



Kay, C. J., Peacock, L., Williams, T., & Gibson, W. C. (2022). Signatures of hybridization in *Trypanosoma brucei*. *PLoS Pathogens*, 18(2), [e1010300]. <https://doi.org/10.1371/journal.ppat.1010300>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1371/journal.ppat.1010300](https://doi.org/10.1371/journal.ppat.1010300)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via PLOS at <https://doi.org/10.1371/journal.ppat.1010300>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

RESEARCH ARTICLE

Signatures of hybridization in *Trypanosoma brucei*Christopher Kay¹, Lori Peacock^{1,2}, Tom A. Williams¹, Wendy Gibson^{1*}¹ School of Biological Sciences, University of Bristol, Bristol, United Kingdom, ² Bristol Veterinary School, University of Bristol, Bristol, United Kingdom* w.gibson@bristol.ac.uk

Abstract

Genetic exchange among disease-causing micro-organisms can generate progeny that combine different pathogenic traits. Though sexual reproduction has been described in trypanosomes, its impact on the epidemiology of Human African Trypanosomiasis (HAT) remains controversial. However, human infective and non-human infective strains of *Trypanosoma brucei* circulate in the same transmission cycles in HAT endemic areas in sub-Saharan Africa, providing the opportunity for mating during the developmental cycle in the tsetse fly vector. Here we investigated inheritance among progeny from a laboratory cross of *T. brucei* and then applied these insights to genomic analysis of field-collected isolates to identify signatures of past genetic exchange. Genomes of two parental and four hybrid progeny clones with a range of DNA contents were assembled and analysed by k-mer and single nucleotide polymorphism (SNP) frequencies to determine heterozygosity and chromosomal inheritance. Variant surface glycoprotein (VSG) genes and kinetoplast (mitochondrial) DNA maxi- and minicircles were extracted from each genome to examine how each of these components was inherited in the hybrid progeny. The same bioinformatic approaches were applied to an additional 37 genomes representing the diversity of *T. brucei* in sub-Saharan Africa and *T. evansi*. SNP analysis provided evidence of crossover events affecting all 11 pairs of megabase chromosomes and demonstrated that polyploid hybrids were formed post-meiotically and not by fusion of the parental diploid cells. VSGs and kinetoplast DNA minicircles were inherited biparentally, with approximately equal numbers from each parent, whereas maxicircles were inherited uniparentally. Extrapolation of these findings to field isolates allowed us to distinguish clonal descent from hybridization by comparing maxicircle genotype to VSG and minicircle repertoires. Discordance between maxicircle genotype and VSG and minicircle repertoires indicated inter-lineage hybridization. Significantly, some of the hybridization events we identified involved human infective and non-human infective trypanosomes circulating in the same geographic areas.

OPEN ACCESS

Citation: Kay C, Peacock L, Williams TA, Gibson W (2022) Signatures of hybridization in *Trypanosoma brucei*. PLoS Pathog 18(2): e1010300. <https://doi.org/10.1371/journal.ppat.1010300>

Editor: Tim Nicolai Siegel, Ludwig-Maximilians-Universität München Biomedizinisches Centrum München, GERMANY

Received: July 19, 2021

Accepted: January 22, 2022

Published: February 9, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1010300>

Copyright: © 2022 Kay et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data generated or analysed during this study are included in this published article, its supplementary information files and associated online resources listed in S1_Text. The data are available from the Dryad

Author summary

Sexual reproduction allows genes from different individuals to be mixed up in the offspring. This is particularly important for disease-causing microbes, because new

repository: DOI <https://doi.org/10.5061/dryad.xd2547djb> and NCBI Sequence Read Archive (SRA <https://www.ncbi.nlm.nih.gov/sra>) Project number: PRJNA795331.

