The genetic map and comparative analysis with the physical map of *Trypanosoma brucei*

Annette MacLeod[1,*], Alison Tweedie[1], Sarah McLellan[2], Sonya Taylor[2], Neil Hall[3], Matthew Berriman[3], Najib M. El-Sayed[4], Michelle Hope[1], C. Michael R. Turner[1,2] and Andy Tait[1]

[1]Wellcome Centre for Molecular Parasitology, Anderson College Complex, 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 G12 8QQ, UK, [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA and [4]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

ABSTRACT

***Trypanosoma brucei* is the causative agent of African sleeping sickness in humans and contributes to the debilitating disease 'Nagana' in cattle. To date we know little about the genes that determine drug resistance, host specificity, pathogenesis and virulence in these parasites. The availability of the complete genome sequence and the ability of the parasite to undergo genetic exchange have allowed genetic investigations into this parasite and here we report the first genetic map of *T.brucei* for the genome reference stock TREU 927, comprising of 182 markers and 11 major linkage groups, that correspond to the 11 previously identified chromosomes. The genetic map provides 90% probability of a marker being 11 cM from any given locus. Its comparison to the available physical map has revealed the average physical size of a recombination unit to be 15.6 Kb/cM. The genetic map coupled with the genome sequence and the ability to undertake crosses presents a new approach to identifying genes relevant to the disease and its prevention in this important pathogen through forward genetic analysis and positional cloning.**

INTRODUCTION

Trypanosoma brucei is a diploid zoonotic protozoan parasite transmitted by tsetse flies. This species has been further subdivided into three morphologically identical subspecies, *T.b. gambiense* and *T.b. rhodesiense* cause sleeping sickness in humans whereas the third subspecies, *T.b. brucei*, is not infective to humans. *T.brucei* also infects cattle and is one of three trypanosome species that causes the economically important disease 'Nagana' in sub-Saharan Africa. There is significant variation both between and within the subspecies in a range of important phenotypes, such as drug resistance and virulence. Identifying genes involved in these phenotypes would be a considerable advance in the study of this important pathogen.

The development of a genetic map for *T.brucei* is crucial to our understanding of the genetic system in this pathogen and opens up the possibility of using forward genetic analysis as a tool to identify genes that determine traits of importance in the transmission, treatment and pathogenesis of the disease (1). Genetic analysis in other parasitic protozoa has been central to identifying the genes and loci that determine drug resistance (2–6) and virulence (7,8).

As no chromosome condensation has been observed in any life cycle stage and no gamete stages identified, the main approach in determining whether *T.brucei* has a sexual cycle and undergoes meiosis has been to undertake classical genetic analysis. Previous work (9–11) has shown that when the tsetse fly vector is co-infected with two genetically different lines of the parasite, the resultant parasites comprise a mixture of the original two parental lines together with hybrids that, by marker analysis, are the equivalent of F1 progeny (12). Recently, further marker analysis on a large number of progeny clones from two genetic crosses, between two *T.b. brucei* strains and between a *T.b. brucei* and a *T.b. gambiense* strain (STIB 247 × TREU 927 and STIB 247 × STIB 386,

respectively) (13) provided formal statistically significant proof of conventional Mendelian inheritance involving meiosis and sygamy. Previous data from other crosses between subspecies (14,15) have shown that a high proportion of progeny were triploid leading to models of genetic exchange involving diploid cell fusion and chromosome loss. In the crosses reported here, either no triploid progeny (247 × 927) or very few (247 × 386) have been found (16). The results presented are consistent with two models of genetic exchange: (i) meiosis to generate haploid gametes which fuse to generate diploid progeny or (ii) fusion of diploid cells followed by meiosis and cell division, also resulting in diploid progeny. These models are both consistent with the observed Mendelian patterns of inheritance and contrast with the diploid fusion followed by random chromosome loss model proposed in some of the earliest studies of *T.brucei* genetics (14,15) and in the related kinetoplastid, *Trypanosoma cruzi* (17). These results mean that *T.brucei* is amenable to genetic mapping, linkage analysis and positional cloning. Here we describe the generation of the first genetic map for *T.brucei*.

