Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits

Andy Tait[1*]**, Dan Masiga**[2]**, Johnstone Ouma**[2]**, Annette MacLeod**[1]**, Juergen Sasse**[3]**, Sara Melville**[3]**, Gabbi Lindegard**[2]**, Anne McIntosh**[2] **and Mike Turner**[2]

[1]*Wellcome Centre for Molecular Parasitology, 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, Glasgow G12 8QQ, UK*
[3]*Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK*

The genome of the African trypanosome, *Trypanosoma brucei*, is currently being sequenced, raising the question of how the data generated can be used to determine the function of the large number of genes that will be identified. There is a range of possible approaches, and in this paper we discuss the use of a classical genetic approach coupled with positional cloning based on the ability of trypanosomes to undergo genetic exchange. The genetics of these parasites is essentially similar to a conventional diploid Mendelian system with allelic segregation and an independent assortment of markers on different chromosomes. Data are presented showing that recombination occurs between markers on the same chromosome allowing the physical size of the unit of recombination to be determined. Analysis of the available progeny clones from a series of crosses shows that, in principal, large numbers of progeny can readily be isolated from existing cryopreserved products of mating and, taking these findings together, it is clear that genetic mapping of variable phenotypes is feasible. The available phenotypes for analysis are outlined and most are relevant to the transmission and pathogenesis of the parasite.

Genetic maps from two crosses are presented based on the use of the technique of AFLP; these maps comprise 146 and 139 markers in 30 and 21 linkage groups respectively. Segregation distortion is exhibited by some of the linkage groups and the possible reasons for this are discussed. The general conclusion, from the results presented, is that a genetic-mapping approach is feasible and will, in the future, allow the genes determining a number of important traits to be identified.

Keywords: *Trypanosoma brucei*; genetics; recombination; genetic mapping

1. INTRODUCTION

The genetic analysis of phenotypes is a key tool with which to exploit the functional description of the genes identified by the *Trypanosoma brucei* genome sequencing projects. There are two distinct genetic approaches to the analysis of gene function. First, reverse genetic analysis based on the 'deletion' of genes either by the knock-out approach using transfection and homologous recombination (Caruthers *et al.* 1993) or by RNAi, whereby the transcript of a particular gene is degraded transiently (Ngo *et al.* 1998; Shi *et al.* 2000) followed by screening for altered phenotype. Second, forward genetic analysis based either on the generation of mutants followed by the identification of the genes involved or classical genetic analysis where the genetic basis of phenotypic variation can be determined. In the mutant approach, mutants are generated, with or without mutagenesis, followed by selection for a particular non-wild-type phenotype. Examples of such selection range from the specific, such as drug resistance (Carter & Fairlamb 1993), to the more general, such as

the inability to differentiate (Tasker *et al.* 2000). Mutants can then be screened for alterations in candidate genes or analysed by proteomics followed by mass spectrometry (Pandey & Mann 2000; Ashton *et al.* 2001) to identify the altered proteins and, from such data, the genes that determine the mutant phenotype. The classical genetic approach can use such mutants, or naturally occurring variation between parasite stocks followed by crosses, to determine the genetic basis (number of loci, dominance, etc.) of the variation using segregational analyses of progeny. This allows a genetic model of the inheritance of the phenotype to be generated and then tested by backcrosses or F2 progeny analysis. The identification of the genes determining the trait of interest requires linkage analysis using molecular markers in a previously established genetic map in order to identify the region of the genome within which a gene (determining a particular phenotype) lies. This essentially involves identifying markers on the genetic map that co-segregate with the different phenotypes in the progeny of a cross, i.e. are closely linked. Prior to the availability of the genome sequence, the region of the genome containing the gene of interest would be isolated by probing a large-insert genomic library with the linked markers, followed by sequence analysis to identify

*Author for correspondence (a.tait@vct.gla.ac.uk).

Phil. Trans. R. Soc. Lond. B (2002) **357**, 89–99
DOI 10.1098/rstb.2001.1050

89

candidate ORFs. This approach has been termed positional cloning. As the resolution of many genetic maps only allows linkage to be established to within several hundred kilobases, several candidate genes are identified and further analysis is needed to define the specific gene. A number of methods are then available to do this and include reverse genetics, expression analysis and higher-resolution mapping within the identified region of the genome. With trypanosomes, the availability of a substantial body of genomic sequence obviates the need to undertake the cloning step as the sequences of the linked markers can be used to assemble the sequence of the genomic region of interest and from this the identification of the ORFs. A further advantage of the trypanosome system is the use of the technique of RNAi, which could be used to test each candidate ORF for loss of phenotype based on degradation of the gene transcript.

Each of the genetic approaches has advantages and disadvantages in terms of analysing gene function and, at the present time, it is difficult to identify a preferred approach. In this context this paper will review and present progress towards the classical genetic approach for the analysis of phenotype in *T. brucei*. It should be noted that equivalent mapping methods have been used in *Plasmodium falciparum* (Su & Wellems 1996; Su *et al.* 1999), *Toxoplasma gondii* (Sibley *et al.* 1992) and *Eimeria tenella* (Shirley & Harvey 2000) to analyse a number of phenotypes.

2. THE GENETICS OF *TRYPANOSOMA BRUCEI*

In order to illustrate the use of genetic analysis it is necessary to briefly review the basic features of the genetic system identified in *T. brucei*. The existence of a sexual cycle in *T. brucei* was established some 15 years ago (Jenni *et al.* 1986) and subsequent analysis has established that mating occurs between life cycle stages within the tsetse fly vector (Sternberg & Tait 1990; Tait & Turner 1990; Gibson & Stevens 1999). Two different cloned lines can be crossed by infecting tsetse flies with a mixture of the lines and then allowing the infection to undergo the life cycle stages in the fly resulting in the production of metacyclic-stage trypanosomes in the salivary glands. These trypanosomes can be isolated individually and each expanded by growth in mice, sampled over time by allowing the infected tsetse to feed on mice and cloning from the resultant infection, Alternatively, all the trypanosomes from the salivary gland can be used to infect mice and single trypanosomes isolated for expansion from the first peak of parasitaemia (figure 1). Analysis of the segregation of a range of markers (RFLP, iso-enzyme, karyotype and minisatellite) in the cloned trypanosomes from such mixed infections has established that mating is not obligatory (Gibson 1989; Turner *et al.* 1990; Gibson & Bailey 1994), the clones are the equivalent of the F1 progeny in a diploid Mendelian system (Sternberg *et al.* 1989; Turner *et al.* 1990; MacLeod *et al.* 1999) and that mating occurs at a high frequency in terms of the number of infected flies that produce hybrid products of mating (Schweizer *et al.* 1988; Turner *et al.* 1990).

