THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0892

The population genetics of *Trypanosoma brucei* and the origin of human infectivity

Annette MacLeod[1], **Andy Tait**[1] **and C. Michael R. Turner**[2*]

[1]*Wellcome Centre of Molecular Parasitology, The Anderson College, University of Glasgow, 56 Dumbarton Road, Glasgow G11 6NU, UK*
[2]*Division of Infection and Immunity, Institute of Biomedical and Life Sciences, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, UK*

The African trypanosome, *Trypanosoma brucei*, is a zoonotic parasite transmitted by tsetse flies. Two of the three subspecies, *T. brucei gambiense* and *T. b. rhodesiense*, cause sleeping sickness in humans whereas the third subspecies, *T. b. brucei*, is not infective to humans. We propose that the key to understanding genetic relationships within this species is the analysis of gene flow to determine the importance of genetic exchange within populations and the relatedness of populations. *T. brucei* parasites undergo genetic exchange when present in infections of mixed genotypes in tsetse flies in the laboratory, although this is not an obligatory process. Infections of mixed genotype are surprisingly common in field isolates from tsetse flies such that there is opportunity for genetic exchange to occur. Population genetic analyses, taking into account geographical and host species of origin, show that genetic exchange occurs sufficiently frequently in the field to be an important determinant of genetic diversity, except where particular clones have acquired the ability to infect humans. Thus, *T. brucei* populations have an 'epidemic' genetic structure, but the better-characterized human-infective populations have a 'clonal' structure. Remarkably, the ability to infect humans appears to have arisen on multiple occasions in different geographical locations in sub-Saharan Africa. Our data indicate that the classical subspecies terminology for *T. brucei* is genetically inappropriate. It is an implicit assumption in most infectious disease biology that when a zoonotic pathogen acquires the capability to infect humans, it does so once and then spreads through the human population from that single-source event. For at least one major pathogen in tropical medicine, *T. brucei*, this assumption is invalid.

Keywords: *Trypanosoma brucei*; sleeping sickness; population genetics; human infectivity; polymorphism

1. INTRODUCTION

Many protozoan parasites are zoonotic (e.g. *Trypanosoma brucei*, *Cryptosporidium parvum*, *Toxoplasma gondii* and *Trypanosoma cruzi*) and transmitted between human and other mammalian hosts by a variety of routes. These parasites are the cause of significant mortality and morbidity and present a challenge in terms of optimizing control strategies given the involvement of multiple hosts, as well as raising the question of how the major sources of human infection can be identified and their significance evaluated. With the increasing availability of genome sequences, a wide range of molecular markers can be readily developed to address some of the outstanding questions and provide the means to track the sources of outbreaks of human infection. Key considerations are the population structure of the parasite under investigation and the role of genetic exchange in generating variation. These issues underpin the utility of any system for defining human-infective parasites and have major implications for the application of chemotherapeutic- or vaccine-based control strategies. With most of the protozoan parasites there is indirect evidence for different genotypes of parasite having different host-specific cycles and in some of the species these have been accorded subspecific status, whereas in others they are referred to as 'types'. The existence of host-specific types significantly complicates the investigation of the population structure, as well as raising a series of questions about the origin and nature of human infectivity. It is often assumed that human infectivity arose once and then spread through the human population, but it is important to keep an open mind on this issue and examine such assumptions carefully and critically.

African trypanosomes comprise a number of morphologically distinguishable genera and species that infect a wide range of different hosts; most are transmitted by biting insects. In terms of human and veterinary medicine the important tsetse-transmitted species are *T. brucei*, *T. congolense* and *T. vivax*, with the latter two species solely infecting domestic animals and wild game. *T. brucei* comprises three classically defined but morphologically identical subspecies (Hoare 1972). The three subspecies were originally defined by their geographical distribution, host distribution, clinical course of disease and virulence of infections in rodents. Thus, *T. b. gambiense* is defined as a human parasite distributed through western and Central Africa, causing chronic disease, while *T. b. rhodesiense* is also a human parasite but distributed throughout eastern and southern Africa,

*Author for correspondence (m.turner@bio.gla.ac.uk).

