TSETSE FLY EVOLUTION, GENETICS AND THE TRYPANOSOMIASES - A REVIEW

E. S. KRAFSUR[1] and IAN MAUDLIN[2]

[1]Department of Entomology, Iowa State University, Ames, IA 50011, USA; [2]School of Biomedical Sciences, The University of Edinburgh

1. Introduction

African trypanosomiasis in humans (HAT) causes high levels of mortality among infected people, and tsetse flies (Diptera: Glossinidae) are the only vectors of HAT caused by *Trypanosoma brucei*.. As discussed below, extensive epidemics have occurred in the past and the disease remains a clear and present danger to diverse populations in sub-Saharan Africa. It is for this reason that HAT has been designated a "neglected tropical disease" by the World Health Organization and a modicum of support for trypanosomiasis research has followed.

Where do the principal trypanosomiasis research interests lie, broadly? A PubMed survey on 22 June 2017 reaching back to the year 1912 disclosed 9,858 citations for '*Trypanosoma brucei'* and 3,025 for '*Glossina*', a ratio of 3:1. Numbers obtained for only the past ten years are: 3,457 for the trypanosomes and 769 for *Glossina*, a ratio of c. 4.5:1. Why might these numbers be so? (1) Trypanosome stocks can be propagated in the laboratory and shared worldwide so laboratories everywhere can undertake research. (2) Tsetse are difficult to culture, requiring space, expertise, and frequent blood feeding. Moreover, tsetse have long life cycles (generation times of at least 50 days at tropical temperatures) and very low reproductive rates, a female producing no more than two puparia in her first 25 days of life. (3) Funding agencies favor molecular studies best performed in modern Western laboratories and it is no accident that laboratory workers do also. Entomological work in the field, on the other hand, does not attract much financial support. Because tsetse are exclusively African, field work by Western scientists becomes extraordinarily expensive, inconvenient, and time-consuming. Promised results from field work are inherently uncertain and such work in not favored in peer review of proposals largely because most scientists are laboratory workers without field experience. But HAT occurs in the field and that is where HAT transmission must be defeated. A limiting factor in field work is adequate sampling. We raise the issue of adequate sampling

recognizing two components: spatial and genetic, both of which apply to tsetse and trypanosomes.

Since the last review of tsetse genetics in relation to the trypanosomiases (Krafsur, 2009) important advances have been made in particular the sequencing of the genomes of *Glossina* spp., *Trypanosoma brucei* sensu lato, *T. cruzi*, and *Leishmania major*. There have also been significant improvements in the control of tsetse and African trypanosomiasis to the point where the World Health Organization (WHO), has indicated that elimination of human African trypanosomiasis has become an achievable goal. The Bill and Melinda Gates Foundation have sponsored several projects related to African trypanosomiasis (human African trypanosomiasis or HAT and animal African trypanosomiasis or AAT) and reduction in its prevalence.

Recent general treatments of tsetse and the trypanosomes they transmit include: for *Glossina* sensu lato.: Wamwiri and Changasi (2016); Benoit et al. (2015); *G. palpalis* group: Solano et al. (2010); De Meeûs et al. (2015); for *G. fuscipes fuscipes*: Aksoy et al., 2013; for *morsitans* group (Krafsur, 2009); for trypanosomes: Ponte-Sucre (2016), Aksoy et al. (2014); and for African trypanosomiasis: Welburn et al. (2016); Auty et al. (2016); Simmaro et al. (2015), Cecchi et al. (2015). Eco-physiological perspectives are offered by Terblanche et al. (2009), Kleynhans and Terblanche (2009) and De Meeûs et al. (2015). A review of mathematical models of human African trypanosomiasis epidemiology is provided by Rock et al. (2015).

2. The trypanosomiases

2.1 *Human African Trypanosomiasis* (sleeping sickness, HAT)

HAT is a severe meningoencephalitic disease that develops after an acute lymphatic infection. HAT comprises two distinct diseases: Gambian sleeping sickness (caused by *Trypanosoma b. gambiense* – gHAT) is a chronic disease that occurs in West and Central Africa. Rhodesian sleeping sickness (caused by *T. b. rhodesiense* - rHAT) is acute and rapidly fatal. It occurs in East and Southern Africa. The two clinical forms of HAT correspond to morphologically identical subspecies of *Trypanosoma brucei*, *T. b. gambiense* and *T. b. rhodesiense*, T. *b. rhodesiense* is a zoonosis and its HAT prevalence is said to be about 5% of the total. *T. b. gambiense* is not a zoonosis. Recent genetic studies on these trypanosomes call into question the longstanding designation of *T. brucei* 'subspecies' (see section *4.1*, *Genetics of Trypanosomes*).

Estimates from geo-referenced epidemiological records of HAT, combined with human population estimates, indicate that approximately 70 million people occupying an area of 1.55 million km^2 are at varying levels of risk of contracting HAT (Simarro et al., 2012). HAT is endemic in 36 African countries and approximately 250 foci have been detected over the past century, many of them long lasting and, therefore, stable.

The annual incidence of HAT in the 1990s was estimated to be 300,000 but that had declined to 70,000 in the early years of the twenty-first century. These data include approximated estimates of unreported cases which, because of the nature of HAT distribution (often in remote and isolated populations), may comprise 40% of these figures (Fèvre et al., 2008). Sustained control efforts have further reduced the number of new cases. In 2009 the number reported dropped below 10,000 for the first time in 50 years, and in 2015 there were 2804 cases recorded, the least in 76 years (WHO, 2017).

2.2. *Animal African Trypanosomiasis* (*Nagana, AAT*)

AAT is endemic in 36 countries in sub-Saharan Africa and caused principally by four species of trypanosomes: *Trypanosoma* (*Trypanozoon*) *brucei* (including its 'subspecies'), *T.* (*Nannomonas*) *congolense*, *T.* (*Duttonella*) *vivax* and *T.* (*Nannomonas*) *simiae*. There are also other, important trypanosome species including *T.* (*Trypanozoon*) *evansi* and *T. (Trypanozoon) equiperdum* both of which were designated as subspecies of *T. brucei* (Lai et al., 2008). AAT causes a mortality in excess of a million cattle yearly (Shaw, 2004). Rock et al. (2014) provide a trenchant summary of AAT.

2.3. *Susceptibility to Trypanosome Infection in Tsetse*

Of the seventy million people estimated to be at risk of HAT, most (81%) are at risk to gHAT of which 80% are in the Democratic Republic of Congo (Simarro et al., 2012). WHO suggest that disease transmission will have stopped altogether by 2030 (Simarro *et al*., 2015) but Welburn et al. (2016) review historical and epidemiological evidence that strongly questions that optimistic assessment. and discussed further in Section 8, below. The link between tsetse and sleeping sickness was established by Bruce following his work on the alarming epidemic in Uganda that caused thousands of deaths around the turn of the 20th C. Bruce concluded that tsetse were the sole means of disease transmission (Bruce *et al.* 1903). Fly belts of different tsetse species extend widely across sub-Saharan Africa (Rogers and Robinson, 2004), but the distribution of HAT is sharply divided East-West between rHAT and gHAT, with the Rift Valley as the boundary (Welburn et al., 2001a). Furthermore, HAT is found in relatively small endemic

foci that have remained remarkably stable for as long as records have been kept (Simarro et al., 2010). It is surprising that the distribution of sleeping sickness cannot be simply overlaid on to distribution maps of tsetse.

Most species of tsetse are capable of vectoring trypanosomes, but none are especially good vectors, evidenced by their low natural infection rates. In a meta-analysis of relevant literature, Abdi et al. (2017) reviewed trypanosome detection methods and prevalence in 12 species of tsetse sampled in the field and seven species in laboratory contexts. The authors concluded that dissection of adults with microscopic examination was the most sensitive method of detecting trypanosomes. The overall (including gut, head and salivary glands) trypanosome prevalence was 10.3% among flies sampled in the field. But few of these would have been infective.

Many factors have been shown to influence tsetse susceptibility to infection (Dyeret et a*l*., 2013) including: (i) fly age (Welburn and Maudlin, 1992) (ii) tsetse genetics (iii) tsetse symbionts (Geiger et a*l*., 2013) and (iv) the tsetse immune system (Weiss et al., 2013). The involvement of tsetse genetics in susceptibility to trypanosome infection was first demonstrated in *G. m. morsitans* (Maudlin, 1982) and then in other Morsitans group sub-species (Moloo *et al*., 1998) with susceptibility shown to be a maternally inherited character. Isofemale lines of *G. m. morsitans* bred over many generations (Maudlin and Dukes, 1985) and when infected the mean midgut infection rates of *T. b. brucei* and *T. b. gambiense* were, respectively, 56% and 55% (Maudlin et al., 1986). This is remarkable when compared with *T. brucei* s.l. infection rates in natural populations of <1% (Harley, 1967). *T. b. brucei* is uninfective to humans but is closely related to *T. b. rhodesiense* (Balmer et al., 2011). Susceptibility was linked to maternally inherited 'rickettsia-like organisms' which were isolated and later classified as the bacterium *Sodalis glossinidius* (Dale and Maudlin, 1999). *Sodalis* were also linked to susceptibility to trypanosome infections in wild (Maudlin and Ellis, 1985; Farikou et al., 2010; Soumana et al., 2013). Experimental work showed that susceptibility to trypanosome infection could be reduced by subjecting successive generations of tsetse pupae from an isofemale line of *G. m. morsitans* to lower temperatures which resulted in a loss of *Sodalis* in the emergent adult (Welburn & Maudlin, 1991). Geiger et al. (2005; 2007) suggested that fly vector competence may be linked to the genotype of *S. glossinidius.* Hamidou Soumana et al. (2014) compared the transcriptome of *Sodalis* strains from trypanosome infected and uninfected tsetse and found that when trypanosomes are taken up by they altered gene expression in host fly symbionts; the most

frequently found categories related to genes coding phage lysozyme proteins. Phage had been observed by electron microscopy in *Sodalis* cultures and tsetse puparia (Welburn, 1991). Hamidou Soumana et al. (2014) concluded that bacteriophage promote refractoriness to trypanosome infection by decreasing *Sodalis* density and triggering the fly's innate immune response. In susceptible, phage would retain a lysogenic cycle and *Sodalis* densities would not be reduced, hence increasing fly susceptibility. This model proposes that a phage (or more than one) is produced from a prophage lying quietly in the *Sodalis* genome that helps to generate the normally refractory phenotype of tsetse. This would require some sort of signal that causes the phage to move from the quiescent lysogenic prophage state to the active fully formed lytic state then burst out of the *Sodalis* producing changes in the fly gut. It remains unclear what precise changes are required in the midgut to kill the trypanosomes there. *Glossina* midgut infection rates can be significantly increased by adding antioxidants to the infective feed, inhibiting the effects of reactive oxygen species (Macleod et al., 2007). It is unclear if such experimental conditions mimic the mode of action of *Sodalis*-related susceptibility in tsetse or simply inhibit oxidative stress in the fly midgut that would lead to a cell death response in trypanosomes (Figarella et al., 2006; van Zandbergen et al., 2010).

2.4. *Distribution of HAT and Tsetse*

It is opaque how the relatively small HAT foci were generated and why they have not, over time, spread across the vast fly belts of Africa. Ford (1971), in his societal analysis of the trypanosomiases, linked the epidemics of sleeping sickness seen across Africa in the 20th C to the introduction of intensive systems of agriculture (e.g. cash cropping of cotton) by colonial regimes that disrupted a centuries old ecological balance that encouraged the spread of tsetse. Krafsur (2009) suggested that tsetse fly/trypanosome genetic co-adaptations have evolved to be closely defined by habitat. The recent spread of rHAT from its ancient focus in the SE towards the NW of Uganda has been attributed to northwestern gene flow between tsetse populations (Aksoy et al., 2013) and evaluated further by Opiro et al. (2017). Echodu et al. (2015) suggested that it was simply movement of *Glossina f. fuscipes* that caused emergence of disease in new foci in Uganda. There is good evidence, however, that the spread northwards of rHAT in Uganda is linked primarily to re-stocking programmes introducing infected cattle, a reservoir of rHAT, into previously disease-free areas (Fèvre et al., 2001; Batchelor et al., 2009), rather than movement of tsetse populations.

The distribution of gHAT across west and south-west Africa presents a more complex problem in tsetse fly biology given that the present parasite populations of *T. b. gambiense* Group 1, estimated from genomes of 75 diverse isolates, have arisen from a single individual within the last 10,000 and more likely 1,000-1,500 years ago (Weir et al., 2016) that expanded across West and Central Africa. The most obvious vehicle for the dispersal of this clone would have been the Bantu expansion, one of the most significant developments in African history, that took place around 4000–5000 years ago (Patin et al., 2017). The Bantu expansion saw several waves of migration moving south through the rain forest from what is now the border between Nigeria and Cameroon then diverging into two branches: one moved south and west, the other to East and south-eastern Africa. The expanding Bantu population would have taken with them their parasites, including clonal populations of *T. b. gambiense* Group 1 which would later be responsible for the catastrophic epidemics of sleeping sickness seen across West Africa in the 19th and 20thC (Simarro et al., 2008). It has been suggested that transmission of this *T. b. gambiense* Group 1 clone does not involve tsetse but rather that the parasite is normally transmitted maternally from infected mothers to their offspring (Welburn et al., 2016). With this model, the single primordial trypanosome would have been transmitted maternally from a Bantu mother in West Africa and then passed down through generations and spread with the Bantu expansion. We now know that *T. b. gambiense* Group 1 infections may be carried for more than 20 years (i.e. for a generation) without significantly impairing the health of the person infected (Kaboré et al., 2011). The importance of cryptic infections was recently addressed (Buscher et al., 2018). Trypanosomes in these asymptomatic carriers may go undetected, cryptically living perhaps under the skin (Capewell et al., 2016), and transmitted vertically from mother to child (Lindner and Priotto, 2010). Did *T. b. gambiense* abandon mating in tsetse salivary glands sometime in the last 10,000 years or did this clone abandon vector flies altogether? (see *4.1.5. Sex, trypanosomes and tsetse*). It is important to observe that all economic analyses of gHAT control assume tsetse transmission (Pandey and Galvani, 2017; Sutherland et al., 2017). Transmission that does not involve tsetse has serious fiscal implications for programmes designed to eliminate gHAT by 2030 (Simarro et al., 2015). Rather than looking for tsetse-related explanations for the limitations of gHAT foci it may, therefore, be more profitable to look at genetic differences between human populations (Bucheton et al., 2011; Courtin et al., 2013).