MATERIALS AND METHODS

Generation of hybrids

A number of independent F1 progeny clones had been generated previously (12,13) from a genetic cross between an isolate from a tsetse fly (TREU 927) and one from a hartebeast (STIB 247). In total 39 F1 progeny of independent mating events were generated from the STIB 247 × TREU 927 cross, which were screened for new informative microsatellite markers spanning the genome. Complete lists of hybrid progeny from this cross is given in ref (13).

Identification of markers and the screening of hybrids

Tandem repeat finder (18) (http://tandem.bu.edu/trf/trf.html, Tandem repeat finder 3.21) was used to screen the available sequence data (19) to identify microsatellites and PCR primers were designed to the sequence flanking each locus, using the primer design program, PRIDE (http://pride.molgen.mpg.de/pride.html, PRIDE 1.2). The primers were used to screen the parental stocks for loci that were heterozygous for allele size differences. To do this, microsatellite markers were amplified from genomic DNA, under the following conditions: 95°C for 50 s, 50°C for 50 s and 65°C for 50 s for 30 cycles, using primer concentrations and the PCR buffer described elsewhere (20). The products were separated on 3% Nusieve (flowgen) agarose gels and visualized under UV. Minisatellite markers, including *MS42*, *292* and *CRAM*, were analysed as described previously (20).

Construction of the genetic map

The map has been based on segregation of alleles in the F1 progeny for loci heterozygous in the TREU 927 parent only. This is because the parental stock STIB 247 was found to be homozygous for 94% of markers. The segregation data for all heterozygous loci from TREU 927 are given in Supplementary Tables 1–11 and http://www.gla.ac.uk/centres/wcmp/research/macLeod/gmpnov.html, (Genetic map of *T.brucei*) and used to construct the genetic map of TREU 927 using

the program MapManager (21) with a Haldane map function, at the highest level of significance for linkage criteria, with a probability of type 1 error $P = 1\,\mathrm{e}^{-6}$. Each marker was linked to the adjacent marker with a LOD score of 5.5 or greater.

RESULTS AND DISCUSSION

The 25 Mb genome of *T.brucei* has been sequenced on a chromosome by chromosome basis by two sequencing centres, The Wellcome Trust Sanger Institute and The Institute for Genome Research (TIGR) and covers all the 11 megabase chromosomes (19,22,23). From the available sequence data we selected 810 mini- and microsatellite loci, which consisted of more than 12 repeat units and were approximately evenly distributed along each of the 11 megabase chromosomes, excluding the subtelomeric/telomeric regions. As *T.brucei* is diploid, heterozygous markers were identified by the PCR amplification of two alleles differing in size. TREU 927 was heterozygous for approximately one fifth of the mini- and microsatellite loci tested. The lack of heterozygosity in STIB 247 is consistent with previous AFLP marker analysis (1) and precludes the construction of a high resolution genetic map of this stock, thus the genetic map was constructed for the genome reference stock TREU 927. All 39 independent F1 progeny from the STIB 247 × TREU 927 cross (13) were genotyped for 182 informative markers, (Supplementary Tables 1–11). The inheritance patterns of these markers in the progeny of the STIB 247 × TREU 927 cross were then used to construct the genetic map of *T.brucei*. The markers fall into 16 linkage groups with the 11 large groups corresponding to the 11 megabase chromosomes of *T.brucei* (Figure 1). All markers, except 3, have been assigned to physical locations. Chromosomes, VII, IX, X and XI each comprise one major linkage group with either one or two apparently unlinked small linkage groups (Figure 1).