Funding: We are grateful to the UK Biotechnology and Biological Sciences Research Council (<https://bbsrc.ukri.org/>) for funding (BB/R016437/1 to WG and TAW; BB/R010188/1 to WG). TAW is supported by a Royal Society University Research Fellowship (URFR\201024 <https://royalsociety.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

combinations of harmful traits can arise, potentially leading to more severe outbreaks of disease. Tsetse-transmitted trypanosomes are single-celled parasites that cause severe human and livestock diseases in tropical Africa. During their developmental cycle in the tsetse fly, trypanosomes can mate and produce hybrid trypanosomes, which have one set of chromosomes from each parent. But polyploid hybrids, with more than one set of chromosomes from one or both parents, are often observed too. Here we have investigated how these polyploid hybrids are formed by comparing the genomes of hybrid progeny with those of their parents. Analysis of the large, paired chromosomes of both diploid and polyploid hybrids showed frequent crossovers, which are the hallmark of meiosis, the special form of division that produces haploid gametes. This indicates that the polyploids were formed after meiosis rather than by fusion of the parental diploid cells. We also investigated the inheritance of two other features of trypanosomes: the large family of variant surface glycoprotein (VSG) genes, and the mitochondrial (kinetoplast) DNA. Hybrid clones had inherited about half the VSG genes from each parent, and also showed biparental inheritance of one component of the kinetoplast DNA, the minicircles. We assessed the relatedness of field-collected trypanosomes by comparing their VSG and minicircle repertoires, together with maxicircle genotype. While most isolates shared few VSGs or minicircles, a group of mostly human-infective strains from Uganda had a large proportion of their repertoires in common. Most of these trypanosomes were probably related by clonal descent, but we also identified that some were hybrids by the mismatch between their maxicircle genotype and their VSG and minicircle repertoires. These signals of hybridization were also detected in some of the other field-collected isolates, suggesting that genetic exchange is widespread in nature. Significantly, the hybridization events involved human infective and non-human infective trypanosomes circulating in the same geographic areas, providing a mechanism for the generation of new, potentially more pathogenic, trypanosome strains causing human disease.

Introduction

Sexual reproduction allows the mixing of genes from different individuals with formation of hybrid progeny. For microbial pathogens this is particularly important, as new combinations of traits such as drug resistance or virulence may be generated, potentially leading to more pathogenic strains and outbreaks of disease. Tsetse-transmitted trypanosomes such as *Trypanosoma brucei* and *T. congolense* are parasitic protists that cause severe human and livestock diseases in tropical Africa. During their developmental cycles in the tsetse fly, these trypanosomes undergo complex cycles of differentiation and proliferation in the fly's alimentary tract, ending up as infective metacyclics that are transmitted to the next host via the fly's saliva. In addition, *T. brucei* undergoes sexual reproduction in the fly's salivary glands, involving meiosis and production of haploid gametes [1–4], though many details remain to be elucidated. Consistent with meiosis, inheritance appears to follow Mendelian rules according to microsatellite analysis and most hybrid clones are diploid like the parental trypanosomes [5,6], but hybrids with high DNA contents, interpreted as triploid or tetraploid, also occur with some frequency [2,5,7,8]. Limited analysis of triploid hybrids has demonstrated the presence of three copies per genome of some housekeeping genes [5,7], but no genome-wide analysis has been carried out to date. Studies of chromosome inheritance after sexual reproduction in other members of the trypanosomatid family have shown that hybrid progeny are often polyploid. For example, in an experimental cross of *T. cruzi*, the hybrid progeny appeared to be the products of fusion

of the diploid parental trypanosomes, though with subsequent genome erosion [9]. Polyploid hybrids are frequently found in experimental crosses of *Leishmania* spp. [10–14]; indeed, of 24 hybrid clones from an *in vitro* cross of *L. tropica*, 19 (79%) were 3N or 4N [15].

In trypanosomes the mitochondrial DNA is tightly packaged into a unique organelle, the kinetoplast. The kinetoplast DNA (kDNA) consists of two sizes of circular DNA molecules: ~25 kb maxicircles, which encode genes required for mitochondrial function, and ~1 kb minicircles, which encode the guide RNAs used to edit maxicircle transcripts; maxi- and minicircles are intercalated into a single, giant network [16–18]. Analysis of the inheritance of kDNA in experimental crosses of both *T. brucei* and *Leishmania* spp. has revealed that hybrid clones have heterogeneous networks consisting of a mixture of parental minicircles; initially, the maxicircles are also heterogeneous, but after several generations of mitotic division and random partition between daughter cells, the relatively small number of maxicircles (~50) becomes homogeneous, so that inheritance of maxicircles appears to be uniparental [19–22]. These observations led to the hypothesis that the parental kDNA networks blend in the zygote, implying that both mitochondrial and cell fusion occur. However, the mechanism remains unknown and the very idea seems inconceivable to some, considering the elegant and highly controlled replication of kDNA [23].

