Whole-Genome Sequencing of *Trypanosoma brucei* Reveals Introgression between Subspecies That Is Associated with Virulence

Ian Goodhead, a Paul Capewell, b J. Wendi Bailey, c Tanja Beament, c Michael Chance, c Suzanne Kay, a Sarah Forrester, a Annette MacLeod, b Mark Taylor, a Harry Noyes, a Neil Hall a

Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom; Wellcome Centre for Molecular Parasitology, Institute of Biodiversity Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom; Molecular and Biochemical Parasitology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

ABSTRACT  Human African trypanosomiasis is caused by two subspecies of *Trypanosoma brucei*. *Trypanosoma brucei rhodesiense* is found in East Africa and frequently causes acute disease, while *Trypanosoma brucei gambiense* is found in West Africa and is associated with chronic disease. Samples taken from a single focus of a Ugandan outbreak of *T. b. rhodesiense* in the 1980s were associated with either chronic or acute disease. We sequenced the whole genomes of two of these isolates, which showed that they are genetically distinct from each other. Analysis of single nucleotide polymorphism markers in a panel of 31 Ugandan isolates plus 32 controls revealed a mixture of East African and West African haplotypes, and some of these haplotypes were associated with different virulence phenotypes. It has been shown recently that *T. b. brucei* and *T. b. rhodesiense* populations undergo genetic exchange in natural populations. Our analysis showed that these strains from the Ugandan epidemic were intermediate between the reference genome sequences of *T. b. gambiense* and *T. b. brucei* and contained haplotypes that were present in both subspecies. This suggests that the human-infective subspecies of *T. brucei* are not genetically isolated, and our data are consistent with genomic introgression between East African and West African *T. b. brucei* subspecies. This has implications for the control of the parasite, the spread of drug resistance, and understanding the variation in virulence and the emergence of human infectivity.

IMPORTANCE  We present a genetic study of the acute form of “sleeping sickness” caused by the protozoan parasite *Trypanosoma brucei rhodesiense* from a single outbreak in Uganda. This represents an advance in our understanding of the relationship between the *T. b. rhodesiense* and *Trypanosoma brucei gambiense* subspecies that have previously been considered genetically distinct. Our data suggest that introgression of West African-derived *T. brucei* haplotypes may be associated with differences in disease presentation in the East African disease. These findings are not only of scientific interest but also important for parasite control, as they suggest that the human-infective *T. brucei* subspecies are not genetically isolated.
tions of both acute (10) and asymptomatic (11) *T. b. gambiense* infections exist from Côte d’Ivoire. In *T. b. rhodesiense* there have been asymptomatic carriers in Botswana (12), mild disease in Zambia and Malawi (13, 14), and reports of severe, acute disease in Uganda (15). In the last case, the isolates collected from a 1989 outbreak in southeastern Uganda displayed a correlation between disease profile and the multilocus enzyme electrophoresis (MLEE) strain group, suggesting that genetic variation may underlie the differences in observed virulence. The most prevalent zymodemes from within these strain groups were Busoga 17 (B17) and Zambezi 310 (Z310). Z310 isolates were associated with a more chronic infection than B17 isolates, and patients were often unaware of being infected due to a lack of a chancre at the site of a tsetse bite. Patients infected with Z310 parasites often presented at clinics with the more serious late-stage disease. B17 patients often presented earlier in the course of infection, especially as chancres were often present and patients had learned to associate these with *T. b. rhodesiense* infections. Those patients that had been observed with late-stage B17 infections had progressed to this stage rapidly, with severe symptoms.

We have used high-throughput sequencing to identify genome-wide single nucleotide polymorphisms (SNPs) in the genomes of one Zambezi (Z310) strain and one Busoga (B17) strain in order to better understand which genetic loci may contribute to the observed differences in virulence from within a single localized outbreak. We have used SNPs discovered by sequencing to further genotype 31 isolates from Uganda, together with additional controls. We present the first genomic sequence data for the East African form of the disease, which can now be compared to the available *T. b. brucei* and *T. b. gambiense* sequences. Furthermore, we have compared the genome sequences of the three subspecies, which revealed evidence of genetic exchange between the East and West African populations. We replicated the human clinical phenotype in mouse models. Our data suggest that certain patterns of introgression may be linked with increased-virulence phenotypes.