By comparing the physical and genetic maps of each chromosome (e.g. see Figure 2), it is clear that the order of markers for all linkage groups is congruent with the physical map (Supplementary Figures 1–11), with only one exception, TB11/12. The relative sizes of the linkage groups measured in centiMorgans (cM) correlates with the corresponding physical length of the regions of the chromosomes represented by the genetic map (Figure 3). The physical maps for chromosomes I to X are single contigs, whereas chromosome XI consists of 3 contigs, which are linked by the genetic map. The genetic map does not extend into the telomeric and subtelomeric regions of each chromosome due to the lack of full genomic sequence and the difficulties in identifying unique sequences in these regions, neither do the physical and genetic maps cover the mini or intermediate chromosomes, which contain only variant surface glycoprotein (VSG) genes and bloodstream expression sites (24). The genetic map spans 1157.5 cM in total, comparable with that of *Plasmodium falciparum*, at 1556 cM (25). There are 122 map intervals, with an average physical distance of an interval being 148 kb, approximately equivalent in size to a BAC clone insert. The marker coverage of the genome is such that there is a 90% probability that at least one marker is within ~11 cM of a given locus.
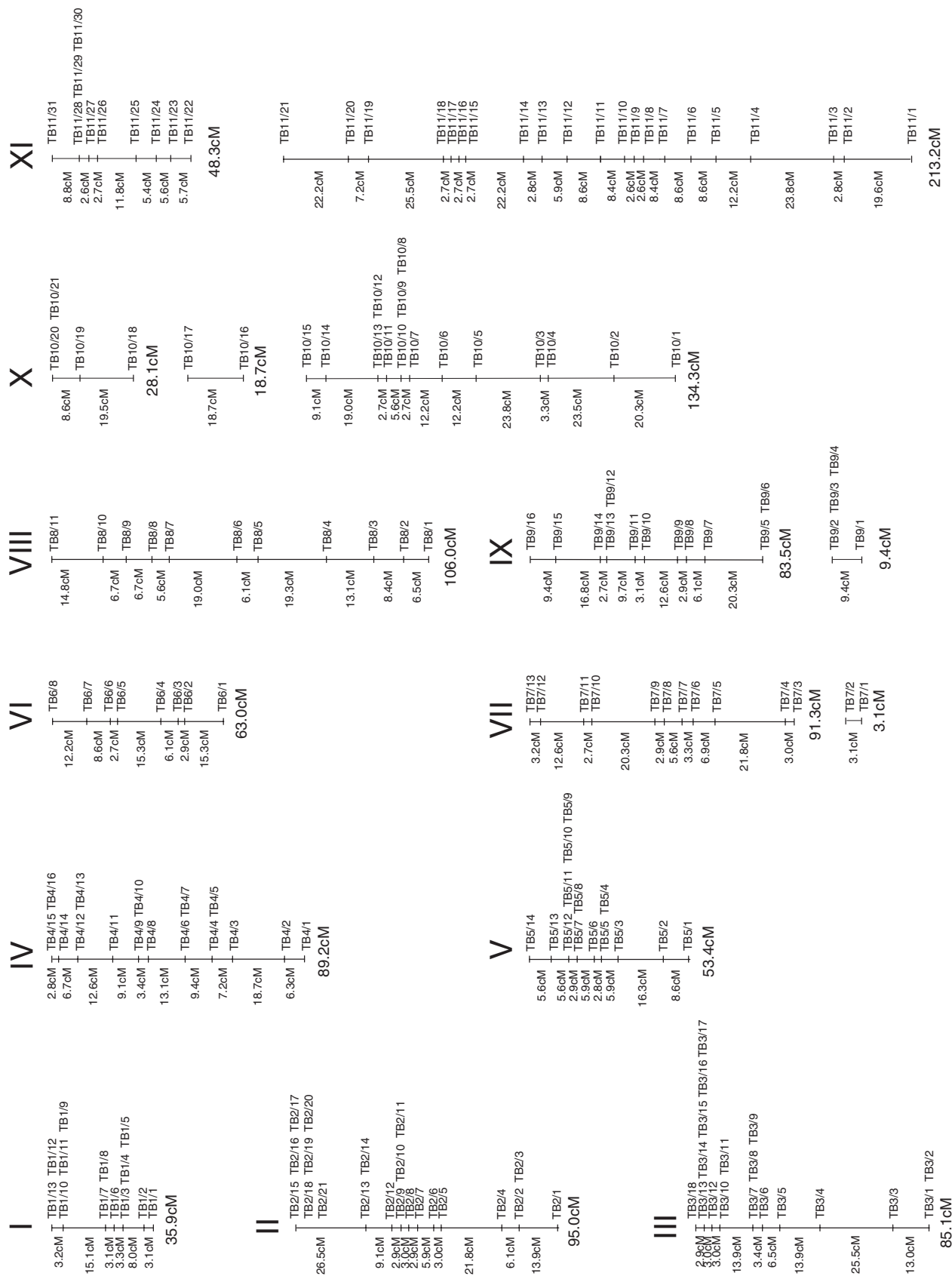
Figure 1. Linkage maps corresponding to the 11 megabase chromosomes (I–XI) of *T.brucei*. All mini- and microsatellite markers (to the right in each map) have been physically assigned to chromosomes except three. Their positions on each chromosome are identified in Supplementary Tables 1–11. The genetic distance between each marker is given in centiMorgans, Haldane corrected. The genetic size of the linkage groups is given below each linkage group.
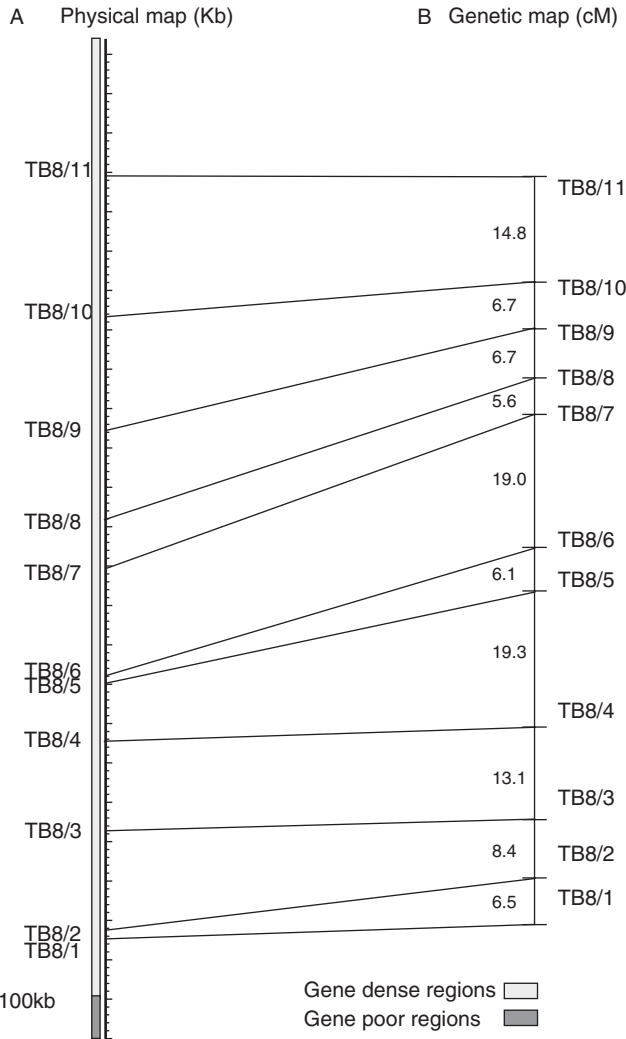
Figure 2. (**A**) The physical map of chromosome VIII of *T.b. brucei* (2.48 Mb) and (**B**) the genetic map of the same chromosome (106.2 cM). The genetic distances between markers are given in cM. Dashed lines indicate the position of all markers on the physical map. The scale bar represents 100 kb. House-keeping genes are contained within the gene dense region, which is marked with a light grey box and gene poor regions are indicated with a dark grey box. Gene poor regions are subtelomeric regions, which mainly compose of repetitive sequences, pseudogenes and gene families.