Notwithstanding the experimental results from the laboratory, it has long been controversial how much sexual reproduction influences the population dynamics of *T. brucei* in nature [24–26]. This is epidemiologically important as both human-infective subspecies, *T. b. gambiense* (*Tbg*) and *T. b. rhodesiense* (*Tbr*), mingle with the non-human-infective subspecies *T. b. brucei* (*Tbb*) in infected mammalian and tsetse hosts, potentially facilitating mating and the generation of new strains of human-infective parasites. For example, transfer of the Serum Resistance Associated (SRA) gene from *Tbr* to *Tbb* generates new strains of the human infective parasite [27] and microsatellite analysis strongly supports the hypothesis that admixture between *Tbr* and *Tbb* has occurred in the past [28]. In contrast, the major group of *T. b. gambiense*, *Tbg1*, appears clonal, backed by comparison of whole genomes [29].

Here we have compared genomes of parental and hybrid progeny from a *T. brucei* cross, which produced presumed triploid and tetraploid progeny as well as the expected diploids (Table 1) [2], with the aim of verifying ploidy and deducing the mechanism of polyploidization. We confirmed the biparental inheritance of kDNA minicircles and of the variant surface glycoprotein (VSG) gene repertoires in hybrid progeny. In contrast to the parental trypanosome strains, hybrid progeny shared similar minicircle and VSG repertoires, and we searched for these signatures of hybridization among a collection of field-derived isolates.

Table 1. Attributes of parental and hybrid clones.

Trypanosome clone	Fluorescence colour	DNA content	Maxicircle type	Microsatellite alleles		
				PLC	XI-53	III-2
J10 RFP	R	2C	J10	ab	aa	ab
1738 GFP	G	2C	1738	cd	bc	cd
FIG2 (SG22 clone 7)	G	2C	J10	bc	ac	ad
FIR1 (SG22 clone16)	R	2C	mix	bc	ac	bc
F1R3N (SG1 clone 18)	R	3C	1738	ad	ac	bd?
F1Y4N (SG1 clone 4)	Y	4C	J10	bd	ab	ad

Fluorescence colour: R = red, G = green, Y = yellow i.e. red and green. Data from [2].

<https://doi.org/10.1371/journal.ppat.1010300.t001>

Results

Genome assembly

Genome data was obtained from four hybrid clones originating from an experimental cross of *T. brucei* J10 and 1738 [2] (**Table 1**). Previous measurements of DNA content showed that two of the hybrid clones (F1R1 and F1G2) were diploid like the parental trypanosomes, while two had high DNA contents consistent with triploidy (F1R3N) or tetraploidy (F1Y4N) [2]. Genome data from the parental lines J10 and 1738 was kindly provided by Adalgisa Caccone (Yale University) [30]. Details of isolates and assembled genomes can be found in **S1 Table**. Assembly of the hybrid isolates yielded similarly sized genomes from 150 bp read data with good contiguity (44.1–46.4 Mbp; N50 length 4959–5590 bp). Parent and field isolates sequenced from 75 bp read data showed more variation, with differences in total assembled size and contiguity likely influenced by the differences in read length and quality of sequencing. The assembled parent genomes were smaller (1738, 33.7 Mbp; J10, 36.9 Mbp), but had good contiguity (N50 length: 1738 13,768 bp, J10 8806 bp).

K-mer and SNP analysis of heterozygosity

We used k-mer frequency analysis of the unassembled reads as an initial guide to heterozygosity, genome complexity and coverage depth [31,32]. This analysis calculates the frequency of unique nucleotide sequences (k-mers) in the genomic reads. Thus, diploid cells will have a k-mer frequency plot with two peaks representing heterozygous and homozygous loci (1x and 2x coverage depth respectively); additionally, in the *T. brucei* genome the 1x peak will also include monoallelic VSG genes. The assembled *T. brucei* Lister 427 genome [33] was used as a reference to align unassembled reads for the identification of heterozygous SNPs in the core regions of the chromosomes excluding subtelomeric VSGs. In principle, any unique reference position in a diploid genome will agree, disagree, or have an observed allele frequency half way between the two (= heterozygous). The number of observed heterozygous sites per genome is a measure of heterozygosity and changes in ploidy would alter the observed allele frequency.