**RESULTS**

**Microsatellite analysis cannot separate the Zambezi and Busoga zymodeme strain groups.** Samples originating from the Ugandan outbreak of HAT from 1989 to 1993, for which clinical presentations have previously been described (15) (see Table S1 in the supplemental material), were taken. Thirty-one strains belonging to nine known zymodemes of *T. b. rhodesiense* were examined at 11 informative microsatellite loci (Table S2). Clustering the genotype data using a neighbor-joining (NJ) method based on Jaccard’s similarity index revealed few branches with strong bootstrap support. There were three major distinct groups of individuals, including a single group containing Z366 isolates as well as a B376 isolate, Z377 isolate, and an additional isolate of unknown zymodeme, which clustered separately (Fig. 1).

**Whole-genome sequencing reveals regions of heterozygosity that correspond with shared alleles between *T. b. gambiense* and *T. b. rhodesiense*.** After sequencing by oligonucleotide ligation and detection (SOLiD), we aligned reads for B17 and Z310 isolates with that of the *T. b. brucei* TREU927 reference strain (Table 1). Comparing SNPs of the sequenced *T. b. rhodesiense* isolates to both *T. b. brucei* TREU927 and type 1 *T. b. gambiense* (DAL972) showed regions within chromosomes 2, 3, 5, 8, and 10 in which heterozygosity in one of the sequenced *T. b. rhodesiense* isolates was associated with alleles shared with the *T. b. gambiense* reference (Fig. 2). The sizes of these regions vary; however, the most striking findings are that ~73% of chromosome 8 of B17 is heterozygous for alleles of *T. b. brucei* TREU927 and *T. b. gambiense* DAL972 and that Z310 is homozygous and similar to *T. b. brucei* TREU927 in the same region, as indicated by a block of blue, representing only one shared allele with the respective reference sequence. Sequence comparison of B17 and Z310 shows that
Experimental infections were carried out for three isolates of each zymodeme in outbred CD-1 mice. Parasitism levels and the overall condition and behavior of the mouse were monitored. Mice infected with B17 isolates survived for longer periods than mice infected with a Z310 isolate despite a higher first peak of parasitemia at 3 to 5 days postinfection (Fig. 4A). B17 isolate-infected mice did not appear ill and never needed to be killed prior to the endpoint of the experiment (38 days postinfection). Z310 parasite-infected CD-1 mice showed a generally higher, more variable level of parasitemia (with a lower initial peak) and showed more-severe symptoms associated with infection; 72% of these mice were humanely killed before the end of the experiment due to deterioration in health. Z310 isolate-infected mice had a significantly higher parasitemia at the time of death than B17 isolate-infected mice (Mann-Whitney, P < 0.001). Mice infected with Z310 parasites lived for a significantly shorter length of time than mice infected with B17 trypanosomes (P < 0.01).

### DISCUSSION

#### Possible origins of the B17 and Z310 genotypes

Uganda is unique in that populations of all three subspecies of *T. brucei* are present in the country. The known ranges of *T. b. gambiense* and *T. b. rhodesiense* strains do not overlap (18); however, it is still possible that recombination occurs between subspecies where the ranges of the animal reservoirs of the parasite overlap (e.g., in the underlying *T. b. brucei* population). It has been reported recently that microsatellite markers are consistent with recombination between *T. b. brucei* and *T. b. rhodesiense* (19), yet there have been no reports of recombination in the field between *T. b. gambiense* and either of the other subspecies. Our SNP markers suggest that West African *T. b. brucei* is more similar to *T. b. gambiense* than to East African *T. b. brucei*. The B17 and Z310 isolates therefore appear to be the products of recombination between East and West African parasites that have expanded clonally in this particular epidemic. Whether this recombination was directly between *T. b. rhodesiense* and *T. b. gambiense* or between *T. b. rhodesiense* and West African *T. b. brucei* cannot be determined from our data. Alternatively, it may be that recombination, followed by losses of heterozygosity at multiple loci, has occurred to generate the different clonal groups.