The Mendelian segregation is most readily illustrated by analysis of the chromosomes by PFGE in the progeny exploiting the size polymorphism of the parental trypanosomes (Tait *et al.* 1993; Melville *et al.* 1998). The karyo-
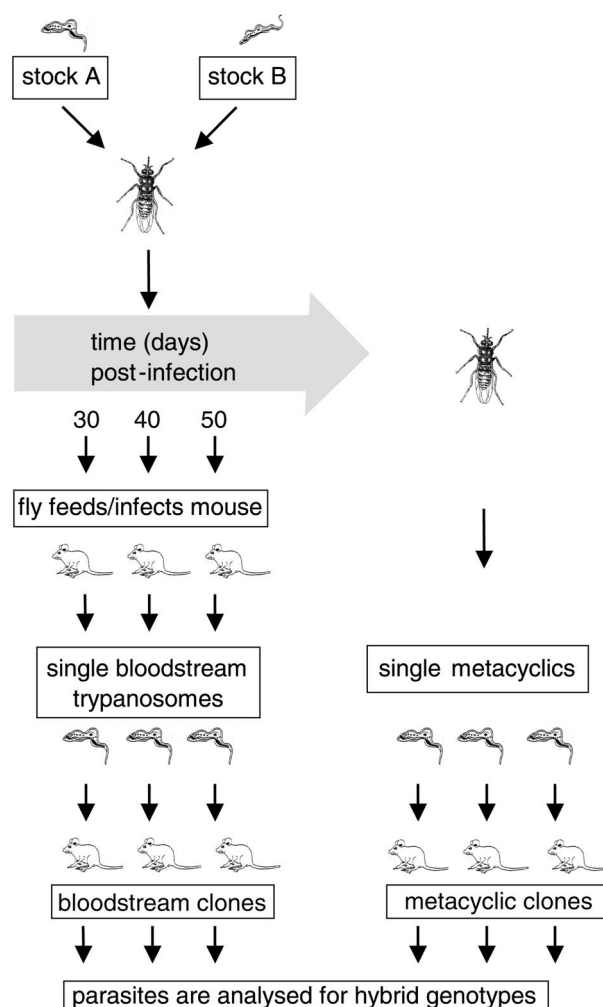

Figure 1. The methods for undertaking crosses in *T. brucei*. Tsetse flies are infected with two different stocks (A and B) of the parasite and, after maturation, clones are isolated either by directly infecting mice with single metacyclic stage trypanosomes or by infecting a mouse with a whole population from the infected fly and subsequent cloning from the resultant bloodstream stage infection. The cloned trypanosomes can then be genotyped using informative polymorphic markers.

type of stocks 247 and 386 has been defined by PFGE followed by Southern blotting with a series of EST probes to identify homologous and non-homologous chromosomes (Melville *et al.* 1998, 1999). The data from these studies show that there is considerable size polymorphism between homologous chromosomes in different stocks of the parasite and that, within a cloned stock, the two homologues of some of the chromosomes also differ substantially in size. The sizes of chromosomes are stable at mitosis, at least at gross level (Tait *et al.* 1993), and therefore these polymorphisms can be used as genetic markers to follow the inheritance of the chromosomes. The inheritance of chromosome I has been analysed by separating the chromosomes of two parental stocks (247 and 386) and a selection of the progeny clones resulting from a cross between them. After PFGE separation, chromosome I is detected by Southern blotting using a gene probe that is chromosome specific. The results of this experiment are illustrated in figure 2; the parental stocks each show two bands of hybridization of different size, which also differ
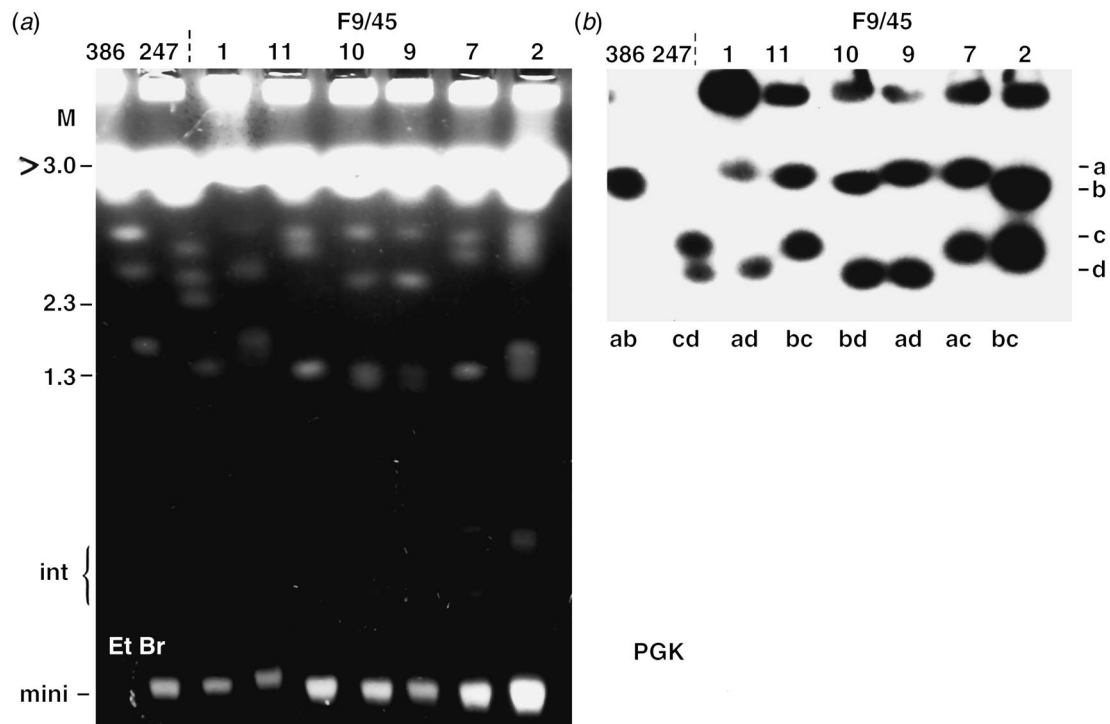
Figure 2. The inheritance of chromosome I in the progeny from a cross between stocks 386 and 247. (*a*) The result of PFGE separation of the chromosomes of the two parental stocks together with six F1 progeny clones. The gel has been stained with ethidium bromide. (*b*) An autoradiograph of a Southern blot of the same gel hybridized with a probe for the gene encoding PGK. The approximate sizes of the chromosomes are indicated in megabases. The designation of the parental stocks and the F1 progeny derived from them are indicated at the top of each track. Each homologue of chromosome I has been designated by a letter (a,b,c,d) and the combinations of these for the progeny indicated in (*b*).

in size between the two stocks. The progeny clones have non-parental karyotypes composed of one homologue of chromosome 1 from each parent and they show all possible combinations of parental homologues (Tait *et al.* 1993). A similar illustration of the inheritance is provided by the segregation of the alleles of the minisatellite loci 292 and CRAM into the progeny; all possible combinations of the parental alleles are observed in the progeny with each progeny clone inheriting one allele from each parent, as would be predicted in a standard diploid genetic system (MacLeod *et al.* 1999). The basic properties of this system of genetic exchange have been reviewed (Sternberg & Tait 1990; Tait & Turner 1990; Gibson & Stevens 1999) and will therefore not be discussed further, other than to conclude that the genetic system has all the properties of that predicted for a standard diploid organism. Analysis of progeny clones from a number of different crosses has shown that some are triploid or possibly trisomic (Gibson *et al.* 1991; Hope *et al.* 1999) and the frequency of these clones in the progeny depends on the particular cross being considered. In the crosses considered here the frequency is low (Hope *et al.* 1999) and the triploid progeny have been excluded from the analyses.