Fig 1 shows the k-mer and SNP frequency plots for the parent and hybrid trypanosomes. As expected, 1738 and J10 have two k-mer peaks, with J10 apparently having greater heterozygosity than 1738, as it has more heterozygous sites in the SNP plot (**Fig 1A and 1B**). Likewise, the diploid hybrid clone F1G2 has two peaks, with a prominent 1x peak (**Fig 1C**); the SNP frequency plot shows a far greater number of heterozygous SNPs than either parent, showing a gain in heterozygosity. Surprisingly, the other diploid hybrid clone (F1R1) did not conform to this pattern, with both the k-mer and SNP frequency plots showing multiple peaks (**Fig 1D**), suggesting that this “clone” consisted of a mixture of trypanosome strains. However, none of our subsequent analyses (see below) revealed the presence of nuclear or kinetoplast DNA from any strain outside the J10 x 1738 cross, and a mixture of two different hybrid clones would have produced anomalously large repertoires of VSG genes and kDNA minicircles. Therefore, our working hypothesis is that only genetic material from F1R1 is present, but that it has been reorganised through selfing. Of note is that the DNA used for genome sequencing here was derived after fly transmission of the original F1R1 clone without re-cloning, on the assumption that clones are transmitted faithfully because selfing only happens infrequently [34–36]. The complexity of the k-mer and SNP frequency plots (**Fig 1D**) suggests that a mixture of F1R1 and a selfed population is present. However, mixtures of F1R1 with a diploid selfed population would give rise to SNP peaks at approximately 0.25, 0.5 and 0.75, whereas the observed peaks are at approximately 0.2, 0.4, 0.6 and 0.8 (**Fig 1D**); in particular, the lack of the 0.5 peak does not fit with any mixture containing a diploid selfed population, suggesting instead that a