The average map unit distance is 15.6 kb/cM, based on the physical distances between genetic markers, but there is considerable variation in the physical size of the centiMorgan within and between chromosomes (Supplementary Figures 1–11). Previously hot and cold spots of recombination had been identified on Chromosome I and II (22,23) and the data presented here show variation in crossover frequency to be a common feature of each of the remaining 9 chromosomes. For example, there is a region of chromosome VII where the physical size of the recombination unit is 1.58 kb/cM and conversely, on chromosome XI, where the recombination unit is 95.64 kb/cM. Hot and cold spots have been observed in many eukaryotes (26), but surprisingly not in *P.falciparum* which has an apparently uniform meiotic crossover activity per physical distance (25). Although local variations in the physical size of a recombination unit contradict the
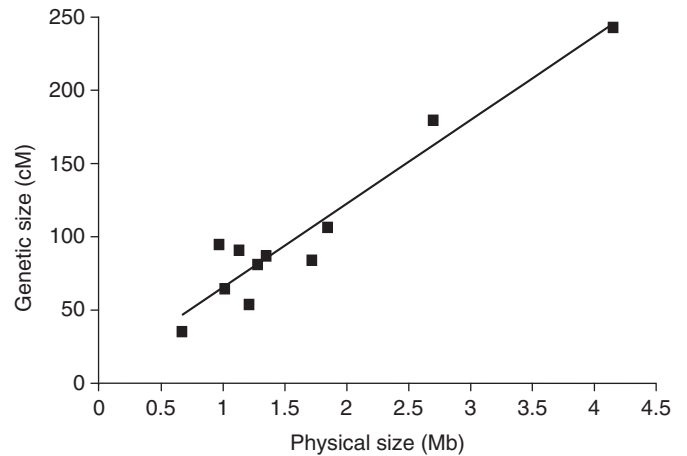