3. GENETIC MAPPING

Two central questions need to be addressed about the feasibility of mapping the genetic basis of particular phenotypes: first, the ability to generate sufficient numbers of progeny clones (F1, F2 and backcross) and, second, the formal demonstration of crossing over coupled with the determination of the physical size of the recombination

unit. The number of progeny affects the ability to determine statistical significance of segregating phenotypes in relation to the number of loci or alleles determining a trait of interest, as well as the level of resolution of a genetic map. The physical size of the recombination unit determines the number of markers required for such a map in order to obtain linkage to within a specified and workable distance from the locus determining a particular trait.

In order to address the first question, an analysis of the number of independent progeny clones produced in three different crosses has been undertaken using three unlinked minisatellite loci (MacLeod 1999; MacLeod *et al.* 2000) as markers. Progeny can be derived by either of two routes, as illustrated in figure 1; the genotypes of clones derived from tsetse flies infected with mixtures of two cloned parental lines have been determined for each of three such mixed infections using both routes for the isolation of clones. Table 1 provides an analysis of the number of F1 progeny clones per cross in this non-obligatory sexual system, the proportion of parental non-mated clones and the proportion of genetically unique progeny clones (the products of independent mating events) derived by the two routes. The clones derived directly from the salivary glands of the tsetse (metacyclic clones) show a relatively low proportion of F1 progeny (average 27%) but over 80% of these are unique. By contrast, the clones derived after the growth of all the salivary gland population in mice prior to cloning (bloodstream clones) show 80% of them to be F1 progeny; however, a lower proportion of these are unique (average 60%) indicating that some clones are the vegetative progeny of the same trypanosome that originally infected the mouse. While

Table 1. *The proportion of F1 progeny clones derived by two different routes from* T. brucei *crosses.*

(The products of crosses between three parasite stocks were derived either by direct cloning of the metacyclic stage from infected tsetse (metacyclic clones) or cloning from the first peak of parasitaemia obtained in mice after infection with the whole salivary gland population of metacyclic trypanosomes (bloodstream clones). Each clone was then genotyped as described in the text. Abbreviations: *n*, number of clones analysed; P, parental genotype; HYB, hybrid genotype; Unique, F1 progeny from different mating events.)

cross	bloodstream clones (%)				metacyclic clones (%)			
	n	P	HYB	Unique	*n*	P	HYB	Unique
247 × 927	15	0.33	0.67	0.7	23	0.57	0.43	0.9
386 × 927	13	0.23	0.77	0.7	11	0.91	0.09	1.0
247 × 386	23	0.04	0.96	0.43	36	0.72	0.28	0.6
average	17	0.2	0.8	0.61	23	0.73	0.27	0.83

there is some variation between the different crosses in the proportions of unique or parental clones, this does not alter the general conclusion that the proportion of progeny is higher in the bloodstream clones and, despite a lower proportion of unique progeny clones, a large number of independent progeny can be generated. Furthermore, these data show that despite the non-obligatory nature of mating, a significant proportion of the trypanosomes are the products of mating. As cryopreserved non-cloned products of mixed-infected tsetse are available, the only limitation on generating a very large collection of progeny is the labour involved in cloning. To date, no F2 progeny have been generated, although backcrosses have been successfully undertaken (Gibson *et al*. 1995; A. Tait, C. M. R. Turner and S. Welburn, unpublished data).

In order to demonstrate that crossing over or recombination occurs and to obtain a first estimate of the size of the recombination unit in this system, the inheritance of three markers on chromosome I, which are a known physical distance apart, has been determined. If recombination occurs it would be predicted that non-parental haplotypes would be observed in the progeny clones as a result of crossing over between the alleles at pairs of loci (it should be remembered that *T. brucei* is diploid). The observation of such non-parental haplotypes constitutes formal proof of the occurrence of recombination, which is an absolute requirement for generating a genetic map, and the frequency of recombination between pairs of loci allows the calculation of the size of the recombination unit. In stock 927, three linked markers have been identified that are heterozygous, such that the inheritance of each homologue can be determined by analysing the inheritance of the alleles in the progeny of crosses between stocks 386 or 247 and 927. Each of a total of 32 progeny clones has been genotyped with the three linked markers, allowing the frequency of each combination of alleles to be determined. This allows the designation of the haplotype in the parental stock, 927, and therefore the definition of the number of progeny in which crossing over has occurred between the pairs of markers. The markers used in this analysis are the minisatellite marker MS42 (Barrett *et al*. 1997) and two CA microsatellites (Sasse 1998). The results of this analysis are illustrated in figure 3. The parental haplotypes are defined by determining the two most frequent classes of allele combinations observed in the progeny and are illustrated diagrammatically, together with the haplotypes observed that are the result of a cross-over between pairs of markers. The frequency of the cross-overs is determined as the proportion in the total number of progeny clones analysed expressed as a percentage and thus yields the genetic distance between the markers in centimorgans. As the physical distance between each marker is known from the genome sequence of chromosome I (http://www.sanger.ac.uk), the physical size of the centimorgan can be simply calculated (figure 3) to give values of 13.8 and 8.5 kb cM^{-1} for the two intervals respectively. The average value obtained (11.2 kb cM^{-1}) for the marker intervals analysed is comparable to the value obtained for *Plasmodium* (Su *et al*. 1999) and thus a 10 cM map of the genome would allow linkage to be determined to within 110 kb of any gene. This is a realizable objective and means that a genetic map is feasible as a tool for positional cloning of loci determining traits of interest.