3. Systematics and Evolution

3.1. *Genetics of Trypanosomes*

3.1.1. *Sampling*

Selection bias (Jamonneau et al., 2003) is an important concern in studying the genetics of trypanosomes because samples from nature are not random, nor is their successful isolation. Unrepresentative trypanosome phenotypes and genotypes may be preferentially (but inadvertently) selected in propagation. Moreover, passage in laboratory animals over many generations without cycling through their natural vectors may introduce another selective bias. Further, many laboratory colonies of tsetse vectors are of longstanding age such that genetic drift or selection may make laboratory vector-parasite relationships unrepresentative. Newer PCR-based methods for sampling in the field trypanosome microsatellite and single nucleotide polymorphism genotypes now allow representative sampling without the necessity of concentrating parasites and passaging through mice, thereby opening the way to examining trypanosome population structure (e.g., Balmer et al., 2011; Koffi et al., 2007, 2009; Duffy et al., 2013; Goodhead et al., 2013).

3.1.2. *Genetics*

In common with other Kinetoplastida (unicellular flagellated eukaryotes), *T. brucei* possess a 'primitive' mitotic apparatus in which the chromosomes do not condense. *T. brucei* s.l. possesses N = 11 large diploid chromosome pairs (of 1-6 Mb each) that possess evolutionarily unique (Akiyoshi and Gull, 2013) kinetochores; 1-5 chromosomes of intermediate size (200-900 kb) and unknown ploidy, contain sub-telomeric expression sites for variant surface glycoprotein (VSG) genes (Van der Ploeg et al., 1984), and *c.*100 transcriptionally silent mini chromosomes also of unknown ploidy (30-150 kb). The intermediate and B chromosomes do not possess obvious kinetochores and their evolutionary origins and significance are unknown (Akiyoshi and Gull, 2013; El Sayed et al., 2000). The mitochondrial genome, however, is condensed and occurs in the kinetoplast and connected to the flagellar basal bodies (Ogbadoyi et al., 2003).

Peacock et al. (2011) directly observed meiosis in clonal and mixed infections of *T. b. brucei* and *T. b. rhodesiense* in tsetse salivary glands by using transfected green and red fluorescent protein expression genes. One hundred nineteen simple sequence repeats (SSRs, or microsatellites) were identified in the *T. brucei* genome and primers were constructed and used to establish a linkage map in *T. brucei gambiense* (Cooper et al., 2008). Eleven linkage groups were thereby demonstrated and linkage maps constructed in *T. b. gambiense*. Synteny

(possessing chromosomal segments in common) was demonstrated between genetic maps of *T. b. brucei* and *T. brucei gambiense*. Moreover, meiosis was genetically demonstrated in *T. b. brucei* and *T. b. rhodesiense* but not in *T. b. gambiense* (Jenni et al., 1986; Peacock et al., 2016), consistent with exclusively clonal reproduction, discussed further in section *4.1.5*.

3.1.3. *Evolutionary Based Classification*

Classification of trypanosomes was addressed by Adams et al. (2010), Gibson (2007), and Stevens and Brisse (2004). The earlier classification was that trypanosomes were placed in subkingdom Protozoa, phylum Sarcomastigophora, class Zoomastigophorea, order Kinetoplastida, suborder Trypanozoon, and family Trypanosomatidae. A new higher classification, however, has been proffered: Domain Eucaryota; supraphylum Excavata (formerly Protista), considered to be basal Eukaryotes and encompassing six phyla; phylum Euglenozoa; subphylum Glycomonada; class Kinetoplastea; subclass Metakinetoplastina; order Trypanosomatida; family Trypanosomatidae (Cavalier-Smith, 2016). The basis for this new classification, lies in cellular diversities (e.g., the ultrastructure of the flagella), and 18S ribosomal DNA. Cavalier-Smith (2016) suggested that further distinct body plans and ultrastructural details remain to be found so further taxonomic revisions may be expected.

Trypanosoma systematics and evolution are also in a state of flux because molecular methods allow the detection of much higher levels of genetic diversity and divergence than traditional taxonomy. Numerous new species and strain-specific genotypes have been recorded and placed in the GenBank database. They have allowed the construction of phylogenetic trees (e.g., Adams et al., 2010, Gibson, 2007). Much genetic variation has recently been detected in trypanosome isolates, and resistance to trypanocidal drugs is now a serious constraint in treating HAT (Fairlamb et al., 2016; Barrett et al., 2011). Stevens and Brisse (2004) offered a comprehensive review of trypanosome phylogenetic relationships. The ten African trypanosome species borne by tsetse were clustered in the Salivarian clade of four subgenera, i.e., *Duttonella*, *Trypanozoon, Nannomonas and Pycnomonas*. Salivarians are so named because they develop in the anterior guts and terminate in the salivary glands of their vectors. Their evolutionary relationships are unknown.

The genomes of *T. brucei*, *T. cruzi* and *Leishmania major* have been sequenced (Berryman et al., 2005; Ivens et al., 2005). A set of CD-ROMs entitled *Trypanosomatids: genomes and biology* are available from WHO/TDR, 20 Avenue, Appia 1211, Geneva 27 Switzerland. The CDs contain the genomic sequences of *T. brucei, T. cruzi* and *L. major*;

software for analysis of these sequences, and the genome papers as published in *Science* (2014). Trypanosomatid genome sequences are also available from TriTrypDB (http://tritrypdb.org; e.g. release 37, 25 April 2018)).

The genus *Trypanosoma* is considered monophyletic, descending from a common ancestor *c.*100 million years ago (Stevens and Gibson, 1999). Nevertheless, the systematics of *Trypanosoma* remain fluid. For example, *T.* (*Trypanozoon*) *equiperdum* and *T.* (*Trypanozoon*) *evansi* were reduced in rank to mere strains of *T. brucei* based on multiple molecular and classical parasitological criteria (Lai et al., 2008). A more recent, and wider, evaluation of likely *Trypanosoma* phylogenies was based on the highly conserved heat-shock protein 70 gene, loci for 18S rDNA and glycosomal glyceraldehyde-3-phosphate dehydrogenase (G3PDH) among 65 strains in 11 *Trypanosom*a species (Fraga et al., 2016). Fraga et al. recognized nine clades; our interest here is in the Salivarian clade and the *Trypanozoon* clade (formerly subgenus) in which there is no clear separation of *T. b. brucei*, *T. b. rhodesiense*, *T. b. gambiense*, *T. evansi* and *T. equiperdum*.

Salivarian trypanosomes are extracellular parasites in their mammalian hosts. A key property of these trypanosomes is their expression of antigenic VSGs that defeat adaptive host immune reactions. VSGs are expressed on the parasite's membranous surface with turnover after cell division in about one per thousand thereby ensuring high levels of immunogenic diversity (Mugnier et al., 2015; Ponte-Sucre et al., 2016).

Hoare (1972) proposed a classification of *Trypanosoma* based on gross morphology, geography, clinical presentation, and host range. But biochemical and molecular criteria have since been applied with subsequent confounding of evolutionary relationships among taxa particularly of *T. brucei* sensu lato. *T. brucei* s.l. was said to consist of three 'subspecies' that were identical morphologically but geographically and clinically distinct (Gibson and Stevens 1999 for review). Those distinctions broke down as sampling became more extensive. The medical and economic importance of *T. brucei* s.l. has led to the application of numerous molecular and biochemical criteria including allozymes, kinetoplast/mitochondrial cytochrome oxidase I (Balmer et al., 2011), microsatellites (Balmer et al., 2006), single nucleotide polymorphisms (SNPs), and genes for 18S ribosomal DNA and G3PDH (e.g., Stevens and Gibson, 1999), all with phylogenetically variable results. One conclusion was that trypanosomes should be grouped only into clades rather than into a consensus defined phylogeny. Such an intermediate result can then await the application of better classification criteria (e.g. conserved

genetic markers). Better criteria are now available by the development of SNPs derived from the whole genome sequencing of two isolates of *T. b. rhodesiense*, one from an HAT focus in Busoga, Uganda and the other a Zambezi strain from southern Africa (Goodhead et al., 2013). The SNPs and SSRs were used to examine haplotypes from 31 Ugandan *T. b. rhodesiense* samples and 32 controls isolated from West African human blood samples that consisted of *T. b. rhodesiense* and 5 type 1 and 12 type 2 *T. b. gambiense*. Some haplotypes were shared leading to the inference that genetic exchange among subspecies had occurred sometime in the past. A more geographically and numerically extensive sampling was performed by Echodu et al. (2015) in a survey of 17 SSR loci among 269 *T. bruce*i samples (78% *T. b. rhodesiense*, defined by the presence of the SRA gene) from Uganda and southwestern Kenya. Evidence of admixture among three genetic clusters was inferred suggesting a small measure of gene flow attributed to genetic recombination. This matter is discussed further in section *4.1.5*.

3.1.4. *Population genetics of* Trypanosoma brucei

Most *T. brucei* s.l. reproduce clonally (Tibayrenc and Ayala, 2002, 2013) although sexual reproduction has been reported in *T. b. brucei* and *T. b. rhodesiense*, as we have seen. Both forms of HAT present a spectrum of clinical symptoms and severities, so establishing firm correlations of trypanosome genotypes with disease phenotypes could provide a valuable means of typing the parasites and diseases without extensive clinical examination.

Particularly well-studied is the *Serum Resistance Associated* gene (SRA) the expression of which confers trypanosome resistance to a human trypanolytic substance termed 'trypanosome lytic factor' and accounts for the infectivity of *T. b. rhodesiense* (Xong et al., 1998). Indeed, SRA defines the *T. b. rhodesiense* 'subspecies'; SRA is not present in *T. b. brucei* nor has it been detected in *T. b. gambiense*. *T. b. rhodesiense* occurs in south and East Africa. SRA is related to the VSG gene family but is intracellular (Bart et al., 2015). The geographical origins of SRA were controversial (e.g. Leak, 1999) and remain unclear (Capewell et al., 2015).

Two types of *T. b. gambiense* were identified that differ phenotypically and genotypically (Gibson, 1986, Hide et al., 1990). *T. b. gambiense* group1 is resistant to the human serum resistance factors trypanosome lytic factor 1 and T. *b. gambiense* group 2 is resistant to trypanosome lytic factor 2 (Capewell et al., 2011) much as in *T. b. brucei*. Genetic research utilizing allozyme loci in numerous trypanosome isolates demonstrated the two immunogenic groups could reliably be distinguished.

As noted in section *4.1.1*., sampling bias makes many measures of trypanosome gene flow unreliable. Another sampling problem is that numerically adequate samples from particular locations and times are unavailable in the large repertoire of available trypanosome samples originating from infected people sampled over many years and locations. The spatial and temporal components of trypanosome population gene frequencies are not well estimated. Associations between *T. b. rhodesiense* genotypes, geography, and clinical presentation were offered by MacLean et al. (2004). De Meeûs and Balloux (2005) estimated a mean F_{IS} = - 0.5, and a mean F_{ST} = 0.29 among *T. b. rhodesiense* samples from Uganda, Kenya and Zambia. F_{IS} estimates random mating among individuals in subpopulations and F_{ST} estimates random mating among subpopulations. Here the negative F_{IS} is indicative of clonal reproduction.

A number of genotypic surveys have been performed on *T. b. gambiense* in West Africa. Koffi et al. (2007) used SSRs and PCR to estimate trypanosome genetic diversities in fresh human blood samples from an HAT focus in Côte d'Ivoire and earlier field and reference samples. Mixed infections (i.e., from superinfected patients) were detected. The great advantage of using PCR-based methods is that selection bias is avoided, i.e., parasite isolation, concentration and propagation methods are unnecessary while simultaneously increasing sampling efficiency and sizes. Later work by Koffi et al. (2009) increased genetic sampling of HAT patients in Côte d'Ivoire and Guinea with important results. Linkage disequilibrium measures the degree of non-random association of alleles at different loci. Strong linkages were detected in all subsamples, suggesting clonal reproduction. Random mating and gene flow are usually measured with Wright's *F* statistics. Koffi et al. (2009) estimated F_{IS} (measuring departures from random mating among individuals in subpopulations) and F_{ST}, measuring departures from random mating among subpopulations thereby providing a measure of differentiation among trypanosome subpopulations. Taken over 3 subpopulations and 3 years of sampling, the mean F_{IS} = - 0.62 indicating clonal reproduction and little, if any, sexual recombination. Moreover, strong linkage disequilibrium also suggested clonal reproduction. Genetic differentiation F_{ST} between Côte d'Ivoire and Guinea was 0.2 – 0.3, suggesting no detectable movement of parasites between the countries. The chief trypanosome vectors were *G. p. palpalis* in Côte d'Ivoire and *G. p. gambiense* in Guinea, raising the possibility of specific vector-parasite coadaptations. Estimation of trypanosome subpopulation size and effective clonal population numbers depend on knowledge of mutation rates and HAT prevalence. *T. b. gambiense* was said to have evolved asexually from a single progenitor some 1,000-1,500 years

ago (CL 750-9,500 years) and its mutation rate 2 x 10^{-8} mutations per nucleotide per year (Weir et al. 2016).