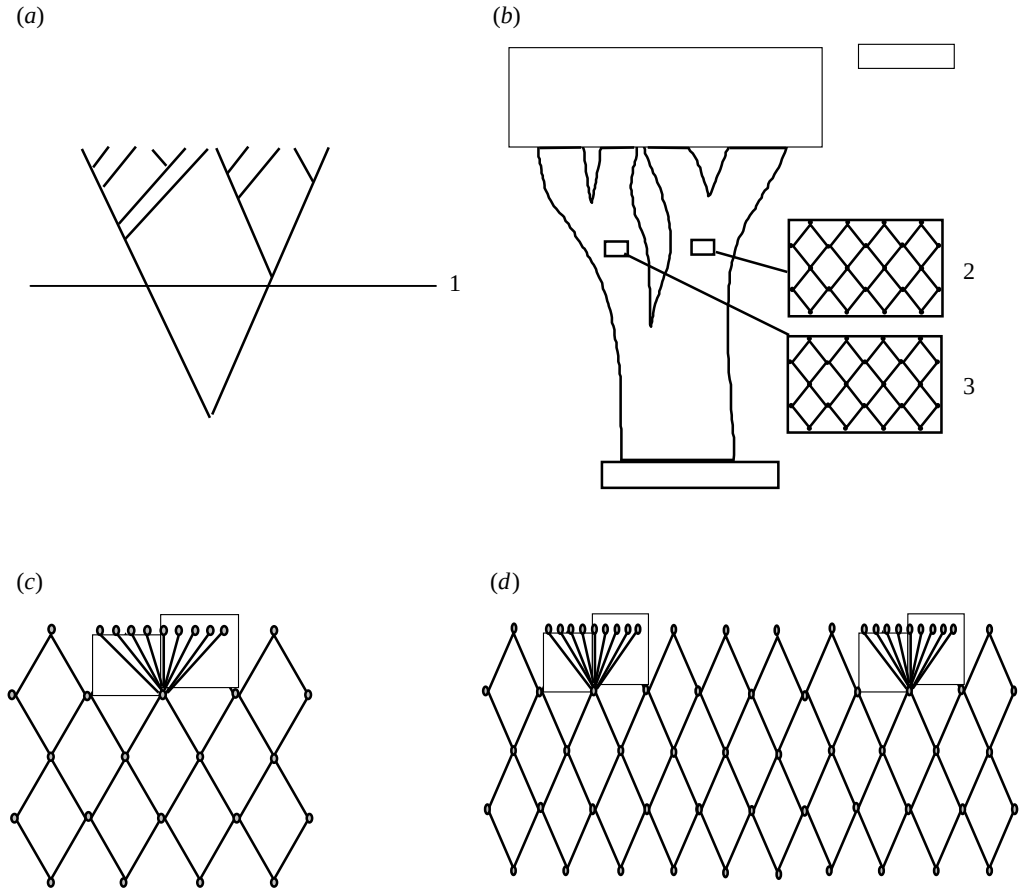

Figure 1. Diagrammatic representations of different population genetic structures (adapted with permission from Maynard Smith *et al*. (1993)). (*a*) *Clonality*. Clonality can be defined as a population in which genetic exchange is limited to such an extent that it is not sufficient to break up associations between alleles at different loci resulting in strong linkage disequilibrium. A dendrogram of genotype similarity would resemble an evolutionary tree, with no recombination occurring between isolates in the same or different branches of the tree. A clonal population does not imply that all isolates from this population are identical. If, for example, samples were taken at time-point 1 two clonal genotypes would have been identified. (*b*) *Panmixia*. In a panmictic population sexual recombination occurs frequently between individuals within the population. It must be remembered, however, that a species can consist of a number of populations, which can be subdivided into subpopulations due to a variety of different factors such as host-specificity and geographical isolation. Within each subpopulation, sexual recombination could occur frequently between isolates from within the same lineage as shown, but rarely, if at all, between different subpopulations. The analysis of polymorphic markers in such subpopulations would reveal a population in agreement with Hardy–Weinberg predictions and no significant linkage disequilibrium. However, if samples from two discrete populations were examined, i.e. if isolates from different subpopulations are combined, for example populations 2 and 3, deviation from Hardy–Weinberg predictions and linkage disequilibrium would be observed. (*c*) *Epidemic*. An epidemic population structure results from a combination of panmictic and clonal effects. The population has a basic panmictic population structure, but occasionally one or more genotypes, which are particularly suited to the environmental conditions, expand clonally to dominate the population. Superficial analysis of such populations would give the appearance of a clonal population structure (i.e. strong linkage disequilibrium and deviation from Hardy–Weinberg predictions) as the expansion of one or two genotypes would obscure underlying frequent genetic exchange in the population. However, reanalysis of the same data once the frequency of the common genotype(s) has been normalized would reveal the underlying nature of the population (i.e. no significant linkage disequilibrium and agreement with Hardy–Weinberg predictions). (*d*) *Different epidemic foci*. It is likely that different subpopulations, for example from different geographically isolated regions, have different clonally expanded genotypes or that over time different genotypes can come to dominate the population.

causing acute disease; the latter is virulent in rodent infections while the former is avirulent. The principal reservoir host for *T. b. gambiense* appears to be the domestic pig and for *T. b. rhodesiense* a large range of domestic and wild animals are potential reservoir hosts. *T. b. brucei* is not human-infective, is distributed throughout sub-Saharan Africa and infects domestic animals and game. These observations have led to an extensive set of investigations into the population genetics of these parasites aimed at determining the relationships between the different classically defined subspecies and the population structure in the different foci of disease. These analyses present an excellent illustration of some of the key issues and questions in parasite population genetics that are common to all the zoonotic protozoan parasites.

Table 1. A comparison of some of the more widely used genetic marker typing systems indicating their relative strengths and weaknesses when used for different purposes.

	typing systems				
purpose	isoenzymes	RFLPs	RAPDs	microsatellites	minisatellites
strain identification	+	+	+++	+++	++++
relatedness	++	++	++	+++	++++
identifying mixed infection	+	+	–	++	++++
population analysis	+++	–	–	+++	++
interspecies analysis	+++	+	–	++	+

2. *T. BRUCEI* POPULATION STRUCTURES

Genetic exchange can occur when *T. brucei* stocks are co-transmitted through tsetse flies in the laboratory, but this is not an obligatory process and both cross- and self-fertilization occur (Gibson & Stevens 1999; Tait & Turner 1990). The extent to which genetic exchange occurs in natural populations, however, is highly controversial. Three types of population structures have been proposed for *T. brucei* based on the extent of sexual recombination in the population: clonal (little genetic exchange), epidemic (some genetic exchange masked by clonal expansion of some strains) or panmictic (randomly mating). Definitions of each of these population structures are given in figure 1.

Tibayrenc *et al*. (1990) proposed the theory of clonality (diagrammatically represented in figure 1*a*) for a number of parasitic protozoa, providing evidence for strong linkage disequilibrium based on isoenzyme data. Evidence for a clonal population structure for *T. cruzi* is very convincing with a few different multilocus genotypes being repeatedly sampled from a number of different countries (Tibayrenc *et al*. 1986). However, the clonality theory may not necessarily be appropriate for other parasitic protozoa. There are several reasons why linkage disequilibrium could be detected other than clonality. Linkage disequilibrium can occur if the population has recently been through a population bottleneck or when isolates from a mixture of different subpopulations (each of which may be randomly mating) are examined and treated as a single population (figure 1*b*). Self-fertilization could also be the cause of linkage disequilibrium as proposed for the malaria parasite *Plasmodium falciparum*, which has an obligatory sexual cycle in the mosquito vector (Walliker *et al*. 1987).

Since it was proposed that *T. brucei* is clonal, a considerable number of studies have been undertaken which have shown that there are departures from Hardy–Weinberg equilibrium and significant linkage disequilibrium when alleles at two loci are considered (e.g. Tibayrenc *et al*. 1993; Mathieu-Daude & Tibayrenc 1994). This is in contrast to the original proposal by Tait (1980) that the population structure was panmictic (figure 1*b*), although given the small sample size it is difficult to exclude agreement with Hardy–Weinberg as having occurred due to chance (Cibulskis 1988). However, it should be remembered that to explain some of the genotypic variation observed using a purely asexual model of reproduction would require postulating multiple identical mutations in different isolates, which is very improbable. These two positions are, to a large degree, resolved by considering an 'epidemic' population structure as proposed by Maynard Smith *et al*. (1993) and illustrated diagrammatically in figure 1*c*,*b*. In this model one or two successful multilocus genotypes expand over a short time-span and mask the underlying occurrence of frequent genetic exchange (Maynard Smith *et al*. 1993). This model was originally developed for bacterial populations and also tested on a *T. brucei* population from Kenya in the original paper. It was subsequently applied to isoenzyme data from a Ugandan population (Hide *et al*. 1994). In both cases the data showed agreement with expectation, i.e. the populations had an epidemic population structure. It is perhaps important to note that the genetic structure proposed by Maynard Smith *et al*. (1993) is that which could be expected of a pathogen during an epidemic, but it is not necessarily *a priori* restricted to that ecological scenario and the genetic structure could also potentially be stable in time and/or space, i.e. the genetic definition of 'epidemic' is different from the ecological definition.

