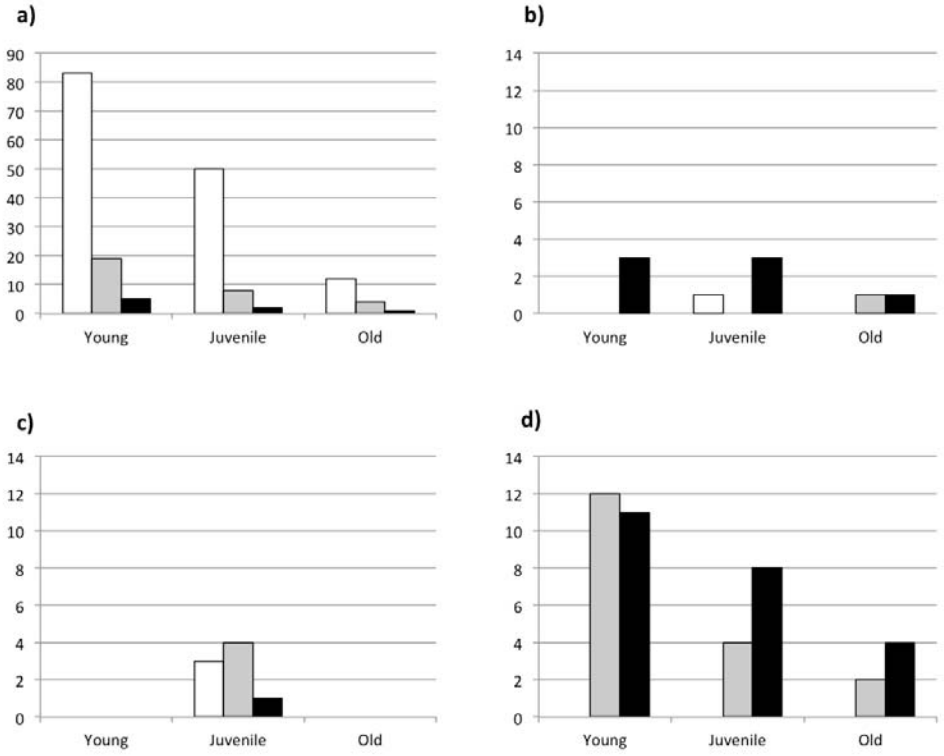

Figure 1. Distribution of flies across sites in relation to age, feeding behaviour and type of hosts fed on, showing flies feeding on: a) single wild hosts; b) single domestic hosts; c) multiple hosts with a wild host dominant; d) multiple hosts with a domestic host dominant. Flies were divided into young, juvenile and old age categories based on wing fray scores. Flies from all age classes at Mukinyo (white bars) fed predominantly on single wild host species (note the resulting larger scale for graph a), whereas feeding on a mixture of domestic and wild hosts was more common at the Shimba Hills sites, Buffalo Ridge (grey) and Zungu Luka (black). Note that only juvenile flies fed multiple times when a wild host was dominant.