Whole genome sequencing of a *T. b. rhodesiense* strain was compared with genomic data from *T. b. brucei* and *T. b. gambiense* (Sistrom et al., 2016). Only the SRA and 3 other genes (among 6,000 open reading frames) were confined to the *T. b. rhodesiense* sample. It was concluded that there is extremely minor genetic differentiation between *Trypanosoma* 'subspecies'. Thus, a plethora of research confirms very close affinities of *T. b*. s.l. (Balmer et al., 2011; Capewell et al., 2011; Echodu et al., 2015; Goodhead et al., 2013; Koffi et al., 2007; Sistrom et al., 2016).

Although genetic recombination between *T. b. brucei* and *T. b. rhodesiense* has been demonstrated, it seems to be rare in nature; clearly they do not have separate evolutionary trajectories but estimation of gene flow among the two 'subspecies' could provide valuable epidemiological information. The question of spatial and temporal genetic differentiation of *T. brucei* s.l. can be directly addressed because most reproduction is clonal, and sexual reproduction uncommon.

3.1.5. *Sex, trypanosomes and tsetse*

Only mature infections, found in the salivary glands and mouthparts of tsetse, are of epidemiological significance for HAT - not all midgut trypanosome infections mature to a salivary gland infection. Factors of epidemiological relevance relating to maturation include: (i) *Fly sex* - male tsetse are more likely to produce mature infections than females (Maudlin *et al*., 1990). This implies that female tsetse would contribute less to vectorial capacity but females live longer than males in the wild (Phelps and Vale, 1978); (ii) *Trypanosome phenotype*, particularly infectivity to humans. Non-human infective *T. b. brucei* are twice as likely to mature to salivary gland infections in tsetse than human infective *T. b. rhodesiense*. Infectivity to humans has apparently evolved at the expense of transmissibility in tsetse (Welburn *et al*., 1995; Milligan *et al*., 1995) with resultant fitness costs (Coleman and Welburn, 2004) and (iii) *Tsetse survival* - salivary gland infections can significantly shorten the life of a fly (Maudlin *et al*., 1998).

In the laboratory, transmission of *T. b. rhodesiense* is straightforward but transmission of *T. b. gambiense* is not: midgut infections of *T. b. gambiense* are easily produced in the laboratory but, salivary gland infections are rarely seen (Maudlin *et al*., 1986; Ravel *et al*., 2006; Janelle *et al*., 2009). In an analysis of field data from Nigeria and Cameroon, the frequency of salivary gland infections was either low (<1.0%) or non-existent (Jordan, 1961). Before the development

of molecular tools, the human infective trypanosomes could not be distinguished from *T. b. brucei* but this is now possible with PCR. Recent surveys using PCR tools have simply reinforced the view that salivary gland infections of *T. b. gambiense* are rare (or undetectable), even in areas with a high prevalence of gHAT: Côte d'Ivoire (Jamonneau *et al*., 2004); Cameroon (Farikou *et al*., 2010) and Angola (Truc *et al*., 2011). *T. b. rhodesiense* by contrast is readily transmitted in the laboratory by a variety of tsetse species. *T. b. rhodesiense* salivary gland infections are also found in the field at low frequency (Okoth and Kapaata, 1986; Kutuza and Okoth 1981). Given the difficulties experienced transmitting *T. b. gambiense* in the laboratory, the mechanisms affecting refractoriness to *T. b. gambiense* salivary gland infection in tsetse are difficult to explore.

A controversial area of research has been the question of a sexual cycle in trypanosomes, first proposed by Tait (1980) who, after examining allozyme variation among trypanosome samples, concluded that trypanosomes undergo random mating and recombination. The centrality of tsetse in this process was shown by Jenni *et al.,* (1986) who found that recombinant hybrids were formed only in the insect vector. A series of experiments has since shown that a complete trypanosome meiotic cycle (including formation of haploid gametes) takes place in tsetse (Peacock *et al.,* 2011; 2014). Mating between trypanosomes takes place exclusively in the salivary glands of infected flies, starting as soon as trypanosomes arrive there from the fly midgut with hybrid production continuing for the fly's lifespan (Gibson, 2015; Peacock *et al.,* 2016). These mating experiments have involved only *T. b. rhodesiense* and *T. b. gambiense* (Group 2) [*T. b. gambiense* Group 2 was found in the 1980/90s in Côte d'Ivoire and Burkina Faso, is indistinguishable from *T. b. brucei,* and may now be extinct (Capewell *et al*., 2013; Weir *et al*., 2016)]. Nevertheless, it may still exist in old laboratory samples.

T. b. gambiense Group 1, by contrast, shows no evidence of mating and haploid gametes have not been observed, even though it appears to possess the genes necessary for meiosis (Jackson et al., 2010) that would be expressed during tsetse infection (Capewell et al., 2013). Given the lack of discernible mating, it is not surprising that *T. b. gambiense* Group 1 was found to be clonal (Morrison et al. 2008; Koffi *et al.,* 2009). The clonal theory of parasitic protozoa proposes that in natural populations of many pathogen populations, recombination is not frequent enough to break a pattern of 'preponderant clonal evolution' (PCE) (Tomasini et al., 2014). PCE leads to widespread identical, or nearly identical, multilocus genotypes that are stable over space and time (Tibaryrenc and Ayala, 2013). The idea of PCE in protozoa is hotly debated but a

recent study sequencing the whole genome of many trypanosomes from different geographical locations provided sound evidence that *T. b. gambiense* Group 1 reproduces exclusively asexually (Weir *et al.,* 2016).

Whether trypanosomes can reproduce sexually is of more than academic interest. It has been suggested that HAT could spread across sub-Saharan Africa because of mating between human infective and uninfective *T. b. brucei* (Gibson et al., 2015), an alarming prospect. In the case of *T. b. rhodesiense*, the SRA gene might be donated to *T. b. brucei* when a tsetse fed from an animal infected with both parasites (or separate blood meals from an animal infected with *T. b. brucei* and from a human infected with *T. b. rhodesiense*); trypanosomes could then mate in the salivary glands (Peacock et al., 2014).

It is not clear, however, that the foregoing hypothetical scenario is a real threat because there are reasons to suppose that trypanosome mating rarely happens in the field. As we have seen, sleeping sickness is confined to discrete geographical foci which have changed little over time (Simarro et al., 2010). We would expect the SRA gene to have spread across the fly belts of Africa wherever *T. b. brucei* exists in livestock or wild animals (i.e. from South Africa to the Sahara), but the SRA gene is not found in trypanosomes to the west of the great Rift valley even though tsetse are not greatly restricted by this boundary. The genetics of *T. b. rhodesiense* isolated in the SE Uganda HAT focus have been extensively studied by using a variety of methods. Duffy et al. (2013) concluded that *T. b. rhodesiense* in the SE Uganda focus exists solely as clonal populations with little or no sexual recombination. Echodu *et al*. (2015) examined 269 trypanosome samples from the same region and found evidence of linkage disequilibrium that supports the clonal nature of *T. brucei* s.l. Despite the observed clonality, Echodu et al. (2015) asserted that there is ongoing genetic exchange (i.e., recombination) between *T. b. brucei* isolates from cattle and humans in SE Uganda, presumably via mating in tsetse. Earlier, Hide et al. (1994), using RFLP, found no evidence of genetic exchange between *T. b. brucei* and *T. b. rhodesiense* in SE Uganda, with human infective trypanosomes grouping independently from populations of *T. b. brucei*; subsequent testing of these trypanosomes with the SRA probe confirmed the divide (Welburn, unpublished). If *T. brucei* s.l. were a random mating population then we would expect the human infectivity gene to be spread randomly throughout *T. brucei* s.l. across this HAT focus but such was not observed. SRA remains confined east of the Rift Valley. The opportunity for mating between trypanosomes in SE Uganda is limited because the salivary gland infection rate is small. This is well illustrated by *G.*

f. fuscipes dissected before, during and after the 1980s rHAT epidemic in Busoga, SE Uganda: at the height of the epidemic, salivary gland infection rates were found to be 0.8% (Okoth and Kapaata, 1986), a figure close to the 0.5% observed in 1976, before the epidemic (Kutuza and Okoth, 1981).

How can we reconcile laboratory data showing frequent mating between trypanosomes with seemingly contradictory field data? One clue lies in laboratory technique. Macleod et al. (2007) found that some antioxidants (glutathione, cysteine, N-acetyl-cysteine, ascorbic acid and uric acid) could increase infection rates in tsetse; for example, in *G. m. morsitans* the addition of 10 mM glutathione to the infective blood meal increased midgut infection rates from 15% (control) to 97%. Experiments involving crosses of trypanosomes routinely use such antioxidants to promote infections in laboratory tsetse (Peacock et al., 2011), as too few experimentally infected tsetse would otherwise develop salivary gland infections (Peacock *et al*., 2014). Adding L-glutathione to the infective feed ensures that almost all *G. m. morsitans* will develop a *T. b. brucei* midgut infection and that 30-60% of these midgut infections mature to form salivary gland infections (MacLeod et al., 2007). We can infer, therefore, that in crossing experiments using antioxidants, Peacock et al. (2011) would have produced salivary gland infections in their flies 100 to 1000 times more likely than the <1% chance of a salivary gland infection developing in wild *Glossina* in Uganda (Okoth, and Kapaata 1986; Kutuza and Okoth, 1981; Hide et al., 1996).

There is also the question of the chance of a wild tsetse imbibing a trypanosome at its critical first feed [the unfed or teneral fly is more likely to develop a midgut infection (Welburn and Maudlin, 1992)]. Trypanosome infected livestock may self-cure but a majority remain infected for months or even years, exhibiting low levels of fluctuating parasitaemia (Masake et al., 2002). In Ugandan cattle, the major reservoir of rHAT in SE Uganda (Welburn et al., 2001b), routine diagnosis of trypanosomiasis using classical parasitological approaches shows very poor sensitivity under field conditions due to very low peripheral parasitaemia in naturally infected animals (Balyeidhusa et al., 2012; Ahmed et al., 2013). This problem led to the adoption of PCR based methods that can detect the presence of parasite DNA equivalent to a single trypanosome in 10 ml host blood (Cox *et al*., 2010, but see Abdi et al., 2017). It follows that susceptible tsetse feeding on cattle are unlikely to pick up a human infective parasite in their first blood meal. This contrasts with the conditions used to infect flies for experiments in the

laboratory, when the infective blood meal may contain approximately 8 x 10^6 bloodstream form trypanosomes ml^{-1} (Gibson *et al*., 2015).

The presence of meiotic indicators (Peacock et al., 2014) in trypanosomes does not necessarily indicate that meiosis takes place in nature [see Weir et al., (2015) for a thorough analysis]. Tibayrenc and Ayala (2012) pointed out that successful experiments of genetic exchange only demonstrates the potential for exchange but say nothing of its frequency in natural cycles. Among human infective trypanosomes in SE Uganda, it appears that recombination is not frequent enough to break the observed predominance of clonal evolution.

3.2. *Glossina*

3.2.1. *Classification*

The hierarchy is, Diptera, Brachycera, Calyptratae (16,000 spp worldwide), Pupiparia (= Hippoboscoidea; 660 spp), Glossinidae (c. 31 nominate species and subspecies). Only one genus, *Glossina* Weidemann 1830, is assigned to the family Glossinidae. The Glossinidae are firmly considered a monophyletic group (Grimaldi and Engle, 2005; Peterson et al., 2007). Chief characters of the Pupiparia include adenotrophic viviparity, a modified developmental pattern in which only one oocyte at a time develops into a mature larva within the mother having been provided nourishment via 'milk glands'. All Pupiparia are blood feeders as adults. Apart from the Glossinidae, there are three families – Streblidae, Nicterobiidae (bat flies), and Hippoboscidae ('keds', ectoparasites of mostly birds and some mammals). Evolutionary relationships were investigated among species of these families by using mitochondrial cytochrome oxidase II, nuclear 18S ribosomal DNA, mitochondrial 16S ribosomal DNA, and mitochondrial cytochrome B (Dittmar et al., 2006; Peterson et al., 2007). *Glossina* spp fit snugly into the Hippoboscoidea but the basal relationships of the four hippoboscid families are ambiguous.

We are unaware of recent taxonomic revisions within the Pupiparia. There is no agreed sister group to the Glossinidae, but an extinct sister group has been postulated from the late Eocene Florissant shale in Colorado, USA and further *Glossina*-like fossils from the Oligocene of Germany (Grimaldi, 1992; Grimaldi and Engel, 2005); the fossils date from ca. 30 – 40 MYA. Flies in the hypothetical sister taxon were about twice the size of modern *Glossina* and it was inferred that their distribution was nearly worldwide. Ford (1971) listed 31 species and subspecies. Thirty-one species and subspecies of extant genus *Glossina* were also recognized by

Jordan (1986, 1993) and Leak (1999); 33 taxa were recognized by Gooding and Krafsur (2005). The subspecies are *G. morsitans morsitans* Westwood 1850, *G. m. centralis* Machado 1970, *G. m. submorsitans* Newstead 1910, *G. palpalis palpalis* Robineau-Desvoidy 1830, *G. p. gambiensis* Vanderplank 1949, *G. fuscipes fuscipes* Newstead 1910, *G. f. martini* Zumpt 1933, *G. f. quanzensis* Pires 1948, *G. pallicera pallicera* Bigot 1891, *G. p. newsteadi* Austen 1929, *G. nigrofusca nigrofusca* Newstead 1910, *G. nigrofusca hopkinsi* Van Emden 1944. Adding to Ford's list is *G. austeni austeni* Newstead, 1912, and *G. austeni mossurizensis* Dias 1956. Note the comparatively late dates of description of some subspecies.