The key questions in trypanosome population genetics are with regard to gene flow within and between populations. Do trypanosomes of different stocks encounter each other as mixed infection in flies sufficiently frequently for sexual recombination to occur? Are trypanosomes in different geographical foci genetically isolated or is there gene flow between these populations? All *T. brucei* isolates infect cattle and wild game, but only some can infect humans—does this restrict gene flow and lead to substructuring of a population by host species? Is there evidence of a single evolutionary event by which *T. brucei* acquired the ability to infect humans?

To discuss these issues we first briefly deal with the different marker systems that have been employed and the important issue of mixed infections. We then review population studies relating to recombination and gene flow in eastern and western Africa separately. Finally we address the issue of the origins of human infectivity.

3. MARKER ANALYSIS

Before considering the evidence supporting each of the proposed population structures, it is important to be aware of the limitations of the various marker systems employed. It is self-evident that the quality of any study in population genetics is limited by the method of sampling and the attributes of the genetic markers. There

is no such thing as a 'perfect' marker system; each methodology has its strengths and weaknesses. As we have illustrated in table 1, each perceived strength and weakness is dependent on the purpose of the study and the availability of materials. Thus, for example, DNA fingerprinting techniques are excellent for determining identity of individuals because of the substantial numbers of loci sampled, but permit limited genetic interpretation of the alleles and loci involved. In contrast, single-locus restriction fragment length polymorphism (RFLP) analysis allows the determination of genotypes, but is laborious if many loci need to be sampled. Neither of these methods permits identification of mixtures of genotypes with any certainty, whereas microsatellite analysis can if the markers are sufficiently polymorphic.

It is not appropriate in this paper to review the various typing methods available other than to note the general trends particular to the study of *T. brucei*. Isoenzyme electrophoresis used in the earlier studies (Tait 1980; Gibson *et al*. 1980; Godfrey *et al*. 1990) gave way to studies using DNA fingerprinting (Hide *et al*. 1990, 1994), RFLP (Paindavoine *et al*. 1989) and random amplified polymorphic DNA (RAPD) (Stevens & Tibayrenc 1995) analysis, if only because these approaches required smaller amounts of material and gave direct access to genotypic information rather than requiring inference from the phenotype. More recently, polymerase chain reaction (PCR)-based techniques such as micro- and minisatellite analysis have come to the fore because they readily permit allelic interpretation and they are technically simple to apply to very small samples (Biteau *et al*. 2000; MacLeod *et al*. 1999). (They do not even necessarily require DNA extraction!) It is clear that with the recent development of the *T. brucei* genome project and new informatics techniques that it is easy to identify DNA repeat sequences, and so it is techniques such as these that hold the future for marker analysis.

A particular difficulty with trypanosome genetics is that field-sampling has rarely been conducted for the explicit purpose of population genetic analysis; most studies were designed with specific epidemiological or clinical hypotheses in mind. Nevertheless, substantial numbers of isolates do exist in cryopreservation banks. When using such collections the basic biology and ecology of *T. brucei* needs to be borne in mind. This is a tsetse-transmitted zoonotic parasite where some strains infect humans and others do not. Thus, for example, to take samples only of human origin will potentially bias any dataset in terms of elucidating gene flow within *T. brucei* populations as a whole, whereas taking samples from tsetse flies will minimize any bias. Also, human-infective *T. brucei* in eastern and southern Africa in particular exists in discrete geographical foci (Hide 1999) and to combine data for isolates from different foci can lead to incorrect conclusions. Furthermore, tsetse flies have particular feeding preferences for different species of mammals (Leak 1999) and so groups of isolates from different host species will potentially access only subfractions of the *T. brucei* population (Mihok *et al*. 1990).

A number of issues surround the handling of isolates before marker analysis is undertaken and these can significantly affect the information content and quality of the data obtained. Marker analysis using isoenzymes, RAPDs or RFLPs requires relatively large quantities of purified parasite material that can only be obtained by amplification of the original isolate in rodents or by culture. It is not certain whether this leads to selection and a loss of genotypic diversity, although it is interesting to note that there are reports of different typing results from the same isolate when more than one independent amplified population is examined (Godfrey *et al*. 1990). If a high proportion of isolates are comprised of a mixture of different genotypes, it is impossible to determine genotype frequencies in a large portion of the sample, thus confounding population genetic analysis. This can be circumvented by cloning each isolate but, as anyone who has undertaken this laborious, time-consuming and costly task will recognize, there are considerable advantages in avoiding this step if possible. Avoiding cloning speeds the analysis of larger numbers of isolates and sample size is important—inadequate sample sizes has bedevilled several studies (Cibulskis 1988, 1992). The counter-argument in favour of cloning is that in its absence there is a possibility of mixed genotypes in any sample, the presence of which could confound interpretation. For example, DNA fingerprinting will mistakenly identify a mixed infection as a new genotype, rather than as the combination of two genotypes.

4. MIXED INFECTIONS

The occurrence of mixtures of different parasite genotypes within one host is an important consideration in *T. brucei* population genetics. If mixtures occur rarely, it seems unlikely that parasites of different genotype will be co-transmitted by tsetse flies and so the opportunities for genetic exchange will be limited. Only three isoenzyme studies have reported the occurrence of mixed genotype infections, which, in part, reflects the lack of sensitivity of the technique and also the lack of genetic interpretation of the enzyme-banding patterns. Cloned lines were established from tsetse isolates from western Africa (Côte d'Ivoire) and typed (Stevens & Tibayrenc 1995), resulting in the identification of two, five and nine different genotypes in each of the isolates, respectively. In a separate study (Letch 1984), two distinct cloned lines from a single pig isolate were reported, also from Côte d'Ivoire. Using the criteria of either cloned lines differing in zymodeme (enzyme profile) pattern from the original isolate or different zymodeme patterns being obtained on separate preparations from the same original isolate (Godfrey *et al*. 1990), 26 isolates from Zambia were shown to consist of mixtures and were derived from human, tsetse or wild game. While it is difficult from the data presented to determine the exact number of isolates screened, this suggests that mixtures occur at a significant frequency. One of the inherent problems in using isoenzyme electrophoresis as a means of detecting mixtures in the absence of cloning is their relatively low polymorphic information content and the difficulty of genetically interpreting some of the patterns. These problems are circumvented by the use of highly polymorphic minisatellite markers.