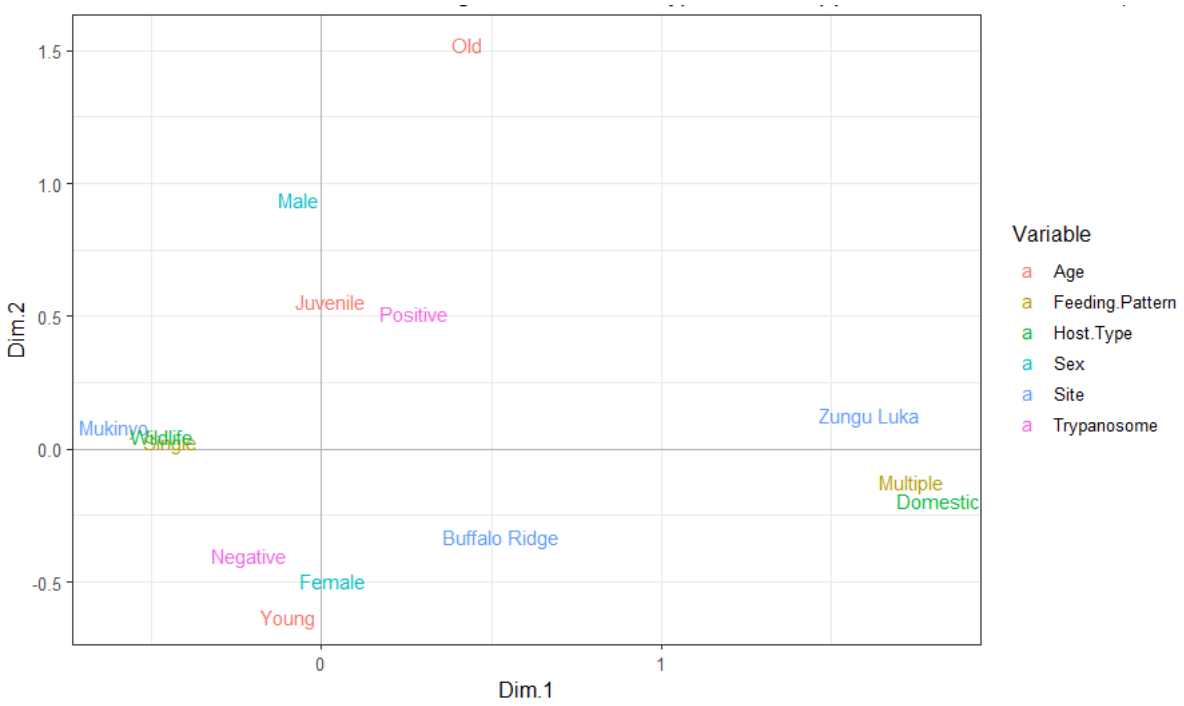

Figure 2. Multiple Correspondence Analysis (MCA), showing associations of dimension 1 (Dim 1; 30.3 % of the variance) and 2 (Dim 2; 16.0% of the variance) in relation to age category (young, juvenile, old), feeding pattern (single or multiple), host type (domestic or wildlife), sex (male or female), site (Buffalo Ridge, Zungu Luka or Mukinyo), and *Trypanosoma spp*. status (positive or negative). This figure clearly shows the strong association between feeding pattern and host type, driven by the differences in fly behaviour at Mukinyo compared to Zungu Luka resolved along dimension 1. Trypanosome status was not explained by variation along dimension 1 but was more related to age and gender of the flies resolved along dimension 2.

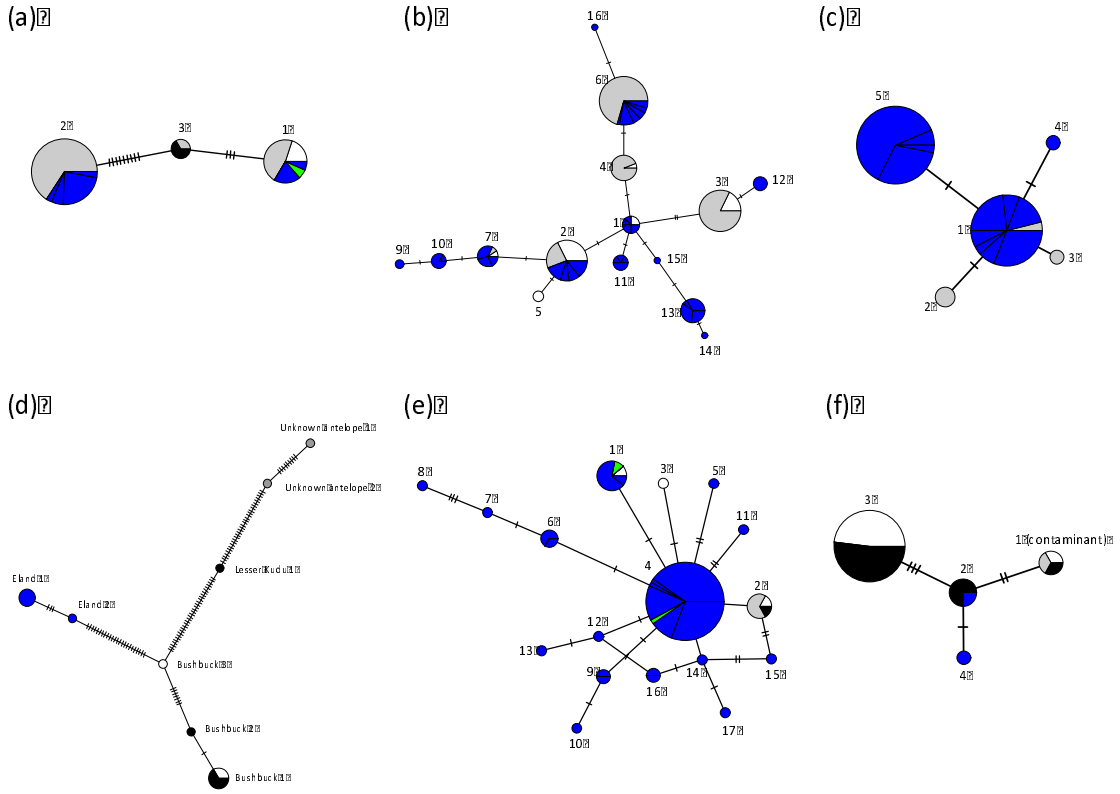

Figure 3. Minimum spanning networks indicating intraspecifc diversity and relative frequency of haplotypes between populations from this study, combined with previous blood meal analyses in Kenya and Tanzania. Intraspecific variation for: a) elephants; b) buffalo; c) giraffes; d) antelope; e) warthogs; f) humans. Note that human type 1 matched the ethnic origin of the main investigator (Asian); samples with this haplotype were considered as contaminants and excluded from analyses. The two antelope sequences labelled “unknown” were found only in single clones, with a more dominant host predominating, and had no close match using BLASTn. Circle sizes are proportional to the frequency of each haplotype (see Supplementary Table 7 for values); notches on branches indicate the number of nucleotide substitutions separating haplotypes; colours represent the population of origin (white = Buffalo Ridge; grey = Mukinyo; black = Zungu Luka; blue = Serengeti (from Auty et al. 2016; Muturi et al. 2011); green = Nguruman (from Muturi et al. 2011).