It has been argued, based on demonstrable reproductive isolation, that many of the subspecies should be elevated to specific status (Endsley and Krafsur, 2006 for *G. morsitans* s.l.; De Meeûs et al., 2015 for the *G. palpalis* s.l. species complex, but see Dyer et al. (2008) for a conservative, contrary view.

There are four subgenera of *Glossina* sensu stricto. and three seem monophyletic and well established (Gooding et al., 1991; Chen et al., 1999; Gooding and Krafsur, 2004, Dyer et al., 2008) although their phyletic relationships with each other are unresolved (Jordan, 1993, Gooding and Krafsur, 2005). The subgenera are: *Glossina* s.s. Weidemann (Morsitans group), *Nemorhina* Robineau-Desvoidy (Palpalis group), and *Austenina* Townsend (Fusca group). Subgenus *Machadomyia*, to which only *G. austeni* s.l. is assigned, seems ambiguous with morphological and DNA sequences shared with *morsitans* and *palpalis* groups. Thus, *G. austeni*'s position is equivocal, indeed anomalous, because it shares some key characters, biochemical loci, and mitochondrial and genomic DNA sequences with both subgenera (Jordan, 1993; Dyer et al., 2008). *Wigglesworthia glossinidia* phyletics based on 16S rDNA and congruent tsetse ITS-2 sequences support the *Machadomyia* designation (Chen and Aksoy, 1999) as do allozyme data (Gooding et al., 1991). Gooding and Krafsur (2005) offer a fuller treatment of *Glossina* phylogenetics.

Tsetse classification was based principally upon the structures of male and female genitalia. Various molecular characters are generally consistent with the morphologically-based classification (Gooding and Krafsur, 2004, 2005). Diverse studies support the monophyletic origins of the principal three *Glossina* subgenera but their evolutionary relationships with each other are opaque. Carlson et al. (1993) examined phyletic relationships of 26 tsetse taxa based on the molecular structures of cuticular alkenes (waterproofing waxes). More recently, Dyer et al. (2008) used mitochondrial and nuclear DNA sequences to study phyletic relationships among 13

Glossina taxa. Three subgenera were supported (but *G. austeni* was unresolved) and cryptic speciation was indicated within *G. p. palpalis*. Dyer et al. (2008) also reviewed the habitat requirements of the three *Glossina* subgenera and summarized literature bearing on their post-Miocene evolution (i.e., since 5.3 MYA). Dyer et al. (2008) critically evaluated available data pertaining to speciation in *G. palpalis* s.l. and *G. morsitans* s.l., recognized significant genetic structuring, but argued that rather more genetic and morphological data from a wider range of natural populations would be desirable before 'subspecies' should be elevated to specific taxonomic rank.

3.2.2. *Tsetse symbionts*

Three bacterial symbionts of tsetse have been described. They are maternally inherited *Wolbachia pipientis* (Proteobacteria: Alphaproteobacteria, Rickettsiales, Anaplasmataceae, Wolbachieae), *Sodalis glossinidius* (Proteobacteria: Gammaproteobacteria, Enterobacterales, Pectobacteriaceae), and *Wigglesworthia glossinidia* (Gammaproteobacteria: Enterobacterales, Erwiniaceae). *W. glossinidia* is found in the guts of tsetse where it provides essential B-complex vitamins to its hosts and is thus an obligate symbiont and indicative of a long evolutionary relationship. *S. glossinidius* is a microaerophilic, secondary symbiont. Its prevalence and density varies greatly within and among species, and it affects tsetse susceptibility to trypanosome infection (Welburn and Maudlin, 1999). *Wigglesworthia* and *Sodalis* are abundant in host tsetse bacteriocytes, found on the anterior gut in a specialized organ, the bacteriome.

Wolbachia systematics are uncertain and its phylogeny seems independent of host phylogeny (Werren, 1997). A strain isolated from *D. melanogaster*, *w*Mel, has been fully sequenced (Wu et al., 2004). Widespread *Wolbachia* horizontal transfer among and within taxa confounds its use in phylogenetic investigations. *Wolbachia* is incorporated into the *G. m. morsitans* genome at multiple, largely heterochromatic sites including the supernumerary and X and Y chromosomes (Brelsfoard et al., 2014). It is intracellular, found in the germ line of infected host populations and may cause unidirectional cytoplasmic incompatibility (CI, manifested as embryonic deaths) between an infected male and uninfected female. Thus, in a mixed population, infected females have a reproductive advantage over uninfected females. There may also be bidirectional CI when mating partners are infected with different *Wolbachia* strains. Such phenomena hypothetically could lead to speciation via reproductive isolation. More complex effects may obtain and deep coevolutionary relationships are evident from its propensity for horizontal transfer, its genotypic and phenotypic variation, and its diverse effects

on host taxa. It has been suggested that *Wolbachia*-induced CI can be used to effect replacement of conspecific vector populations with trypanosome-refractory tsetse, to be discussed further in section 8.

3.2.3. *Tsetse species complexes*

Species complexes are composed of two or more sibling species among which morphological differentiation is difficult or impossible. Keys to detecting sibling species are principally twofold: crosses between colonies that yield infertile progeny and estimates of within-deme population structure that show a deficiency of heterozygotes and yield significantly large F_{IS} estimates. For selectively neutral genes, an excess of homozygotes and paucity of heterozygotes with regard to Hardy-Weinberg expectations defines the Wahlund effect. F_{IS} indicates a departure from random mating within a defined deme as may occur when two or more morphologically identical taxa are treated as a single species. The occurrence of null alleles, however, can also lead to high F_{IS} estimates so careful inspection of alleles is required. The Wahlund effect led to the finding of sibling species in *G. p. palpalis* and *G. p. gambiense*. Mélachio et al. (2015) designed and established a trapping scheme to detect sibling species of *G. p. palpalis* in Cameroon. It consisted of a central surrounded by five traps. A significant F_{IS} estimate would be indicative of two or more breeding units, thus the Wahlund effect.

Intercrosses among *G. morsitans* s.l. may produce normally fecund and fertile females while other crosses show reduced fecundity and longevity among hybrid progeny. Hybrid males from all possible intercrosses are sterile (Gooding 1990, Gooding and Krafsur, 2005). Most intercrosses necessarily depend on using colonized flies and some crosses were between colonies of different geographic origins. Such replication was obtained in some *morsitans* intercrosses and *palpalis* intercrosses and produced disparate results. Sex chromosome aneuploidy varies among *G. p. palpalis* colonies (Southern, 1980) and can complicate the interpretation of hybrid sterility. Males issuing from *G. p. palpalis* and *G. p. gambiense* intercrosses are sterile (Gooding, 1997) and even crosses among diverse *G. p. palpalis* colonies may be sterile (Gooding et al., 2004). The importance of using replicate colonies of any particular taxon is indicated by the variable results in crossing subspecies, e.g., replication arising from differing geographic origins of the colonized tsetse. Molecular criteria for examining phylogeny of *palpalis* group tsetse were utilized by Dyer et al. (2008, 2009).

The wide and discontinuous distribution of *G. pallidipes*, its variable behaviors (Langley et al., 1984) and chromosomal polymorphisms (e.g., Southern and Pell, 1981) may be

suggestive of cryptic speciation but such cross-breeding experiments that have been attempted among colonized geographic strains have not supported the hypothesis. Allozyme, mitochondrial, and microsatellite genetic diversities among numerous geographically diverse field samples provided substantial estimates of F_{ST} (eg, Tables 2, 3, 4) indicating restricted gene flow among populations, discussed further in section *5.4.* No direct evidence was obtained for speciation among the *G. pallidipes* (Krafsur, 2009). On the other hand, Vanderplank (1948) observed that *G. pallidipes* was morphologically similar to *G. longipalpis* Weidemann 1930 and probably an eastern form of it. *G. pallidipes* has a spotty, discontinuous distribution in southern Africa, East Africa, Ethiopia, and the DRC. *G. longipalpis* occurs in Senegal, Guinea and Guinea-Bissau east to Nigeria and Cameroon (Rogers and Robinson, 2005). We are aware of no genetic work on this form and its evolutionary relationship to *G. pallidipes* and other *morsitans* group forms are untested. The allopatric *G. swynnertoni* Austen 1923 was once considered to be a variant of *G. morsitans* s.l. (Ford, 1971) on morphological grounds and also because s*wynnerton*i mated freely with *G. m. centralis* although they were allopatric. Indeed, field matings were obtained by allowing field-collected *G. m. centralis* pupae to eclose in *G. swynnertoni* habitat (Vanderplank, 1948). It was demonstrated that hybrid male progeny were sterile, and hybrid females variably fecund but mostly sterile.

In the *Nemorhina* there are two major clades, each forming a species complex. There are *G. palpalis* subspecies and *G. fuscipes* subspecies. Recent studies have shown that each 'subspecies' *G. p. palpalis* and *G. p. gambiensis* are composed of two or more genetically and geographically reproductively isolated clades (Gooding et al., 2004; Ravel et al., 2007; Dyer et al., 2009; Cordon-Obras et al., 2014; De Meeûs, et al. 2015) and further sampling may disclose even more.

The *G. fuscipes* species complex includes *G. f. fuscipes*, *G. f. martini*, *G. f. quanzensis* which are allopatric (Machado, 1954) but form an unresolved polychotomy (see Fig. 1). Vanderplank (1948) described mating barriers among *G. fuscipes* subspecies, some due to the structures of genitalia; also some reduction in fecundity in male *G. f. fuscipes* X female *G. f. martini*. Beadell et al. (2010) examined 13 SSR loci among 20 Uganda, a Sudanese, a Kenya, and a Democratic Republic of the Congo (DRC) population identified as *G. f. quanzensis*. The variance statistic F_{ST}, a measure of reproductive isolation among the sampled *G. fuscipes* populations, was 0.11 among the Uganda samples, 0.27 ± 0.13 between Uganda and Sudan, 0.24 ± 0.05 between Uganda and the DRC, and 0.24 ± 0.11 between Kenya and Uganda. F_{ST} between

the DRC and Sudan = 0.32, between Kenya and Sudan = 0.40, and between Kenya and the DRC = 0.45, but the foregoing estimates did not suggest speciation. Dyer et al. (2011) examined the genetic basis of *G. fuscipes* subspecies sampled among 13 locations in five countries: two maternally inherited markers were employed: *Glossina* mitochondrial NADPH dehydrogenase 2 gene (*ND2*), and cytochrome oxidase I (*COI*). Also, maternally inherited symbiont *Wigglesworthia glossinidia* genome fragments were amplified and used to establish their phylogenetic relationships vis-à-vis their host *G. fuscipes* s.l. *Glossina* nuclear genomic samples consisted of sequences of ribosomal internal transcribed spacer 1 (*ITS1*) and a panel of five SSRs. With the exception of *ITS1*, sequence data from nuclear, mitochondrial and endosymbiont genomes rejected one or more of the morphological subspecies in tests of monophyly. Thus, strong support was not obtained for Machado's three subspecies and Dyer et al. (2011) concluded that the morphologically defined *G. f. fuscipes* and *G. f. quanzensis* were polyphyletic.

Also in the *palpalis* group is a third clade, consisting of *G. pallicera pallicera* and *G. pallicera newsteadi*. We are aware of no genetic or systematic work on *G. pellicera* s.l. since their descriptions.

Using wild flies from puparia collected in the field and maintained in the laboratory, Vanderplank (1948) was unable to obtain inter-taxon matings in *Austenina* (Fusca group) but was able to do so within the *Nemorhina* and *morsitans* groups. There is a paucity, however, of cross-breeding attempts among diverse *Austenina* (most constituent taxa have not been sampled or colonized). Considering the *Machadomyia*, intercrossing the two *G. austeni* subspecies has not been reported even though colonies of both exist.

3.2.4. *Coevolutionary considerations*

The occurrence of a mycetome/bacteriome/bacteriocytes that harbour symbionts in the anterior tsetse gut epithelium is, *a priori*, evidence of a coevolutionary host response. *Wigglesworthia* endosymbionts are essential for fecundity and tsetse immune system. Ancient co-evolutionary relationships exist among host tsetse, *Wigglesworthia,* and parasitic trypanosomes. Tsetse with no *Wigglesworthia* are sterile and highly susceptible to trypanosome infection (Weiss et al., 2013). It has long been known that teneral tsetse are much more susceptible to trypanosome infection than older, fed tsetse (e.g., Maudlin and Welburn, 1994).

Chen et al. (1999) examined phylogenies of eight tsetse species vis-à-vis their *W. glossinidius* symbionts. To establish the phylogenies independently, they used the nuclear internal transcribed spacer-2 (ITS-2) for the tsetse and small unit 16S ribosomal DNA for *W.*

glossinidius. The two phylogenies were concordant, thereby ruling out horizontal transmission and demonstrating that the association of tsetse and *Wigglesworthia* was established early in the most recent common ancestor of extant *Glossina*. Clearly, these disparate taxa have closely coevolved. The *W. glossinidia YcfW* gene sequence, on the other hand, did not clearly support monophyly of *G. fuscipes* s.l. (Dyer et al., 2011).