A recent study of genotype composition in uncloned *T. brucei* isolates from the salivary glands of tsetse flies used three minisatellite markers and revealed that 36% of samples from Uganda and 47% from Kenya were of

mixed genotype. These results indicate that a significant proportion of tsetse flies harbours more than one genotype and so this important prerequisite for genetic exchange is met frequently (MacLeod *et al*. 1999). The high rate of mixed *T. brucei* infections indicates the possibility of frequent genetic exchange in the field. One interesting question which has emerged from these results is: how do tsetse flies acquire multiple genotypes of trypanosomes? The prevalence of *T. brucei* infections in tsetse flies is *ca*. 0.1–1% (Welburn & Maudlin 1997) and so if the ability of a tsetse fly to acquire an infection is independent of it acquiring a second infection, the predicted frequency of mixed infections is extremely low (0.01–0.0001%). However, tsetse flies are more likely to be infected by trypanosomes during their first feed with the incidence being reduced thereafter (Welburn & Maudlin 1992). The results of the minisatellite analysis imply that the tsetse flies fed on mixed-infected hosts, but the prevalence of mixed infections in mammals is reported to be low at 2–3% (Godfrey *et al*. 1990). In the light of this information, we revisited the data of Godfrey *et al*. (1990) because laboratory crosses have defined the genetic basis for the variation in a number of the isoenzymes used (Tait & Turner 1990; Gibson & Stevens 1999) and it is now possible to reinterpret some of the enzyme-banding patterns. Five of the enzymes used in this analysis were dimeric and their inheritance has been determined so that homozygous and heterozygous patterns can be recognized as well as those that must represent mixtures. In this reanalysis a total of 36 zymodemes can be identified as resulting from mixed infections of a range of hosts, including humans, in isolates from Zambia, Kenya, Ethiopia, Côte d'Ivoire and Tanzania. Thus, such mixtures appear to be quite widespread in terms of both host and geography. The frequency of these is difficult to determine with accuracy as the data are presented in terms of the number of populations rather than the number of isolates and only a proportion of the zymodeme patterns is informative with respect to the identification of mixtures. However, out of a total of 94 zymodemes in Zambia, 22 are clearly the product of mixed infections while out of 54 zymodemes from Côte d'Ivoire 9 are indicative of mixtures. This analysis suggests that the original estimates of prevalences of mixed infections are underestimates.

5. EAST AFRICAN POPULATIONS

Large collections of isolates from different regions of eastern Africa (Lambwe Valley, Kenya; Busoga Region, Uganda; Luangwe Valley, Zambia and others) have been analysed to determine whether the frequencies of the different genotypes observed at each locus conform to those predicted for randomly mating (panmictic) populations (for recent reviews see Gibson & Stevens 1999; Tait 2000). The isolates analysed in these studies were derived from a number of different hosts and include both *T. b. brucei* and *T. b. rhodesiense*. If these classically defined subspecies are genetically isolated, the population studies could be treating two distinct populations as one and so account for the observed linkage disequilibrium. In the absence of independent criteria, isolates from tsetse flies or animals could be either *T. b. brucei* or *T. b. rhodesiense*.

Since the initial proposal by Tibayrenc *et al*. (1990) that *T. brucei* was clonal, a more extensive study of 881 isolates was undertaken (Mathieu-Daude & Tibayrenc 1994) with the isolates subgrouped by host (human, domestic animal, wild game and tsetse). This study made comparisons between the frequency of the most common genotype and that predicted on the basis of panmixia, measured the probability of detecting the number of multilocus genotypes observed and estimated the level of linkage disequilibrium. Using these measures all the groups of isolates showed significant deviation from the expectations based on panmixia except for the group of isolates from wild game where the most common genotype was not overrepresented. These findings suggest that recombination is not occurring at a significant level and so does not disrupt the multilocus genotypes that appear to be stable in time and place, although there are a number of questions about the substructuring of populations by host and geography that need to be resolved before these conclusions can be fully accepted. As a result of the possibility that the overrepresentation of a few multilocus genotypes results from an epidemic population structure, an extensive reanalysis of published isoenzyme data has been undertaken with populations from Zambia, Uganda and Kenya (Stevens & Tibayrenc 1996). This analysis showed that the Ugandan and Kenyan populations have an epidemic structure while those from Zambia have a clonal structure. As these populations were not subdivided into the two subspecies of trypanosomes before analysis there is the formal possibility that these conclusions result from the treatment of two genetically isolated subspecies as a single species (see figure 1*b*). To avoid this possibility, Hide *et al*. (1994) subdivided a population of *T. brucei* from Uganda into the two subspecies using the ability of *T. b. rhodesiense* to resist the lytic effects of human serum as a means of subspecies definition. Analysis of the two subspecies separately demonstrated, on the basis of the value of the index of association obtained, that each subspecies had an epidemic population structure. However, given the low level of variation in the *T. b. rhodesiense* isolates and therefore the small number of different electrophoretic types available, type II errors (observation of apparent differences that are not real) cannot be excluded in relation to the conclusion reached for this subspecies.

MacLeod *et al*. (2000, 2001*a*) investigated the role of genetic exchange in *T. brucei* populations using highly polymorphic minisatellite markers. They avoided the inherent problem of inadvertently examining a geographically substructured population by analysing a well-defined population from a discrete geographical location (Busoga, Uganda). The parasites they examined were primarily isolated from humans and cattle with the latter isolates screened for resistance to lysis by human serum. In this way *T. b. rhodesiense* and *T. b. brucei* subspecies were defined and so were treated separately, avoiding substructuring due to host-specificity. The minisatellite markers showed high levels of polymorphism and from the data obtained it was concluded that *T. b. rhodesiense* is genetically isolated from *T. b. brucei* and can be unambiguously identified by its multilocus genotype. Analysis of the genotype frequencies in the separated *T. b. brucei* and *T. b. rhodesiense* populations