4. VARIANT PHENOTYPES

The value of classical genetic analysis is partly determined by the phenotypes that are available for investigation and the inability of other approaches to easily define their basis. Crosses have been undertaken between three different cloned isolates of *T. brucei* in the Glasgow laboratory (Turner *et al*. 1990) and these stocks have been shown to differ phenotypically in a number of different assays. As a detailed description of these differences will be the subject of other publications, they will only be summarized here. Stock STIB 386 was isolated from a human in West Africa and is therefore human infective by definition and, from marker analysis, is a Type 2 *T. b. gambiense* (Gibson 1986). Stock STIB 247 was isolated from wild game in Tanzania whereas stock TREU 927 was isolated from a tsetse fly in Kenya. Each of these parental lines was screened for a range of phenotypes that include HSR, resistance to arsenical drugs, virulence, infectivity for the tsetse midgut and ability to produce salivary gland infections. Variation between the stocks has been demonstrated in each of these phenotypes thus allowing their genetic basis to be analysed using the progeny from the three crosses. These phenotypes are relevant to the transmission, pathogenesis, treatment and host range of the parasite. The differences are summarized qualitatively in table 2. In addition, two of the parental lines have been selected for resistance to high levels of either an arsenical drug or suramin (Scott *et al*. 1996) and therefore can also
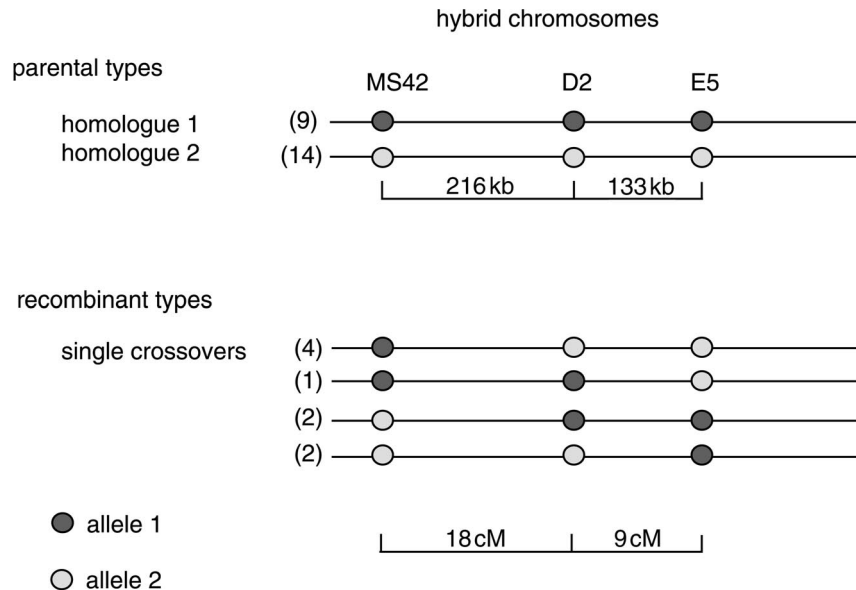
Figure 3. Recombination between loci on chromosome I. The haplotypes of the parental stocks and the progeny from them are indicated diagrammatically for the three heterozygous loci analysed (MS42, E5 and D2) with the different alleles at each locus indicated by shading. The recombinant haplotypes identified in the F1 progeny are illustrated below the parental types. The numbers of parental and recombinant haplotypes observed in the progeny from this cross are indicated in brackets as well as the genetic and physical distance between each marker.

Table 2. Phenotypic differences between the three parental stocks (386, 247 and 927) used for genetic analysis.
(Abbreviations: HSR, human serum resistance/sensitivity (Lindegard 1999); virulence, number of days to parasitaemia $>10^6$ (Turner *et al.* 1995); MG, relative level of tsetse midgut infectivity (MacLeod 1999); TI, transmission index (Welburn *et al.* 1995); Mel^R, resistance to cymelarsan (Scott *et al.* 1996; M. Hope, A. Tait and C. M. R. Turner, unpublished data).)

	phenotype				
stock	HSR	virulence	MG	TI	Mel^R
927	R/S	+++	+++	++	R
247	S	+	+++	+++	S
386	R	+++	+	+	S

be used for genetic analysis using appropriate crosses to drug-sensitive lines.

5. CONSTRUCTION OF A GENETIC MAP

There are a number of different marker systems that could be used to construct a genetic map of *T. brucei*. These include the use of available genome sequence data to identify microsatellites or define regions of the genome for RFLP analysis; such approaches have been used successfully to generate genetic maps of *Plasmodium* spp. at different levels of resolution (Walker-Jonah *et al.* 1992; Carlton *et al.* 1998; Su *et al.* 1999). In addition, it is possible to use a more global system such as AFLP (Vos *et al.* 1995; Masiga *et al.* 2000; Savelkoul *et al.* 1999), which does not require any sequence information as it relies on the sequences of restriction sites. We have used this technique to identify large numbers of polymorphic markers between the three parental lines of *T. brucei*, based on the recognition sites of two pairs of restriction enzymes

*Eco*R1/*Mse*1 and *Hind*III/*Taq*1. This technique involves the digestion of genomic DNA with the pair of restriction enzymes (one six base cutter and one four base cutter) followed by ligation with a pair of adaptors, one of which has the recognition sequence of one restriction enzyme cut site, while the other has the recognition sequence for the second enzyme. This ligation mixture is then subjected to PCR amplification in the 'presence' of a labelled dinucleotide tri-phosphate, using primers specific to the adaptors and the respective restriction-site recognition sequences. The amplified fragments will be labelled and only comprise those fragments from the original digest that were derived by double digestion, i.e. had a combination of both restriction sites. These can then be separated on a standard sequencing gel and detected by autoradiography.

A typical product of AFLP analysis is illustrated in figure 4*a* and shows a comparison of two parental stocks (P1 and P2), which were crossed to yield 15 F1 progeny clones. Comparing the two parental stocks, polymorphisms are detected as the presence of a band in one stock and absence in the other (arrows, figure 4); when the progeny are examined it can be seen that this absence/presence segregates. This can be interpreted as the stock without the band being homozygous for the absence of one of the restriction sites and the stock with a band as being heterozygous for one of the restriction sites (see figure 4*b* for interpretation). In the examples shown, each progeny clone will inherit the absence of the site from one parent but half will inherit the presence of the site from the heterozygous parent and therefore have a band (figure 4). Such segregating markers can then be used to construct a genetic map of the parasite. A feature of the AFLP system is the ability to analyse large numbers of markers simultaneously (figure 4*a*), and the use of markers heterozygous in one parent and homozygous in the other is formally similar to a standard backcross. A
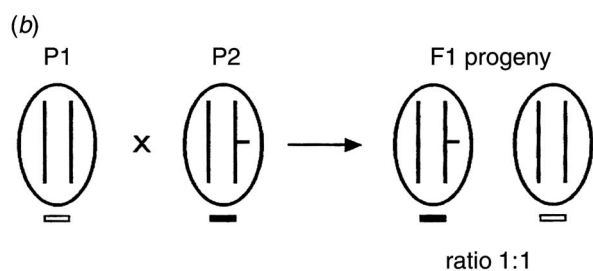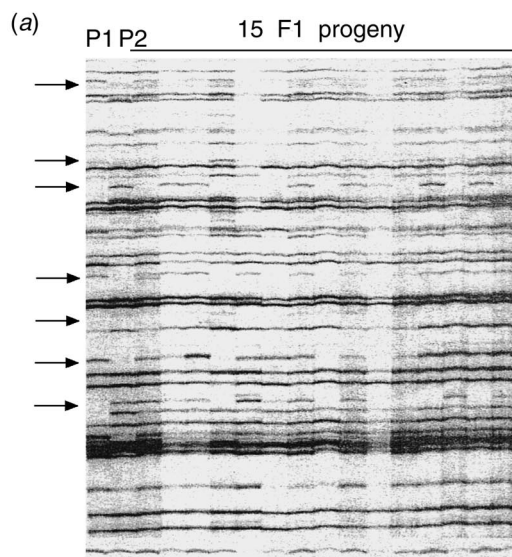
(a)