Secondary endosymbiont *Sodalis glossinidius* has had a much shorter coevolutionary relationship with its host *Glossina* relative to that of *Wigglesworthia. Sodalis* isolated from eight *Glossina* taxa had virtually identical *ftsZ* gene sequences (consistent with the findings of Dyer et al., 2011) and no host fitness effects were observed even after performing transfection experiments (Weiss et al., 2006). *Glossina* susceptibility to trypanosome infection seems to vary with *Sodalis* densities, genotypes, and distribution among flies in nature (Farikou et al., 2010) but relationships need to be clarified.

Most members of *Glossina* species complexes are allopatric, probably testifying to their long separation during which random drift, fluctuating climates and differing selective regimes have promoted divergence. In the case of cryptic forms in *G. palpalis palpalis*, *G. p. gambiensis* and *G. f. fuscipes* we may surmise that morphological differentiation has lagged. But even here, geographic isolation may have been the original mechanism that promoted cryptic speciation.

Assuming a divergence rate of 1.5% per million years (MY) in cytochrome oxidase I, Dyer et al. (2011) estimated time of divergence between a DRC sample and other *G. fuscipes* s.l. to be 0.8 to 1.2 MY. A similar kind of estimate applied to *G. p. palpalis* and *G. p. gambiense* was 2.2 to 4.2 MY (Dyer et al., 2009).

Fig. 1. Hypothetical phylogeny of 31 *Glossina* spp, after Gooding and Krafsur, 2004.

4. Genetics and cytogenetics of tsetse

4.1 *Formal genetics and cytology*

In tsetse there are only six morphological markers (Gooding, 1992; Gooding and Krafsur, 2005). Genetic work related to the sterile insect technique (SIT) and chromosomal rearrangements for population control (e.g., Curtis, 1968; Curtis et al., 1972) proceeded without much formal genetic support. Almost two decades later, isoenzyme and allozyme polymorphisms began to be used as markers in laboratory crosses and population surveys

(Gooding, 1992; Krafsur and Griffiths, 1997; Gooding and Krafsur, 2004, 2005). More recently, restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), microsatellites and single nucleotide polymorphisms (SNPs) have also been employed. DNA sequences are increasingly used as the cost of sequencing DNA has declined. Available *Glossina* genomes in GenBank can be mined for new SSRs and SNPs; indeed, such work has already been accomplished (Gloria-Soria et al., 2016, 2018).

Chromosome mapping studies have shown three linkage groups that correspond to the 'L' and X chromosomes (Fig. 2; reviewed in Gooding and Krafsur, 2004, 2005). There is little or no genetic recombination in males, and this is consistent with other Brachycera (e.g., *Drosophila melanogaster* Meigen, *Musca domestica* L.).

Willhoeft (1997) used fluorescent in situ hybridization methods (FISH) to localize 28S ribosomal DNA to its chromosomes in *G. austeni, G. p. palpalis, G. f. fuscipes, G. m. submorsitans*, *G. pallidipes* and *G. brevipalpis*. Bonomi et al. (2011) also utilized FISH to locate certain microsatellite loci to their respective chromosomes.

We are aware of no further work on the formal genetics and cytology of tsetse so reference here must be made to earlier research. The chromosome complement of tsetse species was briefly reviewed by Leak (1998). In that treatment the autosomal, sex, and 'B' (supernumerary, or 'S') chromosomal complements were not distinguished so we have tabulated what is recorded in Table 1. In the earlier work, it was abundantly clear that polymorphisms in numbers of sex chromosomes (i.e., X and Y) and supernumerary ('B') chromosomes existed within and among populations and species of the Morsitans and Fusca groups but have not been reported in Palpalis group tsetse. Moreover, C-banding also varied within and among taxa. C-bands are differentially Geimsa-stained regions of chromosomes that correspond to relatively condensed regions of heterochromatin. Heterochromatin shows regions of gene inactivation in a chromosome. The original studies on mitotic chromosomes have not been followed up and there is only one complete report on meiosis (in *G. m. morsitans*; Southern et al., 1972a) although elements of meiosis in *G. austeni* (Southern et al., 1972b) and *G. pallidipes* have been reported (Southern and Pell, 1981).

Earlier reports indicated highly variable karyotypes among and within *Glossina* spp. But it soon became plain that the basic mitotic karyotype in *Glossina* is 2N = 4 +XY (Table 1, Fig. 3). A pair of large metacentric or sub-metacentric autosomes (L_1 the larger, L_2 the shorter), an X, and a condensed, heterochromatic Y plus zero to 12 'S' (supernumerary) chromosomes

characterized by repetitive 'satellite' DNA (Amos and Dover, 1981) that vary from acrocentric and subacrocentric (centromere near the chromosome end, or telomere) to metacentric (centromere near the center of the chromosome) and tend to stain intensely because of their greater condensation relative to euchromatin. They are thus termed 'pycnotic' or 'heteropycnotic'. A rather different karyotype was recorded in *G. brevipalpis* (Fusca group) where five chromosome pairs were and two to 12 supernumeraries were observed (Willhoeft, 1997). This remains the only published Fusca group tsetse karyotype.

Tsetse Y chromosomes are highly condensed and therefore heterochromatic (Amos and Dover, 1981; Willhoeft, 1997). Sex determination in *G. palpalis palpalis* is similar to that of *D. melanogaster* and the Y chromosome has no detectable influence on the male phenotype). Instead, sex determination depends on the number of X chromosomes (Maudlin, 1979).

Sex determination in tsetse was examined in neuroblasts of Nigerian *G. p. palpalis* by Maudlin (1979) and reviewed by Southern (1980). As in *Drosophila*, aneuploid karyotypes XXY, XXXY, XXYY were phenotypic females and aneuploid XYY, X0 karyotypes were phenotypical males (sex was determined by dissection; their fertilities could not be assessed). These aneuploids were detected in high frequencies (10.4%) suggesting some kind of balanced polymorphism acting in the sampled populations (Maudlin 1979).

Fig. 2 Goes here (see separate file for the figures).

Fig. 3 Goes here. Diagrammatic *G. austeni* male karyotype, after Curtis et al., 1972. 'L' are metacentric or submetacentric autosomes, 'S' are acrocentric supernumeries, 'Y' and 'X' are heterosomes (sex chromosomes). The S and Y are heterochromatic.

Polytene chromosomes arise by repeated chromosome reduplications without cell division ('endomitosis'). Such endoreduplication results in giant cells with highly banded patterns that usually prove to be highly diagnostic of species and phylogenetic relationships. Polytenes may be found in insect salivary glands, Malpighian tubules, trichogen cells, and ovarian nurse cells. They are particularly well developed in the Diptera. They have been used to uncover speciation in the mosquito *Anopheles gambiae* s.l. (e.g., Lanzaro and Lee, 2013) and are especially useful in finding chromosome rearrangements, e.g., duplications, deletions, inversions

and translocations. Tsetse polytenes have been comparatively studied in *G. m. morsitans, G. m. submorsitans*, *G. austeni*, and *G. pallidipes* by using pupal scutellar trichogen cells that give rise to the apical bristles in pharate adults 18-20 day after pupariation (Gariou-Papalexiou et al., 2002, 2007). Polytenes were also characterized in *G. f. fuscipes* (Pell and Southern, 1976) and Morsitans group taxa (Southern and Pell, 1981). Polytenes present three long chromosomes, two of which correspond to the mitotic and meiotic L_1 and L_2 autosomes and to the X. The S and Y chromosomes appear as a unresolved, heterochromatic networks. Three paracentric inversion(s) were detected in each arm of the X chromosome of laboratory cultures of *G. m. submorsitans* by examining polytenes (Garliou-Papalexiou et al., 2002) and may be inferred from linkage studies (Gooding and Challoner, 1999) via suppression of recombination (Fig. 2). Further chromosome rearrangements have been detected in *G. pallidipes*, *G. m. morsitans, G. austeni*, all by examination of polytenes (Garliou-Papalexiou et al., 2002, 2007).

Few studies of karyotypes in natural tsetse populations have been carried out, but such work as published indicates rather lesser frequencies of supernumerary chromosome than found within laboratory cultures (e.g., Southern and Pell, 1981). Moreover, it seems that frequencies of such chromosomes increase with successive generations in culture. Variation in supernumeries, common in numerous *Glossina* spp, possibly may have effects on reproductive fitness and experimental results. Increase in supernumeries over time, because of a putative relaxation of natural selection in culture (Warnes and Maudlin, 1992), subsequent drift, and the reliance on longstanding tsetse colonies would seem to mandate the examination of their current karyotypes. Supernumeries do not synapse during meiosis so their affinities and evolutionary derivation are opaque, although Amos and Dover (1981) hypothesized that they arise from Y chromosome duplication and accumulation of repetitive DNA. Meiotic nondisjunction could produce the superfluous Y and corresponding XO females were observed in *G. p. palpalis* (Maudlin, 1979). B chromosomes are typically heterochromatic and probably untranscribed. Wild flies seem to have fewer B chromosome than laboratory cultures raising the possibility that lab cultures may display biological properties unrepresentative of wild tsetse. Supernumerary chromosomes have not been recorded in subgenus *Nemorhina*.

The significance is unclear of polymorphic heterosomes among progenies of wild Nigerian *G. p. palpalis* (Maudlin, 1979). Some individual wild females produced one aneuploid and two progeny with balanced (‘euploid’) karyotypes. Thus, here was an instance of meiotic nondisjunction in the wild female but that did not explain the high frequencies of aneuploids. X0

males in *Drosophila* are sterile. Sex determination is polygenic and complex in *Drosophila* and *Musca domestica* (Meisel et al., 2016) and probably so in tsetse because they seem to follow the *Drosophila* model. Maudlin suggested that sex ratio distortion may occur in tsetse if their sex determining mechanism follows the *Drosophila* model; indeed, sex-distortion was later recorded among males in lab cultures of *G. m. submorsitans* (Rawlings and Maudlin, 1984) the locus of which lies on the X chromosome (Gooding, 1986). Garilou-Papalexiou et al. (2002) associated the sex distortion with three paracentric inversion in each arm of the X. Maudlin (1979) and Gariou-Papalexiou et al. (2002) suggested that sex distortion could prove useful for the SIT but this suggestion does not seem to have been examined further. Sex-ratio distortion and chromosome polymorphisms in tsetse require further study; samplings in both nature and in laboratory cultures are indicated.

4.2. *Size of* Glossina *genomes*

Sequencing the *G. morsitans* genome provided an estimate of 600,000,000 base pairs (bp), while *G. palpalis* was estimated at 700,000,000 mbp (international *Glossina* genome initiative 2014; Giraldo-Calderon et al., 2015). *G. f. fuscipes* was estimated to have about 20,749 genes (VectorBase 2014; VectorBase website https://www.vectorbase.org/organisms/glossina-morsitans). The Glossina Genomes Consortium have submitted six *Glossina* species listed at the VectorBase website and includes *G. austeni, G. brevipalpis, G. f. fuscipes, G. m. morsitans*, *G. pallidipes*, and *G. p. gambiensis*. Genome sequence lengths varied from 374,774,708 bp in *G. f. fuscipes* to 315,360,362 bp in *G. brevipalpis*. Full details are provided at the VectorBase website. The availability of these data opens a plethora of new opportunities to study in great detail the molecular biology of tsetse and its adaptations.

5. Tsetse population structures

5.1. *Estimation of population structure*

Population structure is most conveniently estimated by Sewall Wright's *F*-statistics that utilize genetic diversity measures (Tables 2, 3, 4). Two measures are particularly useful: F_{IS} and F_{ST}. For selectively neutral loci, F_{IS} estimates departures from random mating within demes and F_{ST} estimates departures from random mating among demes. Random mating expectations

follow the Hardy-Weinberg rule. F_{ST} can be used to estimate gene flow (i.e., immigration and dispersion) within and among demes.

Many recent studies on vector population structure and gene flow have been conducted from the standpoint of specifying subject populations degree of isolation from conspecific demes that may compromise tsetse suppression programmes. Estimation of effective population sizes *Ne* are thought to be helpful in predicting the likelihood of population recovery from an earlier control programme. Solano et al., 2010; De Meeûs et al., 2014; Aksoy et al., 2013; Vreysen et al., 2013 offer reviews of the general subject.

5.2. *Population structure in subgenus* Nemorhina - *the* palpalis *group*

Palpalis group taxa are among the most effective transmitters of African trypanosomes. Most taxa occupy moist habitats and are riverine in savannas along which dispersal may occur. The *G. palpalis* species complex includes *G. p. palpalis* and *G. p. gambiensis*. *G. p. palpalis* occurs in Benin, Togo, Cameroon, Côte d'Voire and Nigeria south to Gabon, Equatorial Guinea, Democratic Republic of the Congo, Republic of the Congo, and Angola (Rogers and Robinson, 2004). *G. p. gambiensis* is confined to Senegal, The Gambia, southern Mali, Sierra Leone, Liberia, Burkina Faso, Guinea and Guinea-Bissau. Estimates of F_{ST} *in G. palpalis s.l.* varied from essentially 0 to 0.43 (Tables 3, 4). Coexisting within the range *of G. p. gambiense*, *G. tachinoides* afforded an estimate of F_{ST} not significantly different from zero when sampled over a 338km course of the River Mouhoun in Burkina Faso (Koné et al., 2010). PubMed lists 127 citations over the last 10 years that refer to *G. p. palpalis.* Seventy-four are listed for *G. p. gambiensis*.