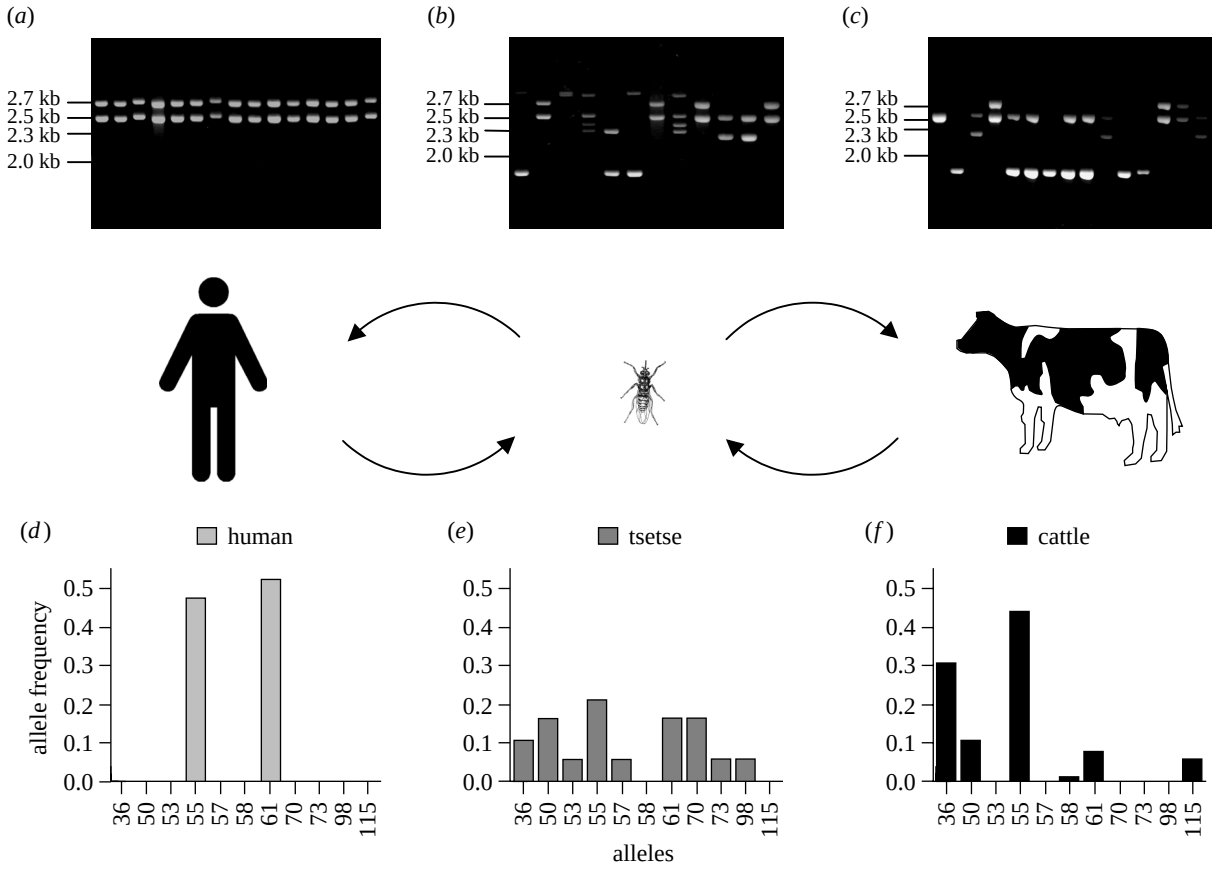

Figure 2. A composite diagram illustrating population genetic structures as observable in different host species. (*a–c*) PCR-amplified products of the *CRAM* minisatellite locus for a collection of isolates from humans, tsetse flies and cattle, respectively. (*d–f*) The equivalent graphs of allele frequencies. (*a*, *d*) These show size polymorphism of the two alleles and that all human isolates are genetically identical at this locus. (*b*, *e*) These show mixed infections to be present (more than two bands per track) and considerable heterogeneity amongst this group of isolates. (*c*, *f*) These show intermediate levels of heterogeneity in cattle.

showed that the former has an epidemic population structure while the latter is clonal; i.e. the *T. b. rhodesiense* population in Uganda consisted of only two multilocus genotypes, which were highly related. Moreover, these *T. b. rhodesiense* genotypes were stable, being observed over a 30-year time-span and thus not restricted to a particular ecological epidemic. These results provide evidence for substructuring due to host-specificity and we propose that this may be a general feature when considering foci of sleeping sickness in eastern and southern Africa. This proposal is illustrated in figure 2 for one minisatellite marker. Inspection of the electrophoretic analysis and the allele frequency distributions in this illustration shows that samples isolated from humans contain only two alleles and one genotype. (Trypanosomes are diploid.) Trypanosomes isolated from cattle are less homogeneous, while the tsetse isolates are the most heterogeneous, containing alleles found in the human and cattle samples and also novel alleles. The host origin of the trypanosome genotypes that are found in the tsetse samples but not in the cattle or human ones is uncertain, but they may derive from wild game. Clearly, if this proposal is correct, then obtaining a large number of samples from a range of different hosts including tsetse flies is the only way to observe the entire picture of *T. brucei* population dynamics.

The importance of population substructuring due to geographical isolation has also been investigated for *T. brucei* by analysing isolates taken from tsetse flies from two separate locations 400 km apart. By comparing samples from the same host, the likelihood of finding alleles in common is maximized. Despite this, the results obtained show that many alleles were unique to each geographical location, supporting the conclusion that these populations are genetically isolated with limited or no gene flow between them (MacLeod *et al*. 2001*b*). This lack of gene flow between trypanosomes from geographically isolated regions has often not been taken into account in previous analyses of the population genetics of *T. brucei* but could account for some of the linkage disequilibrium observed (figure 1*b*) (see for example Tibayrenc *et al*. 1990). This conclusion highlights the need to analyse populations isolated from a single geographical location when addressing the question of whether genetic exchange plays a role in generating diversity between isolates. MacLeod *et al*. (2000, 2001*a*) compared *T. b. rhodesiense* isolates from Uganda with isolates from another focus in Zambia and showed that a completely different genotype was observed in this second focus. This indicates that by using a minisatellite marker system it may be possible to identify the major human-infective genotype for each focus and to use this to track the spread of individual foci.

The clear-cut genetic distinction between the Ugandan and Zambian samples raises the question as to whether other foci are also distinct. Evidence from a Tanzanian focus would suggest that this may well be the case (Komba *et al*. 1998).

6. WEST AFRICAN POPULATIONS

The existence and importance of an animal reservoir for *T. b. rhodesiense* stimulated research on the same question with respect to *T. b. gambiense* using both serum-resistance tests and characterization of isolates by enzyme electrophoresis (Mehlitz *et al*. 1982). Analysis of the zymodemes of a range of isolates from humans, tsetse flies and domestic animals, collected in various countries of western and Central Africa, identified two distinct groups of isolates, which have been referred to as 'type 1' and 'type 2' *T. b. gambiense*, with the latter group showing greater similarity to *T. b. rhodesiense* and *T. b. brucei* (Gibson 1986). The type 1 *gambiense* isolates show characteristic biological features such as low virulence to rodents and stable expression of serum resistance while type 2 are virulent to rodents and can lose the expression of resistance on rodent passage. Furthermore, isoenzyme characterization has identified type 1 *gambiense* in domestic animals, particularly the pig (Gibson 1986).