Figure 3. A comparison of the physical lengths of each chromosome versus the relative sizes of the corresponding linkage groups. The line shown was determined by linear least squares regression analysis, $R^2 = 0.92$.

correlation between physical and genetic length, this appears to average out over the full length of each chromosome (Figure 3).

The *T.brucei* genome contains many copies of VSG sequences, which have been previously associated with mitotic recombination resulting in antigenic variation (27). The genome sequence of *T.brucei* had revealed several regions of tandemly repeated VSG sequences, which mainly consist of pseudogenes (19). It has been suggested that the vast number of these pseudogenes have arisen through hybrid VSG gene formation via mitotic recombination, but the same result could also theoretically arise from meiotic recombination. In order to investigate if these arrays of VSG sequences are associated with meiotic recombination, we examined one of the largest VSG pseudogene arrays in the genome, on chromosome IX, (position 2476002–3055206 bp), by comparing the crossover frequencies in the region of the genetic map covering the VSG array with the map for the rest of the chromosome. The physical size of a recombination unit in the VSG array region is 50.2 kb/cM (measured between positions 2455704 and 2927791 bp), which is markedly larger than the average for the rest of chromosome IX of 15 kb/cM indicating that, surprisingly, this is a region of low meiotic recombination activity, with 3 crossovers identified whereas the average for this chromosome, would predict 12 crossovers ($\chi^2_{1\text{d.f.}}$ 9.3, $P < 0.05$) the null hypothesis being equivalent crossover frequency in both regions of the chromosome. Whether low meiotic recombination frequency is a feature of all VSG arrays has not been determined as yet due to the inherent difficulties in identifying unique sequences/markers within these regions.

The vast majority of markers demonstrate the inheritance of equal numbers of parental alleles in the progeny, in agreement with Mendelian predictions (13). However, there are two regions of the genetic map (on chromosomes III and IX) where there is segregation distortion, i.e. where one haplotype is inherited more frequently in the progeny than predicted (Figure 4). The origin of this distortion is unknown at the present time but could reflect selection acting on the uncloned progeny populations before cloning for alleles at loci in these regions of the chromosomes.
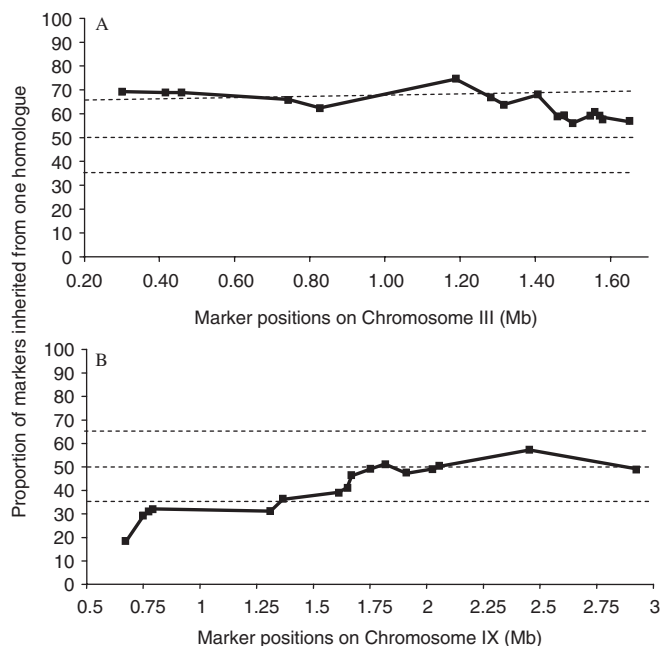
Figure 4. Genotype segregation proportions for markers on chromosome (**A**) III and (**B**) IX. Dashed horizontal lines delimit the approximate 95% probability range for equal segregation of alleles.