G. f. fuscipes is one of the chief trypanosomiasis vectors in Sub-Saharan Africa (Omolo et al., 2009) particularly in Uganda where it transmits both *T. b. gambiensis* and *T. b. rhodesiense* (Aksoy et al., 2013; Albert et al., 2015;). PubMed lists 54 accessions for *G. f. fuscipes* over the last ten years. Five publications were based on a single genetic survey carried out in four Uganda sites (Beadell et al., 2010; Echodu et al., 2011, 2013, Hyseni et al., 2012; Gloria-Soria et al., 2016). We have already reviewed *G. fuscipes* as an allopatric species complex and found that pre-zygotic reproductive isolation among geographically and morphologically designated members of the *G. fuscipes* species complex has not been confirmed with genetic evidence. Indeed, the genetic basis of such differentiation is unresolved and unclear

(Dyer et al., 2011). Five microsatellite loci suggested no significant reproductive isolation among *G. fuscipes* subspecies although F_{ST} = 0.26 was estimated between West DRC and western Kenya *G. f. fuscipes* (Dyer et al., 2011). Population structures and effective population sizes were extensively and intensively studied in Uganda and western Kenya, first by examination of mitochondrial variation in southeastern Uganda and southwestern Kenya (Krafsur et al., 2008) with further, more geographically extensive studies by Beadell et al., (2010), Echodu et al., (2011), Kato et al., (2015), and Opiro et al., (2016, 2017). Microsatellite loci were also employed by the foregoing workers. Effective population number estimates are discussed further in Section 7. Some genetic statistics are summarized in Tables 2 - 5.

In a technical *tour-de-force,* Gloria-Soria et al. (2016) used single nucleotide polymorphisms (SNPs) to investigate linkage disequilibrium, environmental adaptations and susceptibility to trypanosome infection in Uganda *G. f. fuscipes*. A panel of 73,297 SNPs were used also to investigate population genetics and undertake genome-wide association studies. Estimates of genetic differentiation among populations agreed with earlier estimates based on microsatellite and mitochondrial variation. Patterns of linkage disequilibrium were used in efforts to identify likely genome regions responding to selection, including resistance/susceptibility to trypanosome infection and 'local' environments. SNPs were identified that correlated with trypanosome infection and the environments in which the flies were sampled.

Nemorhina species are 'riverine', confined chiefly to habitats comprising strips of riverine woodland along which tsetse disperse, while showing little movement into adjacent savannah (Rogers and Robinson, 2004). Genetic studies designed to estimate gene flow had suggested linear dispersion in dry seasons and wet season dispersion across watersheds (e.g. Koné et al., 2011) but such estimates were indirect and inferential. Bouyer et al. (2009) compared genetic and mark-release-recapture methods on dispersal of *G. p. gambiensis* in a fragmented landscape in Burkina Faso and found the results consistent with one another. The flies were distributed in small, localized subpopulations with short effective dispersal distances of ~ 1km per generation. Vreysen et al. (2013) used mark, release, and recapture methods directly to estimate dispersal of sterilised *G. p. gambiensis* males and females between river basins in Mali in the year 2004. The flies were colonized in Burkina Faso in 1972 and maintained at CIRDES in Bobo Dioulasso. Only one of 56,000 released flies was recaptured in

an adjacent river basin but this was taken as "proof of principle" that dispersion across river basins normally occurs in riverine tsetse and must be recognized in vector control programmes.

5.3. *Population structure in subgenus* Glossina *(the Morsitans group)*

Morsitans group taxa occur largely in wooded savannas of East and southern Africa. PubMed lists 146 accessions published over the last 10 years. Members of this subgenus are the chief trypanosomiasis vectors in East and southern Africa. They include *G. morsitans* s.l., *G. pallidipes*, and *G. swynnertoni*. *G. swynnertoni* occurs principally in north central Tanzania and in a small south western area of Kenya (Rogers and Robinson, 2004; Nagagi et al., 2017). *G. morsitans* 'subspecies' are strongly allopatric. *G. pallidipes* is widely but discontinuously distributed over East and southern Africa (Rogers and Robinson, 2004); many populations seem to occupy very small areas. Patchy distributions also characterize *G. austeni* and *G. brevipalpis* (Rogers and Robinson, 2004).

Morphological differences among the three *G. morsitans* 'subspecies' are small (Vanderplank, 1949; Machado, 1970) but genetic differences are great (Krafsur and Endsley 2002, 2006). Crosses among *G. morsitans* 'subspecies' have been studied for 70 years (Vanderplank, 1948; Curtis, 1972; Gooding, 1990) with unambiguous results: crosses among them produce reduced fecundity, embryonic lethality and sterile, hybrid male progeny (Curtis, 1972; Gooding, 1993).

Genetic studies on three 'subspecies' of *G. morsitans* included populations sampled in Ethiopia, Kenya, Tanzania, Zambia, and Zimbabwe. Markers used included allozymes (Krafsur and Griffiths, 1997), mitochondria (Wohlford et al., 1999; Krafsur et al., 2000, 2001) and both nuclear (SSRs) and mitochondrial genomes (Ouma et al., 2007). Geographically extensive genetic sampling of *G. pallidipes* was performed by using allozymes (Krafsur et al., 1997), microsatellites (Krafsur and Endsley, 2002; Ouma et al., 2006), mitochondrial sequences (e.g. Table 2, Krafsur et al. 2016) and both SSRs and mitochondria (Ouma et al., 2005, 2011).

The foregoing research using allozymes and microsatellites indicated that random mating (i.e., $F_{IS} \approx 0$) was the rule within populations, but that gene flow among populations of subgenus *morsitans* taxa was restricted (i.e., $F_{ST} >> 0$) and attributed to their inherently patchy distributions. As will be discussed further, the effective population numbers of *morsitans* group taxa are small relative to the regions said to be environmentally favourable. Such populations demonstrate high levels of genetic drift, the direct consequence of generation to generation flux from chance alone in selectively neutral allele frequencies. The question arises, are Morsitans

group populations more or less spatially isolated because they are separated by climatically or biologically unsuitable habitats? Alternatively, are the differences in allelic frequencies caused by differing selective regimes? A third possibility is established by the earlier recurrent Rinderpest epizootic that began late in the eighteenth century, wiping out tsetse mammalian hosts (Ford, 1971) so that many extant Morsitans group populations are centered on earlier refugia. The matter is discussed further in section *8.1*.

5.4. *Subgenus* Austenina - *the* Fusca *group*

PubMed lists only 5 accessions (in the last 10 years) that refer to *G. fusca*. Most Fusca group taxa are confined to West African moist, evergreen forested habitats where dense shade is available (Leak, 1999). Fourteen nominate taxa are included in this group, but few are considered economically significant. The principal taxa include *G. brevipalpis* Newstead 1910, *G. nashi* Potts 1955 and *G. tabaniformis* Westwood 1850. The last two were demonstrated to be serious AAT vectors in Gabon and the Democratic Republic of the Congo (DRC) (Leak et al., 1991).

G. brevipalpis is discontinuously distributed from the DRC, with small areas in the northeast corner of South Africa, Mozambique, Tanzania, and Kenya (Rogers and Robinson, 2004). It seems to require dense thickets and heavy shade which governs its otherwise wide distribution over eastern Africa (Leak, 1999). It can be found also in small pockets of Zambia, Ethiopia and Somalia. *G. brevipalpis* colonies are maintained by the IAEA/FAO in Seibersdorf, Austria and its genome is available from VectorBase. *G. brevipalpis* has received the most entomological attention (27 accessions in the last 10 years) largely because it is recorded as a potent vector of AAT in South Africa where there are plans to eliminate it (De Beer et al., 2016).

5.5. *Dispersion in Glossina*

Estimates of tsetse dispersion in terms of emigration and immigration based on measures of gene flow are not at all equivalent to direct measures established by mark, release and recapture methods. The relationship between estimates of gene flow, via F_{ST}, and dispersion was discussed in Gooding and Krafsur (2004, 2005) and Krafsur (2009). Gene flow refers to surviving, reproducing migrants typically integrated over many generations whose genes are detected in the sampled populations.

Rogers (1977) demonstrated experimentally and analytically that simple random diffusion adequately modelled dispersion among *G. f. fuscipes*. By using Rogers' random walk model, Vale et al. (1984) estimated dispersion of about 1 km per day in *G. m. morsitans* and *G. pallidipes* in the Zambezi valley (Zimbabwe). Rogers' diffusion model is in terms of distance moved *d*, step length *s* (distance moved per day) and *x*, the number of steps: $d \approx sx^{1/2}$. Rogers (1977) tested earlier MRR experiments on *G. m. centralis, G. m. submorsitans*, and *G. m. morsitans* with the foregoing model and estimated an overall mean dispersion rate of 253 ± 29.3 m daily. Further development of diffusion equations by Williams et al. (1992) led to an overall estimate a daily diffusion rates of 360 to 1.1 km varying by fly sex and season. Leak (1998, pp150-153) summarised reports on tsetse dispersion and Vale et al. (2014) further refined theory and summarised daily displacement rate of riverine flies (Palpalis group) at about 0.333 and that of savannah flies (Morsitans group) at 1 km.

6. Effective population numbers N_e

6.1. *Definitions*

The effective population number (size) of a population under study is the number of individuals in a theoretical "ideal" population undergoing the same rate of random genetic drift as the studied population (an ideal population is randomly mating, of finite, constant size over successive discrete (i.e., non-overlapping) generations, experiencing negligible selection and gene flow and has an equal sex ratio; see Charlesworth, 2009; Luikart et al., 2010). A much less rigourous definition is that N_e the average number of individuals in a population contributing genes to successive generations. N_e provides a method of calculating the rate of genetic change caused by random genetic drift (Charlesworth, 2009). Additionally, estimates of N_e can provide hypothetical estimates of population densities below which natural populations cannot recover from applied control measures. It must be clear, however, that N_e and census densities are much different measures. N_e tend to be much less than census densities. The measurement of N_e is problematic because it is affected by demographic and genetic phenomena including effective sex ratio, population age structure, inbreeding, temporal variation in population size, variance in reproductive success, gene flow, selection and breeding structure. Moreover, the value of N_e varies across the genome. Thus, estimates vary with the genetic loci employed. They vary also

with the analytical method used to estimate them, a subject beyond the scope of this review. Here we quote data based on temporal methods of estimation (Table 5).

6.2. *Estimation of* N_e

Estimates of harmonic mean effective pop sizes among tsetse varied per class of variation used and the underlying mathematical models employed. But in nearly all cases, reported estimates for tsetse of those few species studied varied from *c.* 20 to ~ twenty thousand flies (Table 5), but a modal value would fall into the low end of the distribution. Now gene flow is generally greater among the subgenus *Nemorhina* than among subgenus *Morsitans* taxa (i.e., their F_{ST} estimates are less, Tables 3, 4). Thus, a modal estimate for *Nemorhina* probably falls in the region of 50 to 2000 and that for subgenus *Morsitans* fall into 50 – 550. For example, an estimate of N_ef (effective number of females) = 400 - 2258 among 3 Kenya and Uganda samples of *G. f. fuscipes* (Krafsur et al. 2008); also in Uganda *G. f. fuscipes*, N_e = 33 - 310 when based on SSRs (Hyseni et al., 2013). Further estimates on microsatellite loci varied among seven *G. f. fuscipes* populations in Uganda 152 to ~20,000 (Echodu et al., 2011). Dyer et al. (2009) estimated N_{ef} = 501 and N_e = 731 in Equatorial Guineas *G. p. palpalis*. Using SSRs, Mélachio et al. (2011) estimated N_e = 20 - 300 in *G. p. palpalis* from Cameroon and the DRC. Solano et al. (2009) estimated N_e = 40 - 1000 among *G. p. gambiensis* in mainland Guinea and N_e =10 – 60 on islands. Harmonic mean estimate for *G. tachinoides* along a 338 km of the River Mouhoun in Burkina Faso where $F_{ST} \approx 0$ was almost 1200 (Koné et al., 2010).

Turning now to the Morsitans group (Table 5), sequential sampling of *G. swynnertoni* in Tarangire, Tanzania afforded an estimated $N_ef \approx 50$ (Marquez et al., 2006). Based on SSRs among *G. pallidipes*, Ouma et al. (2006) estimated N_e = 180 in Lambwe, Kenya, and 551 in Nguruman, Kenya. The foregoing populations were subjected earlier to intensive control procedures from which they had recovered. Also in *G. pallidipes*, a 670 bp fragment of mitochondrial loci *r16S2 + COI*, N_ef = 77 - 1435 among four Ethiopian populations, N_ef = 70 – 407 among ten Kenya samples, N_ef = 66 - 230 among five Tanzania samples, and N_ef = 37 - 67 among four southern African samples. The average N_ef = 369 over the 28 samples. Excluding the eight northernmost and southernmost populations, N_ef = 120. Mitochondrial diversities and effective population numbers decreased from north to south (Krafsur et al., 2016). Microsatellite loci in *G. pallidipes* showed a similar pattern but allozyme loci were equally diverse north to

south, an example of balancing selection (Krafsur, 2002). Consistent with the foregoing, lesser mitochondrial and microsatellite diversities and F_{ST}s were recorded in Southern African than in Tanzanian *G. m. morsitans*.