The conclusion that there are two distinct types of trypanosomes causing disease in western and Central Africa has been further tested using different molecular markers. In a substantive analysis, a collection of 71 isolates from different parts of Africa including Cameroon, Zaire, Congo, Togo, Côte d'Ivoire, Liberia and the Central African Republic were analysed by Southern blotting of restriction digests probed with three VSG genes and two anonymous genomic fragments (Paindavoine *et al*. 1986). The results confirmed and extended the conclusions reached with the isoenzymes analysis and provided a probe (AnTat 1.1) that only hybridized with *T. b. rhodesiense*, *T. b. brucei* and type 2 isolates of human origin from western Africa. Using Jaccards' index of similarity to compare the banding patterns of the digests, it was possible to group the stocks by their similarity and show that the type 1 isolates form a discrete and highly homogeneous group that includes isolates from humans, sheep, pigs and a dog. These conclusions, using some of the same isolates as well as a further collection from humans in Côte d'Ivoire, were also confirmed by Southern blotting with repetitive DNA probes (Hide *et al*. 1990) and RAPDs (Truc & Tibayrenc 1993). Thus it seems clear that there are two distinct types of human-infective trypanosomes in western Africa; one (type 1) being highly homogeneous by marker analysis and the other (type 2) being more similar to East African *T. b. rhodesiense* in terms of its properties. There is no evidence that there are differences in the clinical course of disease between these two types. Type 1 is widely distributed geographically, having been identified in Cameroon, Congo, Zaire, Côte d'Ivoire and Liberia and isolates from these regions show no obvious genetic substructuring. The situation, as regards type 2, is less clear-cut and although the majority of isolates are from Côte d'Ivoire, this may merely reflect the greater number of samples examined from this location.

The population structure of West and Central African *T. brucei* is obviously complicated by the subdivision described above and the issue of the role of genetic exchange in either group has not been extensively studied. Using both isoenzyme electrophoresis and RAPDs, evidence has been presented for linkage disequilibrium (Tibayrenc *et al*. 1990; Truc & Tibayrenc 1993) and, in the case of the isoenzyme data, for departures from Hardy–Weinberg equilibrium leading to the conclusion that, at least for the type 1 isolates, the population is clonal and mating plays a limited role in the generation of variation. The situation as regards the type 2 isolates remains to be elucidated as few isolates have been analysed. In general terms, the current data support the existence of two independently arising human-infective trypanosomes subspecies in western Africa: one is clonal and widely distributed geographically (type 1), while the other shows close similarity to West African *T. b. brucei* (Hide *et al*. 1990) and has similarities in biological properties to *T. b. rhodesiense* from eastern Africa. Whether type 2 has arisen from West African *T. b. brucei* by acquiring the ability to infect humans is an open question but, if correct, suggests that the ability to infect humans has arisen independently not only in different foci in eastern Africa but also in western Africa.

7. HUMAN INFECTIVITY

Several early studies demonstrated that isolates of *T. brucei* from game and cattle were infective for humans whereas others were not and that this ability to infect humans was a stable genetic trait (Heisch *et al*. 1958; Ashcroft 1959 and references therein). The principal limitation on studies such as these was their reliance on human volunteers, which, apart from the very obvious ethical considerations, raised difficulties in terms of sample sizes, adequate controls and repeatability of results. A seminal breakthrough was the development of the blood incubation infectivity test by Rickman & Robson (1970*a*,*b*) to replace inoculation of volunteers. In this assay, trypanosomes are incubated with human or control serum *in vitro* and then inoculated into mice to determine whether the human serum has killed the trypanosomes or not. All current studies on human infectivity in *T. brucei* are based on variants of this simple and robust assay, often replacing the mouse-infectivity component with an *in vitro* assessment of parasite viability.

In the original description of the assay, three phenotypes were identified: sensitive (all trypanosomes were killed by serum and no mice were infected), resistant (all mice became infected), and intermediate (only some mice became infected) (Rickman & Robson 1970*a*,*b*). It has become clear over the intervening years that some isolates are always sensitive, some are always resistant, but some isolates give variable or inconsistent results (Mehlitz *et al*. 1982; Paindavoine *et al*. 1986; Brun & Jenni 1987). For example, it is possible to select for resistance in some isolates that are serum-sensitive. Conversely, prolonged passaging in rodents can lead to serum-resistant isolates becoming sensitive. Thus, it is no longer clear that the 'intermediate' category is a discrete phenotype as originally described, or represents a population of cells of mixed sensitive and resistant phenotype. It is unfortunate

indeed that the classical taxonomy distinction of the three subspecies (Hoare 1972) should be substantially based on a phenotypic characteristic that has turned out to be stable in some but not all isolates.

The instability of the human serum-resistance phenotype in some trypanosome lines has, however, proved very helpful in the investigation of the molecular mechanisms determining human infectivity, as it has enabled the development of isogenic subclones for comparative studies (De Greef *et al*. 1989; De Greef & Hamers 1994; Hager & Hajduk 1997). Using biochemical and molecular approaches to compare such lines, the key host molecule that causes lysis by human serum appears to be haptoglobin-related protein as a component of human high-density lipoprotein termed TLF1 or in a high-molecular weight protein complex, TLF2 (Smith *et al*. 1995; Raper *et al*. 1999). The key molecule in the trypanosome is the serum-resistance associated (SRA) gene product which, if expressed, permits binding of TLF, but not internalization (Hager & Hajduk 1997; Van Xong *et al*. 1998; Milner & Hajduk 1999). In the absence of SRA expression, TLF is internalized by endocytosis and lysis of parasites results. In one study, expression of the SRA gene was determined by its presence in an active variant surface glycoprotein expression site (Van Xong *et al*. 1998). It is perhaps important to note that while comparison of isogenic lines has made the mechanistic analyses more tractable, it implicitly limits the generality of any conclusions drawn with regard to other stocks. Thus, it was noted, for example, that when several different pairs of isogenic lines were compared, SRA expression correlated with resistance in some cases but not all (Rifkin *et al*. 1994). Despite the recent breakthroughs (Van Xong *et al*. 1998; Milner & Hajduk 1999), there is clearly considerable further study required to understand the mechanism(s) of resistance and sensitivity to human serum by *T. brucei* parasites and the role of SRA expression in this process.