Other protozoan parasites undergo recombination to produce new, differentially virulent outbreaks; recombination between two ancestral lineages of *Toxoplasma gondii* has resulted in a pandemic outbreak and the currently circulating clonal genotypes. Furthermore, some progeny from laboratory crosses between these lineages were more virulent than the parental lines (20). It remains to be seen whether other species of African trypanosomes also undergo similar processes, although differential virulence is observed within and between different genetic groups of *Trypanosoma congolense* (21). Certainly, the ability for genetic exchange to

---

**TABLE 1 SOLiD sequencing metrics**

<table>
<thead>
<tr>
<th>Zymodeme</th>
<th>Total no. of sequencing reads (thousands)</th>
<th>Mean coverage at SNP loci</th>
<th>No. of SNPs vs <em>T. b. brucei</em> TREU927/4</th>
<th>No. of SNPs vs alternative <em>T. b. rhodesiense</em> isolate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z310</td>
<td>58.49</td>
<td>53X</td>
<td>132,389</td>
<td>18,607</td>
</tr>
<tr>
<td>B17</td>
<td>61.26</td>
<td>52X</td>
<td>137,665</td>
<td>24,987</td>
</tr>
</tbody>
</table>

*Summary statistics for SOLiD sequencing of two *T. b. rhodesiense* strains. Sequencing reads were aligned to the genome and aligned to the ~26-Mbp *T. b. brucei* TREU927/4 reference sequence. Mean coverage at SNP loci was determined by the SAMtools (32) Pileup package.*
spread phenotypes, such as virulence or drug resistance, throughout a population means that the situation must be monitored, particularly in the case of *T. congoense* and *T. vivax*, due to the impact of the diseases on human welfare (3).

The reassortment of alleles between different subspecies may have an impact not only on virulence but also on other phenotypes, such as host range and, arguably more importantly, drug resistance; there are increasing reports of treatment failures in the major frontline drugs used in the treatment of sleeping sickness, and resistance is relatively simple to produce in the laboratory (22). Given that, in Uganda, the two forms of HAT exist only 100 km apart (18), these data suggest that close monitoring of the circulating genotypes is necessary if the current trend of decreasing infections across Africa is to continue (1).

**Alleles on chromosome 8 that are shared with West African *T. brucei* may underlie differences in virulence between the B17 and Z310 zymodemes.** Analysis of the genome-wide SNP loci shows that the sampled B17 and Z310 genomes are extremely similar; the mean percentage difference between strains is <0.2%. A heatmap of percentage similarity between the sequenced *T. b. rhodesiense* isolates and East African *T. b. brucei* TREU927 (Fig. 2A to C) suggests that variation between the B17 and Z310 zymodemes is largely restricted to differences at heterozygous loci on chromosomes 3, 5, 8, and 10. The largest such locus was on chromosome 8, and the B17 isolate is heterozygous across >70% of this 2.5-Mbp chromosome.

The KASPar SNP genotyping and data from public sequence databases for 31 loci across a total of 63 *T. brucei* subsp. isolates from Uganda, Zambia, and Côte d’Ivoire suggest that all isolates within the B17 population have alleles similar to those of the sequenced isolate (see Table S4 in the supplemental material). However, when the entire population of B17 isolates is sampled, there are only a few sites that have a unique allele that is not shared with Z310 isolates. These alleles are clustered on chromosome 8. While there is apparent linkage disequilibrium between these alleles and other parts of the genome, chromosome 8 is a strong candidate for loci that could underlie differences in virulence between B17 and Z310 parasites.