The foregoing N_e estimates for *Glossina* spp may be compared with African *Drosophila* N_e of 1,125,000, humans of 10,400 (Charlesworth 2009), and many thousands in each of *Anopheles funestus. An. Gambiae,* and *An. arabiensis* mosquitoes (Michel et al., 2006, Hodges et al., 2013). Among 46 *Aedes aegypti* samples from the New World, Africa, and Australia, N_e estimates based on SSRs and SNPs varied from c. 25 - 3,000, averaging 400 – 600 (Saarman et al., 2017). It may be noted that gene flow greatly dominates genetic drift among African *Drosophila* and *Anopheles*. Quite the reverse characterizes *Morsitans* populations studied, where breeding structures are localized and drift greatly dominates gene flow (Krafsur, 2009). Earlier views that particularly *G. m. morsitans* and *G. pallidipes* rapidly and widely disperse (e.g. Hargrove, 2003; Williams et al, 1992) seem to be falsified by the genetic data on Morsitans group and some Palpalis group tsetse populations and this should have important consequences on plans for area-wide vector control because invasion from undetected or identified nearby populations may prove not to be as certain, rapid, or heavy as once thought (e.g. Williams et al., 1992).

7. Tsetse population management and trypanosomiasis control

7.1. *Epidemiology*

In section 2.3 we reviewed some basic epidemiological matters related to HAT. Human African trypanosomiasis caused by *T. b. gambiense* (gHAT) is now targeted by WHO for elimination as a public health problem by 2020 (that is less than one case per 10,000 people per year) with complete elimination by 2030 (Buscher et al., 2018; Franco et al., 2017; WHO, 2014). WHO data show that prevalence of *T. b. gambiense* reached its most recent peak in 1997 (with over 34,000 cases) then incidence fell so that in 2014 there were only 3,797 reported cases and less than an estimated total of 15,000 cases (Franco et al., 2017). In the last 10 years, over 70% of reported cases occurred in the Democratic Republic of the Congo (WHO, 2016a, b). Sutherland et al. (2017) wrote that the WHO aim of elimination of gHAT by 2020 is feasible,

given the continued implementation of vector control. Thus it is assumed that gHAT is almost exclusively vector borne.

We have seen, however, that *T. b. gambiense* is essentially a clonal population, derived from a single progenitor trypanosome (i.e., most recent common ancestor) within the last 1,000-1,500 years before present (CI = 750 – 9500 y) that is evolving strictly asexually (Weir et al., 2016). Transmission of *T. b. gambiense* by tsetse has always presented an obstacle to laboratory studies of this parasite but the underlying assumption remains that gHAT is strictly a vector borne disease upon which WHO disease control policy is founded (Sutherland et al., 2017).

It has recently been suggested that, rather than being exclusively vector borne, *T. b. gambiense* is normally transmitted from mother to offspring during gestation (Welburn et al., 2016a for a thorough review). The very low *T. b. gambiense* salivary gland infection rates in tsetse, tending to zero, are consistent with the hypothesis that vector transmission is rare. We have seen that the incidence of sleeping sickness has declined across West and Central Africa since the 1990's with or without attention to vector control but rather depended on the screening and treatment of patients. On the other hand, longstanding gHAT foci disappeared in parts of West Africa after prolonged droughts that reduced tsetse populations. Buscher et al. (2018) pointed out that gHAT prevalence is underestimated to a greater or lesser extent. They argued that latent, cryptic infections could provide undetected *T. b. gambiense* reservoirs in man and animals thereby allowing for disease resurgence. It seems, therefore, that a complete elimination of gHAT by 2030 is rather optimistic.

7.2. *Choosing control methods*

Some general reviews regarding trypanosomiasis control include Aksoy et al. (2017), Diall et al. (2017), Auty et al. (2016), Torr and Vale (2015), Rock et al. (2015), Tirados et al. (2015), Dicko et al. (2014), Welburn and Maudlin (2012), Févre et al. (2008), and Hargrove (2003). Field-tested tsetse population management methods include targets, traps, insecticidal applications to domestic animals, aerial sprays and trypanocides (Torr and Vale, 2015; Hargrove et al., 2012; Welburn and Maudlin, 2012; Hargrove, 2003, 2004, and Hargrove et al., 2003) and the Sterile Insect Technique (SIT) (Vreysen et al., 2000). Numerous carefully conducted control trials using one or more such methods have recently been performed including a proposal to

eliminate HAT by *T. rhodesiense* (rHAT) in Uganda for which a Public Private Partnership ('Stamp Out Sleeping Sickness' - see http://www.stampoutsleepingsickness.com) was established in 2008 (Welburn et al., 2016b) and a trial performed to treat large numbers of cattle among 20 villages with a single dose of a trypanocide (Hamill et al., 2017; Fyfe et al., 2017). In African settings, fiscal considerations are usually decisive in selecting an appropriate technology. The Restricted Application Protocol (RAP) using insecticide treated cattle (Torr et al., 2007) has been shown to provide the most attractive option for livestock management by smallholder livestock keepers. Effective tsetse control does not require the application of RAP to all animals and this method has an important collateral impact on tick control (Muhanguzi et al., 2015).

The sterile insect technique against tsetse has been convincingly shown to be inefficient and too expensive (Hargrove, 2003; Vale and Torr, 2005; Rock et al., 2013. Vale et al., 20140). The large numbers of sterile tsetse necessary to achieve high rates of sterile matings during the Zanzibar field trial (Vreysen et al., 2000) indicated that competitiveness of sterilised released males for mates was poor. Moreover, the costs of sterile fly production, sexing, packaging, and distribution, together with the large number of trained technicians required to mount an SIT programme put costs above what most African countries and donors would be willing to support. Weekly sterile male releases must continue for at least 14 months because of tsetse longevity, long generation time, and slow reproduction. In contrast, effective deployment of targets and/or insecticidal applications can erase a natural target population in a single month. The finding that secondary and tertiary inseminations may occur (Bononi et al., 2011) does not seem to argue against the SIT because such secondary matings could be between initially fertile females and sterile males. Altogether lacking are data that relate experimental sterile fly dosage to sterile mating rates and target tsetse population density responses.

7.3. *Hypothetical methods*

Hypothetical methods include combining conventional methods with the SIT (e.g. Curtis and Adak, 1974; Bourtzis et al., 2016; Shaw et al., 2013; Muhanguzi et al., 2015) and an imaginative paratransgenic method (Aksoy, 2003, Aksoy et al., 2008, Alam et al., 2011; Medlock et al., 2013, Gilbert et al., 2016) in which symbionts are genetically engineered to produce trypanolytic substances in host tsetse (see *8.1 Paratransgenesis and Incompatible Insect Technique*). The Incompatible Insect Technique utilizes *Wolbachia* to cause embryonic lethality

via cytoplasmic incompatibility (CI) and happens when infected males inseminate uninfected females (unidirectional CI). Bi-directional CI is expressed when individuals with different incompatible strains of *Wolbachia* are crossed. The strength of CI expression varies among strains of *Wolbachia* and is proportional to the densities of infection (Doudoumis et al., 2013; Schneider et al., 2015). In principle, the intracellular, maternally transmitted reproductive parasite *Wolbachia* could be used, via CI, to replace a natural vector tsetse population with a trypanosome-resistant one. Cytoplasmic incompatibility was demonstrated in *G. m. morsitans* (Brelsfoard et al., 2014; Alam et al., 2011).

A brief examination of tsetse commensals and their chief properties was reviewed earlier principally from a coevolutionary point of view. A number of those properties can hypothetically be used to manage tsetse populations to interrupt transmission of pathogenic trypanosomes. Indeed, experiments have been carried out to develop a trypanocidal expression system in the secondary symbiont *Sodalis* (De Vooght et al., 2014). The use of tsetse symbionts and pathogens in area-wide *Glossina* pest management was reviewed (Gilbert et al., 2016; Abd-Alla et al., 2013). Here we examine genetically-based methods for control vis-à-vis the use of targets, traps, and insecticidal treatments.

Wolbachia-induced CI hypothetically can be used to 'drive' a desirable genotype into a natural population, thereby replacing it (Doudoumis et al., 2013). Indeed, there is compelling evidence for such a drive mechanism in nature that explains the replacement in California of one strain of *Drosophila simulans* by another. For specific applications, it is necessary to learn the distribution and genotypes of *Wolbachia* strains in a target tsetse species. *Wolbachia* distribution and prevalence is patchy within and among tsetse species (Cheng et al., 2000; Charlat et al., 2003; Alam et al., 2011; Doudoumis et al., 2012; Wamwiri et al., 2013). Diverse *Wolbachia* genetic variation and distribution has been reported among Uganda *G. f. fucipes* (Symula et al., 2013; Doudoumis et al., 2013). To use CI operationally, it is important to learn the specific reproductive effects of any particular *Wolbachia* strain on tsetse taxa, not a simple, inexpensive, or quick undertaking. But there is progress: Schneider et al. (2013) provided a new, more sensitive and accurate screening tool, VNTR-141 (Variable Number Tandem Repeat locus) for estimating *Wolbachia* prevalence and 'fingerprinting' strains.

Given optimum assumptions, the time must be estimated, in terms of number of generations, necessary to achieve an effective reduction in trypanosome infected tsetse. A number of simulations have been offered (Vale and Torr, 2005; Vale et al., 2015; Alam et al.,

2011; Medlock et al., 2013; Gilbert et al., 2016) and some show that at least three years of intensive, efficient effort must be made to achieve goals. Much shorter periods for effective control can be obtained when proven, conventional means are employed (e.g., Hamill et al., 2017; Fyfe et al., 2017; Vale et al., 2014; Vale and Torr, 2005).

7.4. *Consequences of Glossina population structure*

Is insecticide resistance likely to occur where and when area-wide IPM is employed? Maudlin et al. (1981) demonstrated a limited potential for resistance in *G. m. morsitans* by showing that only 0.7% of topically applied DDT was detoxified to DDE, an inconsequential conversion rate. But how likely is 'potential' insecticide resistance in *Glossina* to evolve into operationally significant resistance in nature, where insecticidal selection pressures can be very strong? That likelihood was addressed by Rock et al. (2015) and found to be negligible because tsetse are extraordinarily susceptible to insecticides. Moreover, their capacities to respond to such high selective pressures are small because of long generation intervals and low reproductive rates. Thus an extremely rare insecticide-resistant fly is most unlikely to found an economically significant population of resistant progeny. Moreover, as already discussed, tsetse effective population numbers N_e typically are small (of the order 10^1 to 10^4) compared, for example, with those of mosquitoes, *D. melanogaster* and, indeed, most other insect pest taxa. Mutation rates of an allele from susceptible to resistant approximate to $< 10^{-7}$. The product of two small numbers is much smaller. In unselected populations, resistant phenotypes usually are reproductively less fit than susceptible phenotypes. So the probabilities of resistant tsetse in nature are vanishingly small with little on which insecticidal selection pressures can act unless the mutant allele conferring resistance is dominant. Moreover, tsetse reproduction rates are small; a female must live at least 25 days to produce two progeny and generation times approximate to over 50 days. Because of their small effective population numbers and low mutation rates, tsetse populations are unlikely to survive the selection pressures that insecticidal applications bring.

8. Discussion and Prospectus

8.1. *Paratransgenesis and Incompatible Insect Technique*

The concept of paratransgenesis requires the creation of symbiotic bacteria genetically engineered to express trypanolytic substances (e.g. *Sodalis*, Aksoy et al., 2017) that may be 'driven' into natural tsetse populations via CI. Thus, at its simplest, a fertile, robust line of *Glossina* must be created that contains the engineered symbionts and carries also a strong CI-inducing strain of *Wolbachia*. The hypothetical advantage is, a natural population of vectors may be replaced with a population of entirely refractory forms, thereby eliminating risk of acquiring African trypanosomiasis. Likely problems in execution include: (1) the prevalences and infection densities of *Wolbachia* vary in and among populations; (2) different strains of *Wolbachia* have different effects on host reproductive fitness; (3) monitoring the progress of population replacement would require intensive, extensive, and careful monitoring; (4) evolving resistance of trypanosomes to trypanocidal substances; and (5) the long time-course necessary to achieve replacement due to tsetse generation time.

It is hard to see how estimates of effective population sizes are necessary in planning and execution of conventional vector control methods. Census population densities typically are very much greater than N_e and standing populations must be dealt with. In the estimation of N_e, there is no satisfactory treatment of populations in which individuals reproduce more or less continuously (Charlesworth, 2009). Because tsetse reproduce continuously, their generations overlap continuously and compromise estimates of N_e. This problem applies also to N_e estimates for *Drosophila* and *Anopheles* mosquitoes even though their N_e estimates are very much greater than those of *Glossina* spp. Estimates of *Glossina* N_e may be off by two or three orders of magnitude if their mutation rates are like that estimated in African *Drosophila* (i.e., 2.8×10^{-9} per site per generation based on whole genome sequencing; Keightley et al., 2014). Mutation rate estimates for *Glossina* were based on only two mitochondrial loci, so they may not be comparable. Nevertheless, *Glossina* densities plainly are very much less than the taxa to which we have compared them, gene flow among their populations is not great, their reproductive rates are comparatively much less, so it seems small effective population sizes are likely the rule and not exceptional.

Small populations are affected more by drift than are large populations and tend to possess lesser genetic diversities than larger populations (Kimura et al., 1963). There is a further consideration: populations derived from colonizing episodes or recovered from bottlenecks hypothetically show much smaller N_e than their ancestral populations. Many extant tsetse populations may be derived from a few survivors or colonizers from surviving populations,

particularly among Morsitans group taxa. It seems the great rinderpest epizootic of the late 19th and early twentieth centuries eliminated many tsetse populations and greatly reduced densities in others (by killing off their sources of blood meals); moreover, the effect of the rinderpest on tsetse densities seemed to have been greater in southern than in northern Africa (Krafsur et al. 2016. Published F_{ST} estimates suggest that dispersing Palpalis group flies, (e.g., riverine taxa *G. p. palpalis*, *G. p. gambiense*) along watercourses show higher rates of gene flow (i.e., smaller F_{ST} estimates) than do savannah forms (Morsitans group taxa, *G. morsitans s.l., G. pallidipes, G. swynnertoni*). Practical and theoretical evidence, however, provide a contrary view of dispersal rates (Vale et al., 2014).