(b)

ratio 1:1

Figure 4. An illustration of the inheritance of AFLPs in a cross between two stocks of *T. brucei*. (*a*) The cloned stocks 247 (P1) and 927 (P2) were used as parents in a genetic cross to generate 15 diploid F1 progeny that are all products of independent mating events. DNA from each of these cloned trypanosomes were analysed by amplification using a specific pair of primers as described by Masiga *et al.* (2000), the products separated on a sequencing gel and detected by autoradiography. Bands of the same mobility in different tracks mark equivalent loci. Some bands are non-polymorphic and are present in all tracks while others are polymorphic, with a band visible for one parent but not the other and do not show segregation into the F1. Other bands are both polymorphic and show segregation into the F1 (marked by arrows) and can thus be used for genetic mapping. (*b*) An illustration of the allelic interpretation of the segregating polymorphic bands on which the construction of the linkage map is based.

total of 700 AFLP markers are polymorphic between the three stocks and are used for this analysis, and 343 of these segregate in the F1 progeny of crosses. Only nine of these markers were heterozygous in stock 247 which is consistent with previous analysis showing that this stock is homozygous at most loci (Sternberg *et al.* 1989; Tait *et al.* 1996) and making it equivalent to a highly inbred line. The AFLP marker linkage analysis has been carried out for two crosses—those between stocks 247 and 927 (map 1) using 15 F1 progeny and stocks 247 and 386 (map 2) using 17 F1 progeny. All the progeny used are the products of independent mating events based on marker analysis (Sternberg *et al.* 1989; Turner *et al.* 1990; MacLeod 1999).

The construction of a genetic map essentially involves the identification of markers that co-segregate and are
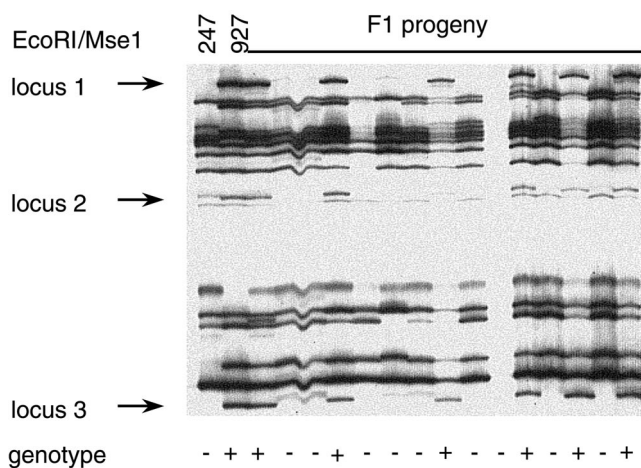

Figure 5. Part of an AFLP autoradiogram showing genetic linkage. A general description and interpretation of the inheritance of AFLP polymorphisms is given in figure 4. In this particular example, three polymorphic bands (arrowed) show the same pattern of segregation in the F1 progeny, thus indicating that their inheritance is tightly linked and they will thus occupy identical or similar positions in a genetic map.

therefore linked, coupled with measurement of the frequency of recombination between such markers to provide a measure of the distance between them. An example of linkage between three loci is shown in figure 5 where the three AFLP markers (indicated by arrows) clearly co-segregate and, as a result of the lack of crossovers, these would be defined as being at the same locus. Linkage between the markers for each cross was determined by multipoint analysis using MAPMAKER (Lander *et al.* 1987; http://www.mcbio.med.buffalo.edu/MMQTX.html). Results are shown at a stringency of $p = 0.05$ with the Haldane correction of map distance. The data obtained are summarized in table 3 for each of the maps. It is clear that the number of linkage groups (30 groups for cross 247×927; 21 groups for cross 386×927) is some two- to three-fold greater than the number of diploid chromosomes ($n = 11$) and suggests that there are 19 gaps in map 1 and 10 gaps in map 2. Generating further markers or screening larger numbers of progeny could close these gaps. Details of these two maps are presented in table 4, showing an average genome coverage of 469 cM. A high proportion of the informative markers are linked but there are a number that do not fall into linkage groups and, as no attempt has been made to physically link any of the markers, it is possible that some of these are not part of the diploid genome. These could be derived from the intermediate or mini chromosomes and possibly from the mitochondrial genome (maxicircle). It is assumed that the larger linkage groups represent the physically large chromosomes based on the higher probability of detecting markers in these. This could be verified by probing amplified markers from each linkage group onto PFGE-separated chromosomes or by hybridization to physically mapped P1 or BAC clones. Preliminary results show that this is feasible and a number of the linkage groups have been anchored to specific PFGE-separated chromosomes or arrayed P1 library clones using PCR-amplified AFLP fragments (Masiga *et al.* 2000). This work is still in progress.

Table 3. *Summary of data for two genetic maps, each generated from crosses using a different pair of parents. Part of map 2 is illustrated in figure 6. The full maps are available at http://www.gla.ac.uk/ibls/II/cmrt/*

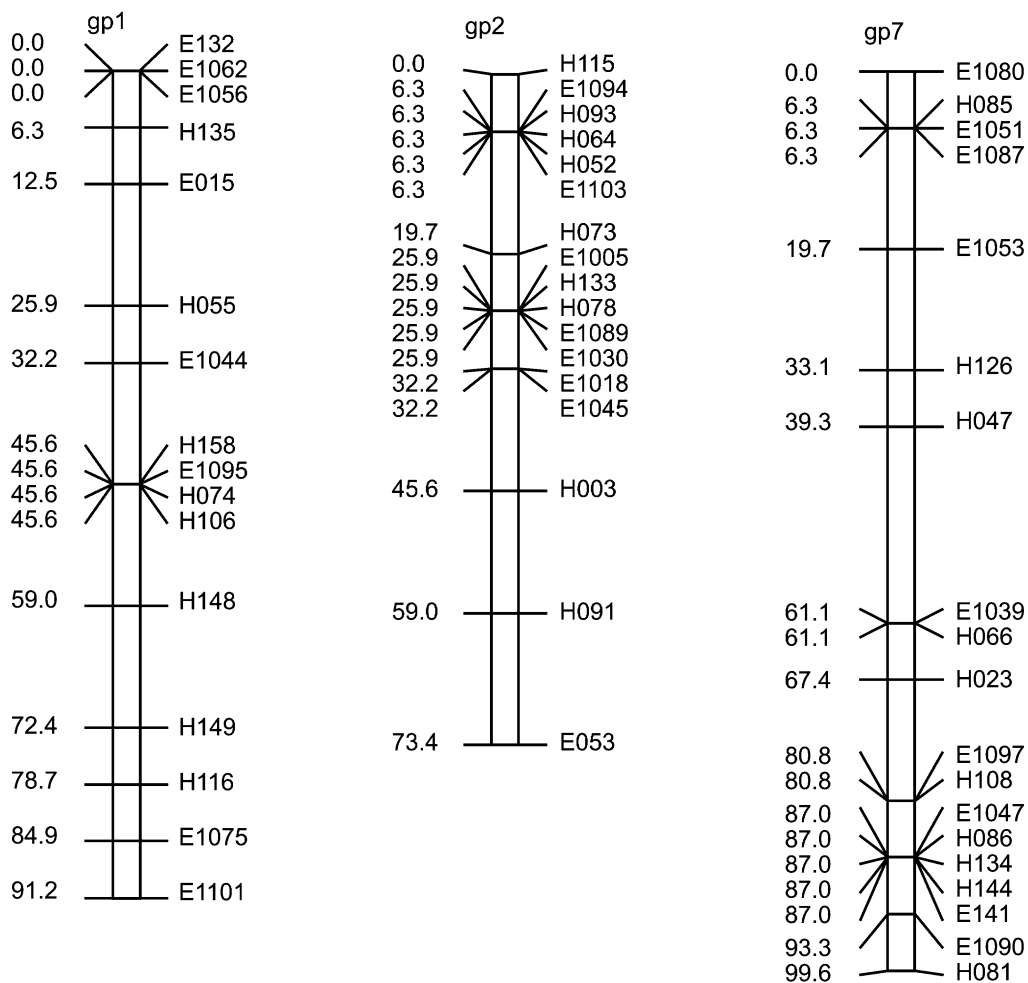| genetic cross (map)	group	number of markers	number of positions	genetic distance (cM)
247 × 927 (map 1)	larger groups:			
	1	12	6	36.9
	2	12	5	30.4
	4	17	5	30.9
	8	12	9	60.6
	20	5	5	28.6
	smaller groups:	2–11	1–4	0–25.1
	total:			
	30	146	86	424.8
247 × 386 (map 2)	larger groups:			
	1	16	11	91.2
	2	17	8	73.4
	3	7	4	33.1
	4	10	4	27.3
	7	19	11	99.6
	8	12	6	38.4
	16	4	4	32.2
	smaller groups:	2–7	1–3	0–26.8
	total:			
	21	139	74	513.3