Relating the trait of human infectivity to the population genetics of *T. brucei* has been undertaken only rarely. In a detailed study of isolates from cattle, tsetse and humans in Busoga region, Uganda (Hide *et al*. 1994), the human serum resistance of each isolate and a molecular fingerprinting approach and cluster analysis showed that *ca*. 20% of cattle harboured human-infective parasites (a value in line with other studies (Hide 1999)) and these were genetically distinguishable from the non-infective isolates. Nevertheless, all isolates in this geographical focus were clearly more closely related to each other than to isolates from other parts of Africa. Recently, this study has been extended to a number of isolates from Uganda, Kenya and Zambia and used a combination of multilocus genotyping of three highly polymorphic minisatellite loci and minisatellite variant repeat mapping by PCR of one of these loci, *MS42*. This approach accesses a greater degree of polymorphism and is amenable to cladistic analysis to trace stock relationships and evolution (MacLeod *et al*. 2000, 2001*a*). Importantly, minisatellite markers are amenable to allelic interpretation, whereas DNA fingerprinting is not. The data from this analysis show that human-infective isolates from Uganda and Zambia are genetically very different and suggest that these two populations have different evolutionary origins (MacLeod *et al*. 2000, 2001*a*).

8. CONCLUDING REMARKS

Population genetic analysis indicating multiple origins of human infectivity in *T. brucei* poses problems for taxonomists, geneticists and those seeking to understand the molecular mechanisms underlying human infectivity. With regard to taxonomy and genetics, it seems on the basis of current evidence that the subspecies *T. b. rhodesiense* is genetically inappropriate. Human-infective isolates are best considered as host-range variants of the local *T. b. brucei* populations that have become genetically isolated. Furthermore, human-infective isolates are genetically distinct when comparing different geographical locations.

DNA fingerprinting and isoenzyme electrophoresis data split *T. b. gambiense* isolates into type 1 and type 2 populations and some data suggest the type 2 isolates are closely related to West African *T. b. brucei* isolates. Both types are only distantly related to Ugandan and Zambian human-infective isolates but there is clearly an urgent need to revisit the population genetics of West African trypanosomes using highly polymorphic markers that permit allelic interpretation.

When taken together, our current best estimate from the extant literature would suggest that *T. brucei* has acquired the ability to infect humans on four separate occasions during its evolution, giving rise to *T. b. rhodesiense*, Uganda; *T. b. rhodesiense*, Zambia; *T. b. gambiense* type 1 and *T. b. gambiense* type 2. As further *T. b. rhodesiense* foci are investigated and when *T. b. gambiense* populations are reanalysed this number may need modification. Also, it would be helpful to extend the range of genetically unlinked mini- and microsatellite markers available for screening (MacLeod *et al*. 1999; Biteau *et al*. 2000) to guard against the formal possibility of type II errors in current interpretations. Perhaps there is a salutary lesson here for infection disease biology in general. In many studies of the evolutionary origin of a human pathogen there is an implicit assumption that there was a single event with subsequent spread through the human population. In the best-studied protozoan pathogen, *P. falciparum*, which is the principal cause of malaria, it would appear that this is indeed the case (Rich *et al*. 1998; Conway *et al*. 2000). If *T. brucei* is considered as a case study, however, we would suggest that the assumption is dangerously simplistic.

A practical limitation in studies to determine molecular mechanisms of human infectivity is that they necessarily focus on a very limited number of isolates, often only one. There is always some risk in generalizing the results from one isolate to an entire species; this risk is amplified considerably in *T. brucei* because there is no *a priori* reason why different evolutionary events should have come about using the same molecules and mechanisms. For example, care is needed in generalizing as to whether the expression of the SRA gene (Van Xong *et al*. 1998) is the sole mechanism determining human infectivity in *T. brucei*, or one of several mechanisms each used in different foci. On a positive note, many of the materials and reagents (isolate collections, molecular probes, etc.) are available to render such studies tractable now that the need has been recognized.

C.M.R.T. and A.T. gratefully acknowledge The Wellcome Trust for its continued financial support of work in its laboratories.

REFERENCES

Ashcroft, M. T., 1959 The Tinde experiment: a further study of the long-term cyclical transmission of *Trypanosoma rhodesiense*. *Annls Trop. Med. Parasitol*. **52**, 137–146.

Biteau, N., Bringaud, F., Gibson, W., Truc, P. & Baltz, T. 2000 Characterisation of *Trypanozoon* isolates using a repeated coding sequence and microsatellite markers. *Mol. Biochem. Parasitol*. **105**, 185–201.

Brun, R. & Jenni, L. 1987 Human serum resistance of metacyclic forms of *Trypanosoma brucei brucei*, *T. brucei rhodesiense* and *T. brucei gambiense*. *Parasitol. Res*. **73**, 218–223.

Cibulskis, R. E. 1988 Origins and organization of genetic diversity in natural populations of *Trypansoma brucei*. *Parasitology* **96**, 303–322.

Cibulskis, R. E. 1992 Genetic variation in *Trypanosoma brucei* and the epidemiology of sleeping sickness in the Lambwe Valley, Kenya. *Parasitology* **104**, 99–109.

Conway, D. J. (and 11 others) 2000 Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol*. **111**, 163–171.

De Greef, C. & Hamers, R. 1994 The serum resistance-associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. *Mol. Biochem. Parasitol*. **68**, 277–284.

De Greef, C., Imbrechts, H., Matthyssens, G., Van Meirvenne, N. & Hamers, R. 1989 A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. *Mol. Biochem. Parasitol*. **36**, 169–176.

Gibson, W. C. 1986 Will the real *Trypansoma b. gambiense* please stand up. *Parasitol. Today* **2**, 255–257.

Gibson, W. C. & Stevens, J. 1999 Genetic exchange in the Trypanosomatidae. *Adv. Parasitol*. **43**, 1–45.

Gibson, W. C., Marshall, T. F. C. & Godfrey, D. G. 1980 Numerical analysis of enzyme polymorphism. A new approach to the epidemiology and taxonomy of trypanosomes of the genus *Trypanozoon*. *Adv. Parasitol*. **18**, 175–245.

Godfrey, D. G., Baker, R. D., Rickman, L. R. & Mehlitz, D. 1990 The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. *Adv. Parasitol*. **29**, 1–44.

Hager, K. M. & Hajduk, S. L. 1997 Mechanisms of resistance in African trypanosomes to cytotoxic human HDL. *Nature* **385**, 823–826.

Heisch, R. B., McMahon, J. P. & Manson-Bahr, P. E. C. 1958 The isolation of *Trypanosome rhodesiense* from a bushbuck. *Br. Med. J*. **2**, 1203–1204.

Hide, G. 1999 History of sleeping sickness in east Africa. *Clin. Microbiol. Rev*. **12**, 112–125.

Hide, G., Cattand, P. & Le Ray, D. 1990 The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. *Mol. Biochem. Parasitol*. **39**, 213–226.