From the segregation analysis of markers, one progeny clone, F532/53 mcl 1, appears to be trisomic for chromosome I, having inherited both alleles from parental stock TREU 927 for all chromosome I markers analysed. Analysis of the inheritance of markers on other chromosomes, however, indicates that this hybrid has inherited only one homologue from each parental stock, clearly demonstrating that this progeny clone is trisomic for chromosome I, but not triploid. This is the first case of trisomy as opposed to triploidy reported in *T.brucei* and probably arose due to chromosomal non-disjunction at meiosis. The frequency of trisomy in this genetic cross is 2.5% and, while this clone cannot be used for mapping purposes for chromosome I, it is informative for all other chromosomes and so was included in the panel of informative hybrids.

Thirty-nine progeny clones typed for the 182 markers typed has generated a dataset of 6797 scored alleles (Supplementary Tables 1–11), and has identified two spontaneous mutation events, generating novel sized alleles distinct from parental alleles. These mutation events occurred at loci TB9/9 and TB9/13, giving a mutation frequency for each locus of 0.027 mutants/alleles genotyped. This generates an overall mutation frequency for all markers combined of 0.0003 mutants/alleles genotyped. These finding raises the question of the origin of the mutant allele. As these two mutant alleles replaced the original parental allele they could not have arisen during vegetative growth of the progeny clone but arose before the cloning process, possibly at meiosis. In both cases the mutant alleles had gained repeats generating alleles larger than the parental alleles and for mapping purposes it was assumed the parental allele closest in size to the mutant allele was the progenitor allele.

The high resolution genetic map we have generated for *T.brucei* for this pathogen has opened up the possibility of identifying genes that determine traits of importance by genetic analysis and positional cloning (1). The next challenge is to exploit this new toolset to understand better traits such as human infectivity (28) drug resistance or virulence (1) that all contribute to the severity of sleeping sickness and Nagana.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Tait,A., Masiga,D., Ouma,J., MacLeod,A., Sasse,J., Melville,S., Lindegard,G., McIntosh,A. and Turner,C.M.R. (2002) Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits. *Philos. Trans. R. Soc. Lond., B*, **357**, 89–99.
2. Su,X., Kirkman,L.A., Fujioka,H. and Wellems,T.E. (1997) Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell*, **91**, 593–603.
3. Ferdig,M.T., Cooper,R.A., Mu,J., Deng,B., Joy,D.A., Su,X.Z. and Wellems,T.E. (2004) Dissecting the loci of low-level quinine resistance in malaria parasites. *Mol. Microbiol.*, **52**, 985–997.
4. Culleton,R., Martinelli,A., Hunt,P. and Carter,R. (2005) Linkage group selection: rapid gene discovery in malaria parasites. *Genome Res.*, **15**, 92–97.
5. Fidock,D.A., Nomura,T., Talley,A.K., Cooper,R.A., Dzekunov,S.M., Ferdig,M.T., Ursos,L.M., Sidhu,A.B., Naude,B., Deitsch,K.W. *et al.* (2000) Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell.*, **6**, 861–871.
6. Sidhu,A.B., Verdier-Pinard,D. and Fidock,D.A. (2002) Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcrt mutations. *Science*, **298**, 210–213.
7. Su,C., Howe,D.K., Dubey,J.P., Ajioka,J.W. and Sibley,L.D. (2002) Identification of quantitative trait loci controlling acute virulence in *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA*, **99**, 10753.
8. Martinelli,A., Cheesman,S., Hunt,P., Culleton,R., Raza,A., Mackinnon,M. and Carter,R. (2005) A genetic approach to the *de novo* identification of targets of strain-specific immunity in malaria parasites. *Proc. Natl Acad. Sci. USA*, **102**, 814–819.
9. Jenni,L., Marti,S., Schweizer,J., Betschart,B., Le Page,R.W., Wells,J.M., Tait,A., Paindavoine,P., Pays,E. and Steinert,M. (1986) Hybrid formation between African trypanosomes during cyclical transmission. *Nature*, **322**, 173–175.
10. Sternberg,J., Turner,C.M., Wells,J.M., Ranford-Cartwright,L.C., Le Page,R.W. and Tait,A. (1989) Gene exchange in African trypanosomes: frequency and allelic segregation. *Mol. Biochem. Parasitol.*, **34**, 269–279.
11. Gibson,W.C. (1989) Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Parasitology*, **99**, 391–402.
12. Turner,C.M., Sternberg,J., Buchanan,N., Smith,E., Hide,G. and Tait,A. (1990) Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitology*, **101**, 377–386.