Little genetic work has been accomplished with *Fusca* group tsetse, with the exception of descriptions of the mitotic karyotype of *G. brevipalpis* and the assignment of 28S ribosomal DNA loci to physical chromosome L_1 (Maudlin, 1970; Willhoeft, 1997). There are 14 species in the group, most of which are confined to forests (Leak, 1998; Robinson & Rogers, 2004). As today's principal trypanosome vectors are brought under control, it may be anticipated the other, now secondary and tertiary vectors may become increasingly important.. Their rates of doing so, however, are likely to be slow and discontinuous. Low effective population numbers insure slow evolutionary responses (Charlesworth, 2009). It may also be anticipated that, in the long term, adaptation to new vectors may occur via mutation in *T. brucei* s.l. and genetic recombination in *T. b. brucei* and *T. b. rhodesiense*.

8.2. *Opportunities for further genetic work*

8.2.1. *Glossina spp*

Regional, replicate cultures of HAT vectors need to be established to allow variances to be estimated in tsetse responses to experimental treatments. Earlier work in finding differences between two conspecific cultures of *G. pallidipes* (e.g. Langley et al., 1984; Southern and Pell, 1981) and *G. m. morsitans* (Jordan et al., 1977) cannot reliably be confirmed because there was no replication therefore no estimate of experimental error for treatment effects. A further advantage of establishing replicate, regional tsetse cultures is that reciprocal crossing of such lines would provide an efficient way of finding cryptic species via hybrid fertilities.

Colonization of *G. longipalpis* would allow explicit genetic comparisons to be made with its eastern relative, *G. pallidipes*. *G. pallidipes* occurs in Mozambique and Somalia but has not been sampled there for genetic work.

Further sampling of genomic and cytological diversity among *Glossina* taxa would help to explain the polychotomies observed in their phylogeny (e.g. *G. fuscipes* s.l., see Fig. 1). Indeed, beyond the exceptional work of Dyer et al. (2008, 2009, 2011) the further work needed on tsetse phylogenies recommended by Gooding and Krafsur (2004) has not yet been undertaken as workers rush to define and justify the establishment of area-wide control programmes.

The restricted gene flow among Morsitans group tsetse could be tested further by mark, release, and recapture trials. Rogers (1977) seminal work on *G. f. fuscipes* during 1971 in Uganda and *G. m. morsitans* and *G. pallidipes* at Rekomitjie field station in Zimbabwe (Vale et al., 1984), require replication in new environments. Although such work is expensive and tedious, it is necessary to confirm and help to rationalize genetic inferences on dispersion and target population isolation.

The rate and accuracy of detection of *Wolbachia pipientis* has recently improved. The parasite has been sequenced (Wu et al., 2004) which opens up opportunities to study *Wolbachia* diversities vis-à-vis that of its hosts. Further surveys of *Wolbachia* in natural and colonized *Glossina* populations are indicated because of the role CI might have had in *Glossina* speciation and race formation.

8.2.2. *Trypanosoma*

The availability of genomic markers and sequenced genomes will allow further geographic hic and replicate sampling of trypanosomes infecting man, animals, and tsetse. Temporal changes in trypanosome gene frequencies can be evaluated (after considering the sampling bias in earlier parasite isolations already discussed) in an effort to detect temporal evolutionary changes. Evaluation of fly genomes vis-à-vis trypanosome genomes in the long-established, stable HAT foci might lead to testable hypotheses regarding parasite-vector co-adaptations.

8.2.3. *Epidemiology of Gambian HAT*.

Correlation of human genotypes vis-à-vis gHAT resistance and susceptibility, (see 2.4. *Distribution of HAT and Tsetse*), could provide important insight into their coevolution and epidemiology. Welburn et al. (2016) reviewed and discussed the importance of human genetics with respect to gHAT and rHAT.

FIGURES

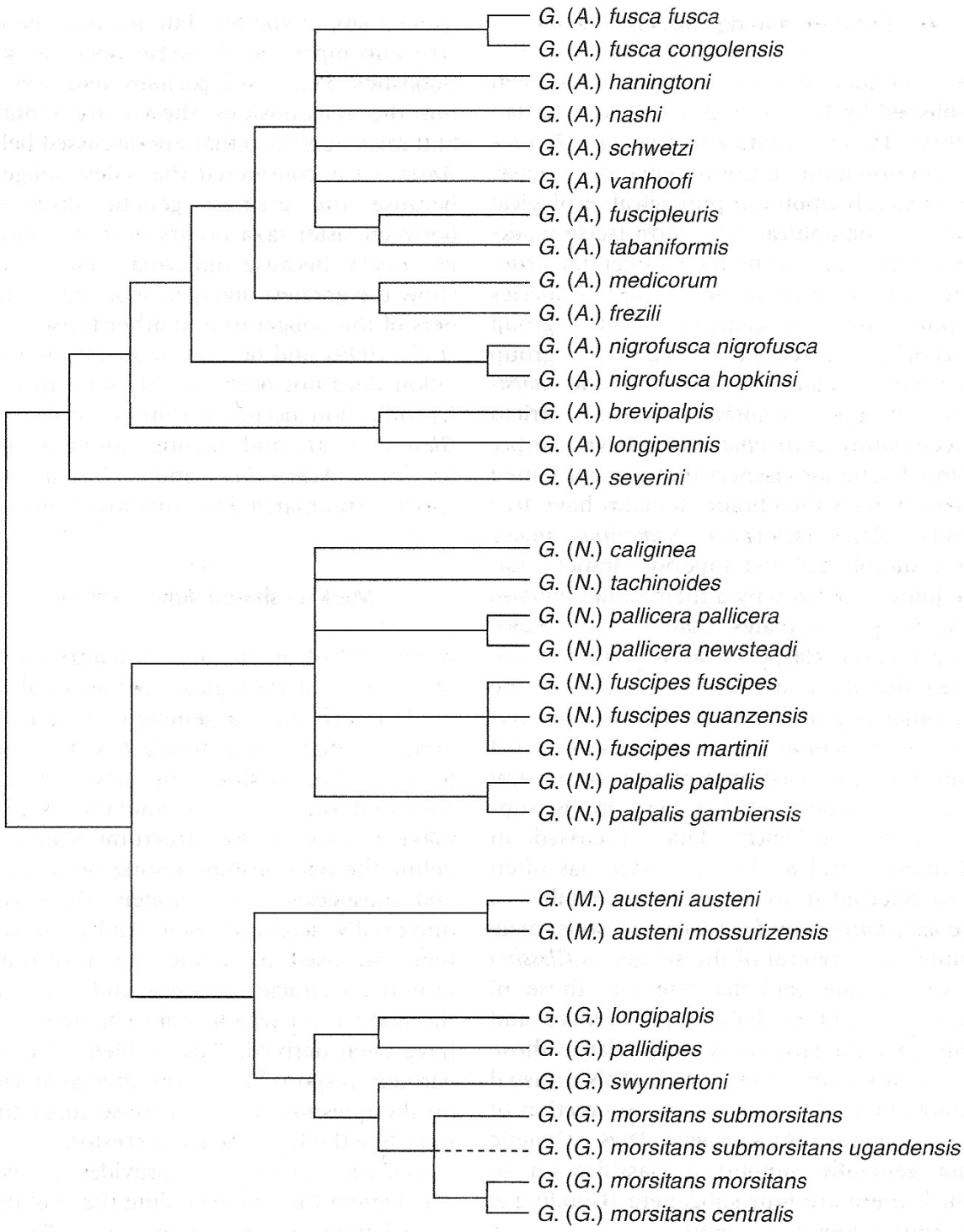

Fig. 1. Hypothetical phylogeny of 31 *Glossina* spp, after Gooding and Krafsur, 2004.

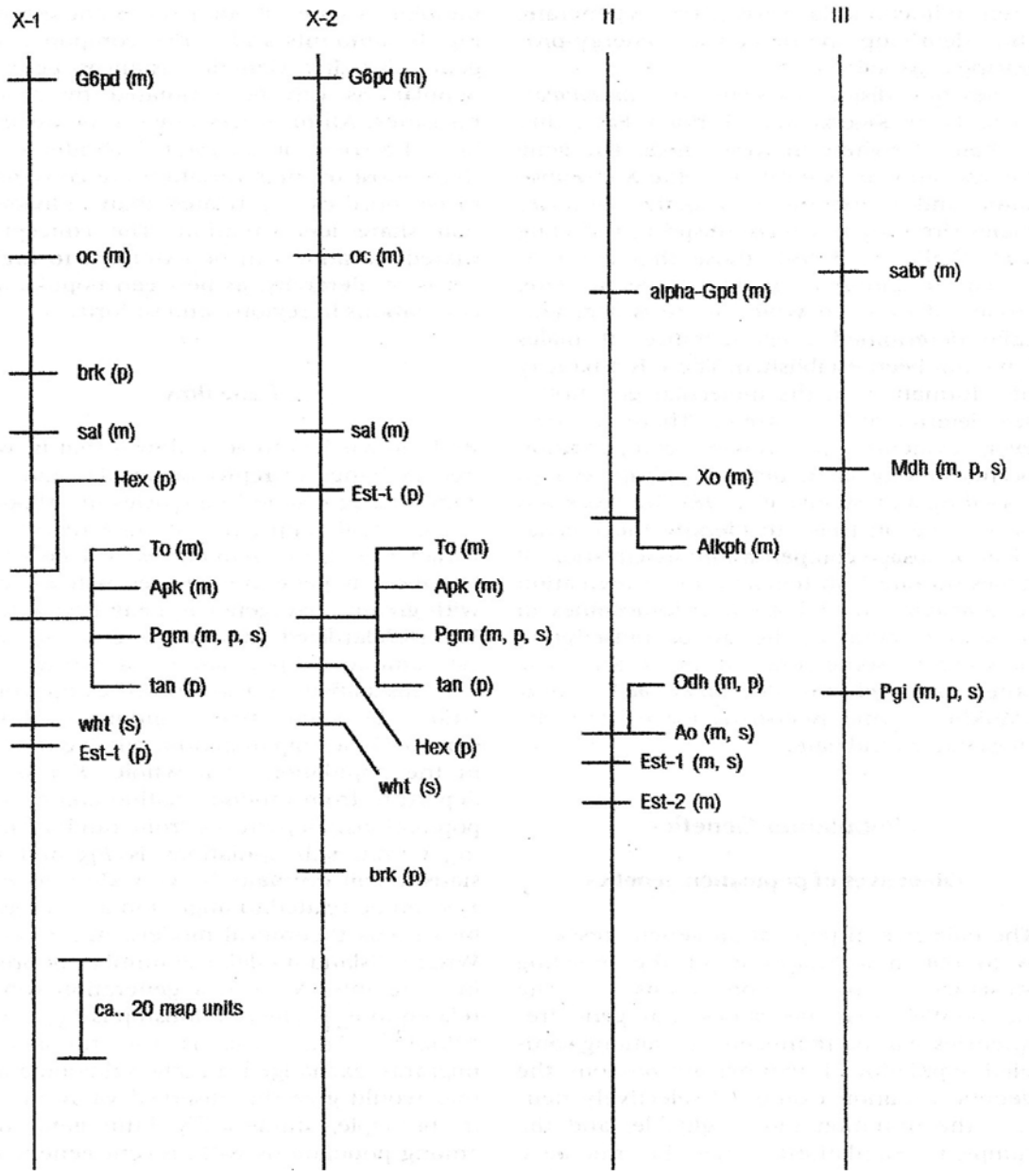

Fig. 2. Diagram is from Gooding and Krafsur, 2004. Composite linkage map for *morsitans* and *palpalis* group tsetse showing morphological and allozyme loci on the X chromosome and two autosomes. Two X chromosomes, originally from two *G. m. submorsitans* colonies, are figured because they similar to that of *D. melanogaster* and the Y chromosome has no detectable influence on the male phenotype). Instead, sex determination depends on the number of X chromosomes (Maudlin, 1979). X-1 and X-2 probably differ by an inverted segment. Taxonomic abbreviations are (m) *G. m. morsitans*, (s); *G. m. submorsitans*, and (p) *G. p. palpalis*.

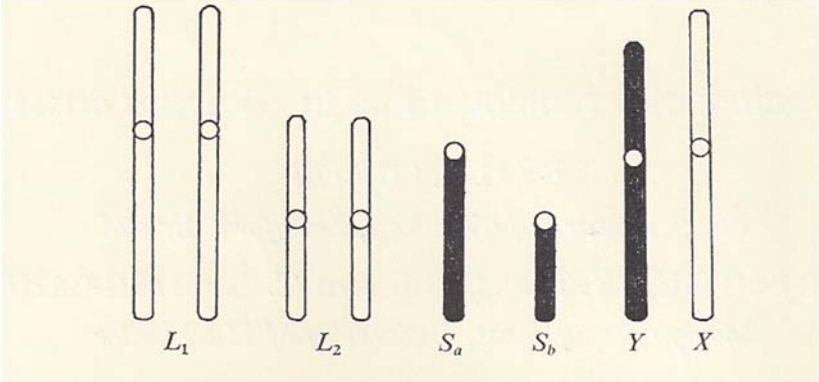

Figure 3. Diagrammatic *G. austeni* male karyotype, after Curtis et al., 1972. 'L' are metacentric or submetacentric autosomes, 'S' are acrocentric supernumeries, 'Y' and 'X' are heterosomes (sex chromosomes). The S and Y are heterochromatic.