Examining chromosome 8 SNPs on a neighbor-joining tree (Fig. 3B) suggests that B17, EATRO3, and EATRO2340 parasites are heterozygous on chromosome 8, with shared alleles from both type 1 *T. b. gambiense* and *T. b. brucei* TREU927 (Fig. 3D). The Z310 zymodeme, by contrast, has few shared alleles with *T. b. gambiense* DAL972 on chromosome 8 and is more similar to the East African *T. b. brucei* TREU927.

Only the isocitrate dehydrogenase (ICD) isoenzyme consistently differentiates between the Zambezi and Busoga strain groups. The ICD gene on chromosome 8 is in the region that is heterozygous for *T. b. brucei* TREU927 and type 1 *T. b. gambiense*
alleles in B17 isolates. The ICD gene contains 3 nonsynonymous SNPs (nsSNPs) that are predicted to modify the charge and isoelectric point of the predicted protein and consequently the mobility of the ICD isoenzyme. ICD mobility determines placement in the Zambezi and Busoga strain groups, which correlates with virulence (15). Therefore, the isoenzyme data are consistent with the SNP genotype data, both of which implicate the heterozygous region of chromosome 8 as a major locus contributing to the observed differences in virulence.

Conclusions. Our analysis demonstrates that the T. b. brucei subspecies causing HAT have undergone genetic exchange in natural populations, since the East African B17 and Z310 parasites share alleles with West African type 1 T. b. gambiense. The associated differences in disease progression in isolates with differentially derived haplotypes has clear implications for parasite control and diagnosis, as other important traits, such as human serum resistance or drug resistance, may move between parasite groups. It will be important to identify how common this process is and where recombination occurs in the field.

Full-genome sequencing has been able to identify subtle genetic differences between parasite groups that were not apparent from microsatellite typing; however, other polymorphisms, such as small insertions and deletions, could not be detected by the sequencing chemistry. Nevertheless, the SNP data presented in
this study have been used to generate a panel of KASPar SNP markers that can now be employed to identify shared alleles and candidate loci underlying phenotypic differences. These techniques could be useful tools for further screening of field isolates in future studies.

MATERIALS AND METHODS

Trypanosome stocks. All parasite samples were from preexisting collections at the Liverpool School of Tropical Medicine and the University of Glasgow and have been described previously (15–17). All parasites were isolated from blood samples collected for diagnostic purposes and collected in Glasgow and have been described previously (15–17). All parasites were from preexisting collections. Isolates collected from blood samples generated from infected laboratory crosses were used in future studies.

Candidate loci underlying phenotypic differences. These techniques could be useful tools for further screening of field isolates in future studies.

Parasitemia in T. b. rhodesiense-infected mice. Line graph of mean parasitemia in CD-1 mice infected with B17 and Z310 zymodeme T. b. rhodesiense parasites ± standard errors (25 fields, thick film, 400× objective). Three isolates of each of the two zymodemes were each used to infect 12 mice (total of 72 mice). We killed 12 mice per week (two per parasite isolate), plus any mice that were killed due to moderate symptoms.