Figure 6. Linkage maps for the three largest groups generated using the genetic cross, 247 × 386 and AFLP analysis. On the left of each group is the cumulative genetic distance in centimorgans (Haldane corrected) and on the right the list of markers. Those markers prefixed 'H' were generated using restriction enzymes *Hind*III and *Taq*1; markers prefixed 'E' were generated using restriction enzymes *Eco*R1 and *Mse*1. The complete maps are given at http://www.gla.ac.uk/ibls/II/cmrt.

Table 4. *Summary statistics for the two AFLP-generated genetic maps for* T. brucei, *each derived from crosses using a different pair of parents. Informative markers were those that were polymorphic between the two parents and showed segregation into the F1.*

	map 1	map 2
cross	247 × 927	247 × 386
F1 progeny (no.)	15	17
informative markers (no.)	196	156
markers linked (%)	146 (75)	139 (89)
linkage groups (no.)	30	21
loci linked (no.)	86	74
genome coverage (cM)	424.8	513.3

The linkage maps were drawn using DRAWMAP (Van Ooijin 1994) and the maps of three linkage groups from map 2 (247 × 386) are illustrated in figure 6. The groups do not show a high level of marker clustering suggesting that the markers are relatively evenly distributed across the genome. Markers prefixed 'E' were generated using *EcoR*1/*Mse*1, which will preferentially cut in AT-rich regions, whereas markers prefixed 'H' were generated by *Hind*III/*Taq*I, which will cut in regions with a balanced base composition. The interspersed distribution of these two different classes of marker throughout the linkage groups indicates that there is no bias for mapping any AT-rich regions of the genome, further strengthening the conclusion that the linkage analysis is providing a broad coverage of the genome.

Using the data generated, it is possible to determine whether there is any bias in the inheritance of markers from either parent and the results of this analysis are illustrated in figure 7 for the five larger linkage groups from map 1. Three of the groups (1, 2 and 8) show no, or limited, segregation distortion while the other two (4 and 20) show a strong bias towards a preferential inheritance of markers from parent 927. The origins of this distortion are unknown at the present time but could reflect selection on the progeny clones for alleles at loci on these particular chromosomes. Alternatively it could be due to chance, given the relatively small number of progeny clones. If the former explanation is correct it offers a further approach to defining genes determining particular traits, as one could specifically apply selection and then identify those chromosomes showing segregation distortion.

6. GENETIC ANALYSIS OF PHENOTYPE

To date, there has been limited published analysis of the inheritance of specific phenotypes except for a preliminary analysis of HSR (Gibson & Mizen 1997). In our laboratory, we have undertaken the analysis of the inheritance of HSR and resistance to an arsenical drug (cymelarsen) as well as a preliminary analysis of infectivity for tsetse salivary glands. The results of these experiments are, as yet, incomplete and will be published in full elsewhere. The ability to determine the genetic basis of these phenotypes is dependent on their mode of inheritance. In the simplest scenario, where the phenotype is determined by alleles at a single locus and one or both parents are heterozygous for the relevant alleles, the segregation of the

phenotype in the F1 progeny would be predicted and linkage analysis is relatively straightforward provided backcross or F2 progeny are available to confirm any genetic model. In the case where both parents are homozygous but different for alleles determining a phenotype, a backcross (or F2) would be required, as no segregation will be observed in the F1. Such backcrosses have been undertaken in our own and other laboratories (Gibson *et al.* 1995) and await full analysis. More complex genetic scenarios (multiple loci and alleles) could potentially confound classical genetic approaches unless there are only one or two loci with major effects on phenotype. The initial analysis of three phenotypes (HSR, cymelarsen resistance and tsetse salivary gland infectivity) suggests that one or two loci are involved in determining each phenotype, making them amenable to a genetic approach. For the other two phenotypes available for analysis (virulence and tsetse midgut infectivity—table 2) no segregation data have been obtained as yet and so it is not possible to ascertain the feasibility of genetic analysis at the present time.

7. CONCLUSIONS AND DISCUSSION

The classical genetic analysis of phenotypic variation in *T. brucei*, coupled with linkage analysis and positional cloning, is one approach to identifying the function of genes in this parasite and is complimentary to other approaches using reverse genetics and mutational analysis. In this paper we have presented data showing that it is possible to generate a genetic map of this diploid parasite, that it is relatively straightforward to generate large numbers of F1 progeny from crosses and that the physical size of the recombination unit is sufficiently small ($11 \, \text{kb cM}^{-1}$) to allow the positional cloning of candidate genes determining a trait using a low-resolution (10 cM) map. Compared to the genetic maps generated for *Plasmodium falciparum* and *Toxoplasma*, the current *T. brucei* map is incomplete and requires the analysis of additional markers as well as anchorage to either physical maps or PFGE-separated chromosomes. Technically this is straightforward although whether further AFLP markers should be used or, given the recent availability of extensive sequence from the genome projects, micro- and minisatellite markers are preferable is largely a question of convenience. While the latter class of markers may be preferable, given that their location within the genome will be known, AFLP markers could be particularly useful when attempting to provide higher-resolution maps of specific intervals defined by linkage studies.