Hide, G., Welburn, S. C., Tait, A. & Maudlin, I. 1994 Epidemiological relationships of *Trypanosoma brucei* stocks from south east Uganda: evidence for different population structures in human infective and non-human infective isolates. *Parasitology* **109**, 95–111.

Hoare, C. A. 1972 *The trypanosomes of mammals*. Oxford, UK: Blackwell.

Komba, E. K., Kobona, S. N., Ambwene, A. K., Stevens, J. R. & Gibson, W. C. 1998 Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania. *Parasitology* **115**, 571–579.

Leak, S. G. A. 1999 *Tsetse biology and ecology: their role in the epidemiology and control of trypanosomiasis*. Oxford, UK: CABI.

Letch, C. A. 1984 A mixed population of trypanozoon in *Glossina palpalis palpalis* from Ivory Coast. *Trans. R. Soc. Trop. Med. Hyg*. **78**, 627–630.

MacLeod, A., Turner, C. M. R. & Tait, A. 1999 A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. *Mol. Biochem. Parasitol*. **102**, 237–248.

MacLeod, A., Tweedie, A., Welburn, S. C., Maudlin, I., Turner, C. M. R. & Tait A. 2000 Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic and epidemic population genetic structures. *Proc. Natl Acad. Sci. USA* **97**, 13 442–13 447.

MacLeod, A., Welburn, S., Maudlin, I., Turner, C. M. R. & Tait, A. 2001*a* Evidence for multiple origins of human infectivity in *Trypanosoma brucei* revealed by minisatellite variant repeat mapping. *J. Mol. Evol*. **52**, 290–301.

MacLeod, A., Turner, C. M. R. & Tait, A. 2001*b* The detection of geographical substructuring of *Trypanosoma brucei* populations by the analysis of minisatellite polymorphisms. *Parasitology*. (In the press.)

Mathieu-Daude, F. & Tibayrenc, M. 1994 Isozyme variability of *Trypanosoma brucei* s.l: genetic, taxonomic and epidemiological significance. *Exp. Parasitol*. **78**, 1–19.

Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. 1993 How clonal are bacteria? *Proc. Natl Acad. Sci. USA* **90**, 4384–4388.

Mehlitz, D., Zillman, U., Scott, G. M. *et al*. 1982 Epidemiological studies on the animal reservoir of Gambiense sleeping sickness. III. Characterisation of *Trypanozoon* stocks by isoenzymes and sensitivity to human serum. *Tropenmed. Parasitol*. **33**, 113–118.

Mihok, S., Otieno, L. H. & Darji, N. 1990 Population genetics of *Trypanosoma brucei* and the epidemiology of human sleeping sickness in the Lambwe Valley. *Parasitology* **100**, 219–233.

Milner, J. D. & Hajduk, S. L. 1999 Expression and localisation of serum resistance associated protein in *Trypanosoma brucei rhodesiense*. *Mol. Biochem. Parasitol*. **104**, 271–283.

Paindavoine, P., Pays, E., Laurent, M. *et al*. 1986 The use of DNA hybridization and numerical taxonomy in determining relationships between *T. brucei* stocks. *Parasitology* **92**, 31–50.

Raper, J., Fung, R., Ghiso J., Nussenzweig, V. & Tomlinson, S. 1999 The identification and characterisation of a novel trypanosome lytic factor from human serum. *Infect. Immun*. **67**, 1910–1916.

Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. 1998 Malaria's eve: evidence of a recent population bottleneck throughout the world population of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95**, 4425–4430.

Rickman, L. R. & Robson, J. 1970*a* The testing of proven *T. brucei* and *T. rhodesiense* strains by the blood infectivity test. *Bull. WHO* **42**, 911–916.

Rickman, L. R. & Robson, J. 1970*b* The blood incubation infectivity test: a simple test which may serve to distinguish *Trypanosoma brucei* from *T. rhodesiense*. *Bull. WHO* **42**, 650–651.

Rifkin, M. R., De Greef, C., Jiwa, A., Landsberger, F. R. & Shapiro, S. Z. 1994 Human serum-sensitive *Trypanosoma brucei rhodesiense*: a comparison with serologically identical human serum-resistant clones. *Mol. Biochem. Parasitol*. **66**, 211–220.

Smith, A. B., Esko, J. D. & Hajduk, S. L. 1995 Killing of trypanosomes by the human haptoglobin-related protein. *Science* **268**, 284–286.

Stevens, J. R. & Tibayrenc, M. 1995 Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterized by RAPD analysis and isoenzymes. *Parasitology* **110**, 181–186.

Stevens, J. R. & Tibayrenc, M. 1996 *Trypansoma brucei* s.l.: evolution, linkage and the clonality debate. *Parasitology* **112**, 481–488.

Tait, A. 1980 Evidence for diploidy and mating in trypanosomes. *Nature* **287**, 536–538.

Tait, A. 2000 Trypanosomosis. In *Molecular epidemiology of infectious diseases* (ed. R. C. A. Thompson), pp. 258–282. London: Arnold.

Tait, A. & Turner, C. M. R. 1990 Genetic exchange in *Trypanosoma brucei*. *Parasitol. Today* **6**, 70–75.

Tibayrenc, M., Ward, P., Moya, A. & Ayala, F. J. 1986 Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. *Proc. Natl Acad. Sci. USA* **83**, 115–119.

Tibayrenc, M., Kjellberg, F. & Ayala, F. J. 1990 A clonal theory of parasitic protozoa—the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc. Natl Acad. Sci. USA* **87**, 2414–2418.

Tibayrenc, M., Neubauer, K., Barnabé, C. & Guerrini, F. 1993 Genetic characterisation of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proc. Natl Acad. Sci. USA* **90**, 1335–1339.

Truc, P. & Tibayrenc, M. 1993 Population genetics of *Trypanosoma brucei* in central Africa: taxonomic and epidemiological significance. *Parasitology* **106**, 137–149.

Van Xong, H., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van Den Abeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P. & Pays, E. 1998 A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* **95**, 839–846.

Walliker, D., Quakyi, I. A., Wellem, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Cororan, L. M., Burkot, T. R. & Carter, R. 1987 Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666.

Welburn, S. C. & Maudlin, I. 1992 The nature of the general state in *Glossina* and its role in the acquisition of trypanosome infection in tsetse. *Annls Trop. Med. Parasitol.* **86**, 529–536.

Welburn, S. C. & Maudlin, I. 1997 Current trends in parasite–vector interactions. In *Trypanosomiasis and leishmaniasis: biology and control* (ed. G. Hide, J. C. Mottram, G. H. Coombs & P. H. Holmes), pp. 315–326. Wallingford, UK: CABI.