FIG 4 Parasitemia in T. b. rhodesiense-infected mice. Line graph of mean parasitemia in CD-1 mice infected with B17 and Z310 zymodeme T. b. rhodesiense parasites ± standard errors (25 fields, thick film, 400× objective). Three isolates of each of the two zymodemes were each used to infect 12 mice (total of 72 mice). We killed 12 mice per week (two per parasite isolate), plus any mice that were killed due to moderate symptoms.
STIB247 (accession number ERX000726). Similarly, sequencing reads for 
T. b. brucei (STIR247), a type 1 T. b. gambiense isolate (DAL972) (33), and a type 2 T. b. gambiense isolate from the Ivory Coast (STIB386) were downloaded directly from the Wellcome Trust Sanger Institute (WTSt) FTP website (ftp://ftp.sanger.ac.uk/pub/pathogens/Trypanosoma/brucei /T.b.gambiense_sequences/). All publicly available sequence data were generated using an Illumina genetic analyzer, except data for DAL972, which was sequenced using dideoxynucleotide (Sanger) sequencing (33). Sanger sequence read lengths exceed the maximum read length permissible by the Bowtie aligner. Therefore, Sanger reads were artificially split into 50-bp reads using a Perl script and treated as per next-generation sequencing data in order to align all data using the same alignment software. The Illumina genetic analyzer and artificial 50-bp Sanger sequencing reads were aligned to the TREU927/4 reference sequence using Bowtie (30). SNPs between the downloaded genomes were extracted using the Pileup feature in the SAMtools package (32). SNP loci were identified where an SNP was present in at least one genome and where there was >5-fold coverage in all genomes studied. For the SOLiD-sequenced genomes of T. b. rhodesiense B17 and Z310 isolates, the mean coverage at SNP loci was ~52-fold (Table 1). Due to the high copy number and variability in genes coding for variable surface glycoproteins (VSG), any SNP located within the boundaries of a VSG coding sequence were removed. A total of 109,495 non-VSG SNP loci were identified.

Selection of SNP loci for KASPar genotyping. In order to genotype a wider range of T. brucei subsp. isolates from both Uganda (T. b. rhodesiense) and West Africa (T. b. brucei/T. b. gambiense), duplicate whole-genome amplification (WGA) reactions (Illumina GenomePhip v2 DNA amplification kit; GE Healthcare) were performed on DNA extracted from stables (Ugandan samples) or from Whatman FTA punches (West African samples). DNA was extracted and WGA reactions were performed according to the manufacturer’s instructions as previously described.

Fifty nonsynonymous SNP loci (25 homozygous and 25 heterozygous) were selected from the MySQL database of all SNPs between B17 and Z310 isolates for subsequent typing of the remaining isolates. Loci were evenly distributed across all 11 megabase chromosomes, and all had a low (<0) BLOSUM50 score, indicating that they might modify protein activity. A 100-bp window surrounding the SNPs was extracted from a consensus sequence for the B17 and Z310 genomes using a Perl script and submitted to K Biosciences Ltd. (Hoddesdon, United Kingdom) along with DNA from a total of 63 isolates for SNP genotyping using their proprietary KASPar platform (http://www.kbioscience.co.uk/). Of the 50 SNP loci selected, 31 loci were successfully genotyped (Table S5). These assays are available to other users on application to KBiosciences.

Genome-wide SNP analyses. Two thousand seven hundred eighty-seven genome-wide, nonsynonymous SNPs between the T. b. rhodesiense B17 and Z310 genomes were compared to T. b. brucei (TREU927/4) and T. b. gambiense (type I, DAL972). A plot showing homozygosity and heterozygosity at these loci is shown in Fig. 2. Additionally, a bootstrapped (based on 1,000 replicates) Jukes-Cantor neighbor-joining (NJ) tree was created using SplitsTree for genome-wide SNP loci (Fig. 3A) and for those situated on chromosome 8 (Fig. 3B) (28).

Nucleotide sequence accession number. Sequence data have been submitted to the European Nucleotide Archive under study accession number ERP001836.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org /lookup/suppl/doi:10.1128/mBio.00197-13/-/DCSupplemental.

Table S1, DOCX file, 0.1 MB.
Table S2, DOCX file, 0.1 MB.
Table S3, DOCX file, 0.2 MB.
Table S4, XLSX file, 0.1 MB.
Table S5, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

I.G. was supported by a University of Liverpool Ph.D. studentship. Much of the work was funded by a Royal Society Wolfson Merit Award grant to N.H. S.K. was supported by an MRC award to N.W. A.M. was supported by a Wellcome Trust career development fellowship (reference number 079703/Z/06/Z) and a Wellcome Trust senior fellowship (reference number 095201/Z/10/Z).

REFERENCES