The availability of progeny from backcrosses, and the data showing that the larger-scale isolation of F1 progeny from existing cryopreserved products of crosses is achievable, allows the genetic map and the genetic analysis of traits to be undertaken with greater resolution and will allow more robust statistical analysis of the data. In this context it is important to note that the current genetic analysis of *P. falciparum* is based on 35 independent progeny, although methods have been developed for the large-scale isolation of further progeny (Kirkman *et al.* 1996). The two genetic maps presented here are based on a total of 32 progeny as are the data on recombination in chromosome I. The availability of a large panel of progeny would, potentially, allow the analysis of more complex
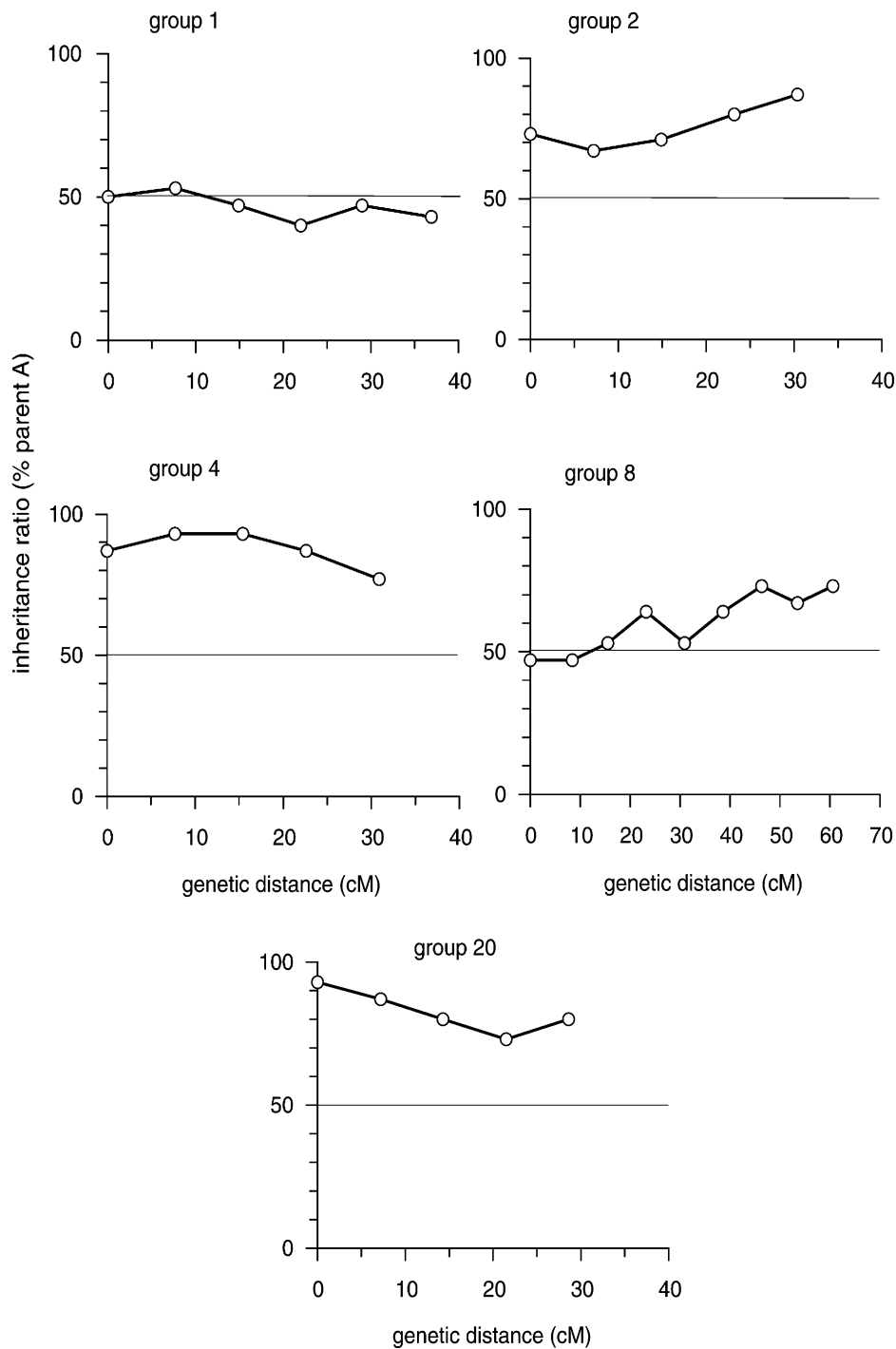
Figure 7. Segregation data for five of the larger linkage groups in the map made for the genetic cross, 247 × 927. These data show the proportion of alleles (%) inherited from each parent at each marked locus in the linkage groups. The predicted inheritance for these markers in a Mendelian system would be an equal probability of inheritance from each parent as is seen in groups 1 and 8. Severe segregation distortion is particularly evident in groups 4 and 20 and could result from selection, as discussed in the text.

traits determined by more than one locus. One of the major restraints on the use of classical genetic analysis of protozoan parasites is the complex requirements and time taken for making crosses as well as the lack of any simple *in vitro* systems. This has meant that these studies rely on a few crosses, although the restraints are less significant in the case of the rodent malaria species. In fact, the availability of a number of crosses between different parental stocks (Gibson & Stevens 1999) and the resultant progeny

makes *T. brucei* unusual compared to the other protozoan parasites in this respect. However, it should be emphasized that the only trypanosome crosses for which there are genetic maps are those described in this paper although, in principal, maps for the other crosses could be generated using the approaches described here.

In the context of the total number of genes within the protozoan parasite genomes and the need to define their function, classical genetic analysis cannot be considered

high throughput given the reliance on existing phenotypic variation for analysis. Thus, large-scale mutant screens using RNAi or gene knockouts will have a significant role to play in defining gene function. However, such systems also have restraints, particularly in relation to identifying informative phenotypes that go beyond defining genes as essential or non-essential. The advantage of classical genetic analysis is that the phenotype is available and in a number of the phenotypes being analysed, it is difficult to see how reverse genetics would directly identify the genes involved. Furthermore, with *T. brucei*, the variant phenotypes studied are highly relevant to understanding the pathogenesis and transmission of the disease.

REFERENCES

Ashton, P. D., Curwen, R. S. & Wilson, R. A. 2001 Linking proteome and genome: how to identify parasite proteins. *Trends Parasitol.* **17**, 198–202.

Barrett, M. P., MacLeod, A., Tovar, J., Sweetman, J. P., Tait, A., LePage, R. W. F. & Melville, S. E. 1997 A single locus mini satellite sequence which distinguishes between *Trypanosoma brucei* isolates. *Mol. Biochem. Parasitol.* **86**, 95–100.

Carlton, J., Mackinnon, M. & Walliker, D. 1998 A chloroquine resistance locus in the rodent malaria parasite *Plasmodium chabaudi. Mol. Biochem. Parasitol.* **93**, 57–72.

Carter, N. S. & Fairlamb, A. H. 1993 Arsenical-resistant trypanosomes lack an unusual adenosine transporter. *Nature* **361**, 173–176.

Caruthers, V. B., Van der Ploeg, L. H. & Cross, G. A. 1993 DNA-mediated transformation of bloodstream form *T. brucei. Nucleic Acids Res.* **21**, 2537.

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

Gibson, W. C. 1989 Analysis of a genetic cross between *T. brucei rhodesiense* and *T. b. brucei. Parasitology* **99**, 392–402.