13. MacLeod,A., Tweedie,A., McLellan,S., Taylor,S., Cooper,A., Sweeney,L., Turner,C.M.R. and Tait,A. (2005) Allelic segregation and independent assortment in *Trypanosoma brucei* crosses: proof that the genetic system is Mendelian and involves meiosis. *Mol. Biochem. Parasitol.*, **143**, 12–19.

14. Gibson,W. (1995) The significance of genetic exchange in trypanosomes. *Parasitol. Today*, **11**, 465–468.

15. Gibson,W. and Stevens,J. (1999) Genetic exchange in the trypanosomatidae. *Adv. Parasitol.*, **43**, 1–46.

16. Hope,M., MacLeod,A., Leech,V., Melville,S., Sassa,J., Tait,A. and Turner,C.M.R. (1999) Maintenance of diploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange. *Mol. Biochem. Parasitol.*, **104**, 1–9.

17. Gaunt,M.W., Yeo,M., Frame,I.A., Stothard,J.R., Carrasco,H.J., Taylor,M.C., Mena,S.S., Veazey,P., Miles,G.A., Acosta,N. *et al.* (2003) Mechanism of genetic exchange in American trypanosomes. *Nature*, **421**, 936–939.

18. Benson,G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic. Acids Res.*, **27**, 573–580.

19. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E., Haas,B. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**, 416–422.

20. MacLeod,A., Turner,C.M.R. and 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.

21. Manly,K.F., Cudmore,J.R.H. and Meer,J.M. (2001) Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome*, **12**, 930–932.

22. Hall,N., Berriman,M., Lennard,N.J., Harris,B.R., Bart-Delabesse,E.N., Hertz-Fowler,C., Gerrard,C.S., Atkin,R.J., Barron,A.J., Bowman,S. *et al.* (2003) The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. *Nucleic. Acids Res.*, **31**, 4864–4873.

23. El-Sayed,N.M.A., Ghedin,E., Song,J., MacLeod,A., Bringaud,F., Larkin,C., Wanless,D., Peterson,J., Hou,L., Taylor,S. *et al.* (2003) The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic. Acids Res.*, **31**, 4856–4863.

24. El-Sayed,N.M., Hegde,P., Quackenbush,J., Melville,S.E. and Donelson,J.E. (2000) The African trypanosome genome. *Int. J. Parasitol.*, **30**, 329–345.

25. Su,X.-Z., Ferdig,M.T., Huang,Y., Huynh,C.Q., Lui,A., You,J., Wootton,J.C. and Wellems,T.E. (1999) A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum. Science*, **286**, 1351–1353.

26. Petes,T.D. (2001) Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.*, **2**, 360–369.

27. Barry,J.D. and McCulloch,R. (2001) Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv. Parasitol.*, **49**, 1–70.

28. Turner,C.M., McLellan,S., Lindergard,L.A., Bisoni,L., Tait,A. and MacLeod,A. (2004) Human infectivity trait in *Trypanosoma brucei*: stability, heritability and relationship to sra expression. *Parasitology*, **129**, 445–454.