Gibson, W. C. & Bailey, M. 1994 Genetic exchange in *Trypanosoma brucei* evidence for meiosis from analysis of a cross between drug-resistant transformants. *Mol. Biochem. Parasitol.* **64**, 241–252.

Gibson, W. C. & Mizen, W. H. 1997 Heritability of the trait for human infectivity in genetic crosses of *T. brucei* spp. *Trans. R. Soc. Trop. Med. Hyg.* **91**, 236–237.

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

Gibson, W. C., Garside, L. & Bailey, M. 1991 Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *T. b. rhodesiense* and *T. b. brucei. Mol. Biochem. Parasitol.* **45**, 77–90.

Gibson, W. C., Kanmogne, G. & Bailey, M. 1995 A successful backcross in *Trypanosoma brucei. Mol. Biochem. Parasitol.* **69**, 101–110.

Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A. & Turner, C. M. R. 1999 Analysis of ploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange. *Mol. Biochem. Parasitol.* **104**, 1–9.

Jenni, L., Marti, S., Scweizer, J., Betschart, B., LePage, R. W. F., Wells, J. M., Tait, A., Paindavoine, P., Pays, E. & Steinert, M. 1986 Hybrid formation between African trypanosomes during cyclical transmission. *Nature* **322**, 173–175.

Kirkman, L. A., Su, X. & Wellems, T. E. 1996 *Plasmodium falciparum*: isolation of large numbers of parasite clones from infected blood samples. *Exp. Parasitol.* **83**, 147–149.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.

Lindegard, L. A. G. 1999 Human serum resistance in *T. brucei.* PhD thesis, University of Glasgow.

MacLeod, A. 1999 Genetic analysis of *T. brucei.* PhD thesis, University of Glasgow.

MacLeod, A., Turner, C. M. R. & Tait, A. 1999 A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hyper variable 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 *T. brucei*: reconciliation of clonal panmictic and epidemic population genetic structures. *Proc. Natl Acad. Sci. USA* **97**, 13 442–13 447.

Masiga, D., Tait, A. & Turner, C. M. R. 2000 Amplified restriction length polymorphism in parasite genetics. *Parasitol. Today* **16**, 350–353.

Melville, S. E., Leech, V., Gerrard, C. S., Tait, A. & Blackwell, J. M. 1998 The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol. Biochem. Parasitol.* **94**, 155–173.

Melville, S. E., Gerrard, C. S. & Blackwell, J. M. 1999 Multiple causes of size variation in the diploid megabase chromosomes of African trypanosomes. *Chromosome Res.* **7**, 191–203.

Ngo, H., Tschudi, C., Gull, K. & Ullu, E. 1998 Double-stranded RNA induces mRNA degradation in *T. brucei. Proc. Natl Acad. Sci. USA* **95**, 14 687–14 692.

Pandey, A. & Mann, M. 2000 Proteomics to study genes and genomes. *Nature* **405**, 837–846.

Sasse, J. 1998 The development of genetic markers for the *T. brucei* genome. PhD thesis, University of Cambridge.

Savelkoul, P. H. M., Aarts, H. J. M., De Hass, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J. L. W., Schouls, L. & Lenstra, J. A. 1999 Amplified-fragment length polymorphism analysis: the state of an art. *J. Clin. Microbiol.* **37**, 3083–3091.

Schweizer, J., Tait, A. & Jenni, L. 1988 The timing and frequency of hybrid formation in African trypanosomes during cyclical transmission. *Parasitol. Res.* **75**, 98–101.

Scott, A. G., Tait, A. & Turner, C. M. R. 1996 Characterisation of cloned lines of *Trypanosoma brucei* expressing stable resistance to MelCy and Suramin. *Acta Tropica* **60**, 251–262.

Shi, H. A., Dijkeng, A., Mark, T., Wirtz, E., Tschudi, C. & Ullu, E. 2000 Genetic interference in *T. brucei* by heritable and inducible double stranded RNA. *RNA* **6**, 1069–1076.

Shirley, M. W. & Harvey, D. A. 2000 A genetic linkage map of the apicomplexan protozoan parasite *Eimeria tenella. Genome Res.* **10**, 1587–1593.

Sibley, L. D., LeBlanc, A. J., Pfefferkorn, E. R. & Boothroyd, J. C. 1992 Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii. Genetics* **132**, 1003–1015.

Sternberg, J. & Tait, A. 1990 Genetic exchange in African trypanosomes. *Trends Genet.* **6**, 317–322.

Sternberg, J., Turner, C. M. R., Wells, J. M., Ranford-Cartwright, L. C., LePage, R. W. F. & Tait, A. 1989 Gene exchange in African trypanosomes—frequency and allelic segregation. *Mol. Biochem. Parasitol.* **34**, 269–280.

Su, X. & Wellems, T. E. 1996 Towards a high resolution *Plas-*

modium falciparum linkage map: polymorphic markers from hundreds of simple sequence repeat. *Genomics* **33**, 430–444.

Su, X., Ferdig, M. T., Huang, Y., Huynh, C. Q., Liu, A., You, J., Wootton, J. C. & Wellems, T. E. 1999 A comprehensive genetic map of the human malaria parasite *Plasmodium falciparum. Science* **286**, 1351–1353.

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

Tait, A., Buchanan, N., Hide, G. & Turner, C. M. R. 1993 Genetic recombination and karyotype inheritance. In *Genome analysis of protozoan parasites* (ed. S. P. Mozaria), pp. 93–108. Nairobi: ILRAD Press.

Tait, A., Buchanan, N., Hide, G. & Turner, C. M. R. 1996 Self-fertilisation in *T. brucei. Mol. Biochem. Parasitol.* **76**, 31–42.

Tasker, M., Wilson, J., Sarkar, M., Hendriks, E. & Mathews, K. 2000 A novel selection regime for differentiation defects demonstrates an essential role for the stumpy form in the life cycle of the African trypanosome. *Mol. Biol. Cell* **11**, 1905–1917.

Turner, C. M. R., Sternberg, J., Buchanan, N., Smith, E., Hide, G. & Tait, A. 1990 Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitology* **101**, 377–386.

Turner, C. M. R., Aslam, N. & Dye, C. 1995 Replication, differentiation, growth and the virulence of *Trypanosoma brucei* infections. *Parasitology* **111**, 289–300.

Van Ooijen, J. W. 1994 DRAWMAP: a computer program for drawing genetic linkage maps. *J. Heredity* **85**, 66–69.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. 1995 AFLP: a new technique for DNA finger-printing. *Nucleic Acids Res.* **23**, 4407–4414.

Walker-Jonah, A., Dolan, S. A., Gwadz, R. W., Panton, L. J. & Wellems, T. E. 1992 An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favoured linkage groups in a genetic cross. *Mol. Biochem. Parasitol.* **51**, 313–320.

Welburn, S. C., Maudlin, I. & Milligan, P. J. M. 1995 Trypanozoon: infectivity to humans is linked to reduced transmissibility in tsetse I. Comparison of human serum resistant and human serum sensitive field isolates. *Exp. Parasitol.* **81**, 404–408.