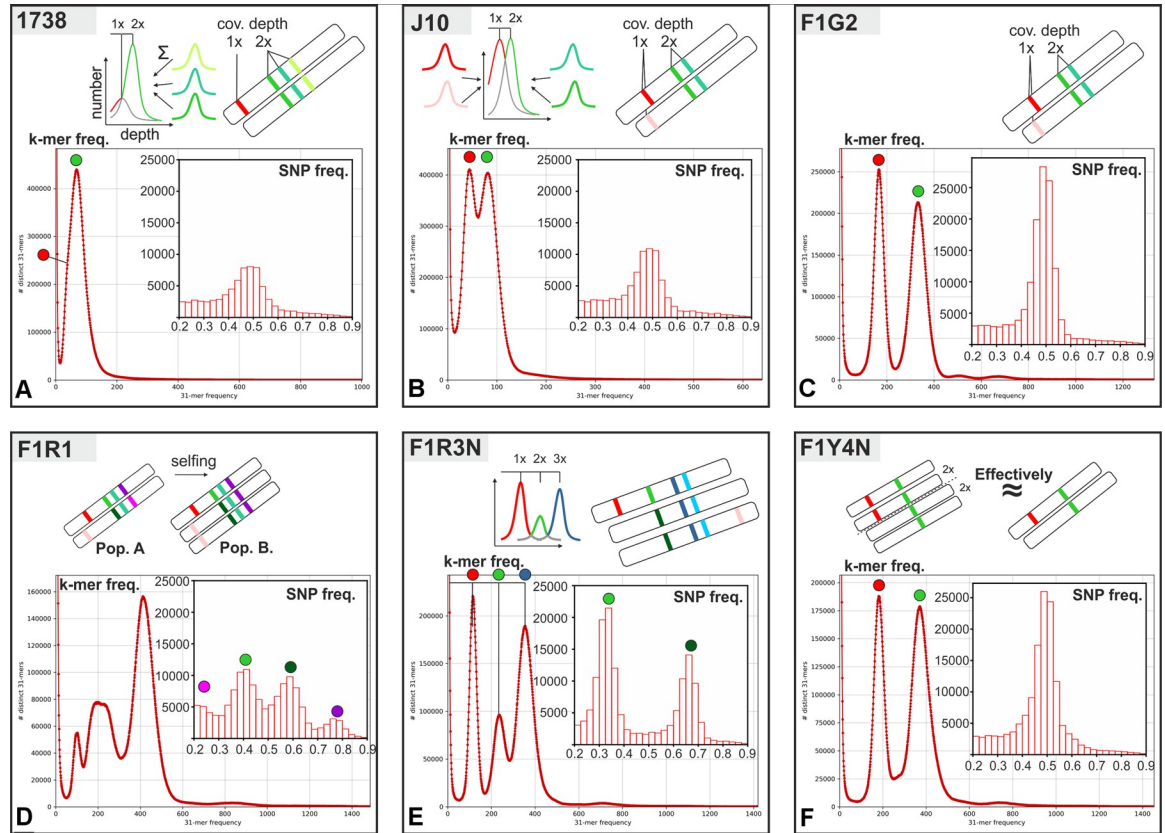


Fig 1. K-mer and SNP analyses reveal ploidy and heterozygosity of hybrid clones. Unassembled reads from parents (1738 and J10) and hybrid progeny (F1G2, F1R1, F1R3N, F1Y4N) were analysed by k-mer and SNP frequency; the plots are linked schematically to coverage depth via a genome model in the inset cartoons. The k-mer frequency plots show the number of k-mers with a specific coverage depth; equally spaced peaks correspond to multiples of k-mer incidence within the whole genome. The SNP frequency plots show the number of detected heterozygous SNPs within the core chromosomal regions, excluding VSGs in subtelomeric regions, and their relative observed allele frequency (proportion of observed reads with alternate value). Inset cartoons illustrate the inheritance of loci on one pair of homologues, linked by coloured dots to particular peaks on the frequency plots. Our working hypothesis is that F1R1 is a mixture of the original F1R1 clone (population A) and a selfed, possibly triploid, population (population B) formed during fly transmission of F1R1 after its original isolation; the 0.2:0.8 peak ratio is produced by chromosomal crossing over (see text for further explanation).

<https://doi.org/10.1371/journal.ppat.1010300.g001>

triploid selfed population is present, and interestingly this hypothesis is supported by the Smudgeplot [37] ploidy analysis (S1 Fig).

Despite the uncertainties in interpretation, we kept F1R1 in the analysis, because the original hybrid is present and makes up a substantial portion of the population; no information about its genomic repertoire has been lost and the presence of a selfed population only changes the observed allele frequency/coverage depth. Furthermore, this is an example of an interesting biological event and demonstrates the loss of heterozygosity on selfing.

For the two hybrid clones with raised DNA contents, the k-mer frequency plot for F1R3N has three peaks, and the SNP frequency plot has observed allele frequencies approaching 0.33 and 0.66 (Fig 1E), in accord with results generated for a triploid *T. congolense* isolate [38], while F1Y4N is indistinguishable from F1G2 (Fig 1F). F1Y4N had a DNA content consistent with tetraploidy, but microsatellite analysis detected only two alleles at each locus (Table 1), suggesting genome endoreplication rather than fusion of four independent gametes [2]; this observation is supported by the observation of only a single peak at 0.5 in the SNP frequency

analysis. To support our overall interpretation of ploidy in the hybrid clones, we carried out complementary analysis using Smudgeplot [37], included as a supplementary figure (S1 Fig).

Inheritance of parental chromosomes by hybrid progeny

SNP analysis was further used to identify the two parental homologues for each of the 11 megabase chromosomes, which carry housekeeping genes [39], and track their inheritance in the four hybrid progeny clones (Figs 2 and S2). As expected, the pattern of inheritance in the diploid progeny clone F1G2 is Mendelian, as there is one homologue from each parent and evidence of at least one crossover event for several chromosomes; the pattern for the original F1R1 hybrid was able to be resolved and is similar, supporting the interpretation from the k-mer analysis (Fig 1D) that this “clone” is a mixture of F1R1 and a minor selfed population, rather than another trypanosome strain or hybrid clone. The tetraploid clone F1Y4N follows the same pattern, as does the triploid clone F1R3N for chromosomes inherited from parent 1738 but not from parent J10 (Fig 2). Instead SNPs from both J10 homologues are present over extensive regions of seven of the 11 chromosomes (1, 4, 6, 7, 9, 10, 11), demonstrating that both homologues have been inherited and that this happened after meiotic crossing over had occurred.

The way the 3N hybrid was formed can be inferred from these observations. Firstly, the presence of a single 1738 homologue for each of the 11 chromosomes in F1R3N, nine of which have crossovers, indicates the contribution of a haploid 1738 genome that was a product of meiosis. The J10 partner also contributed a post-meiotic genome, as there is evidence of

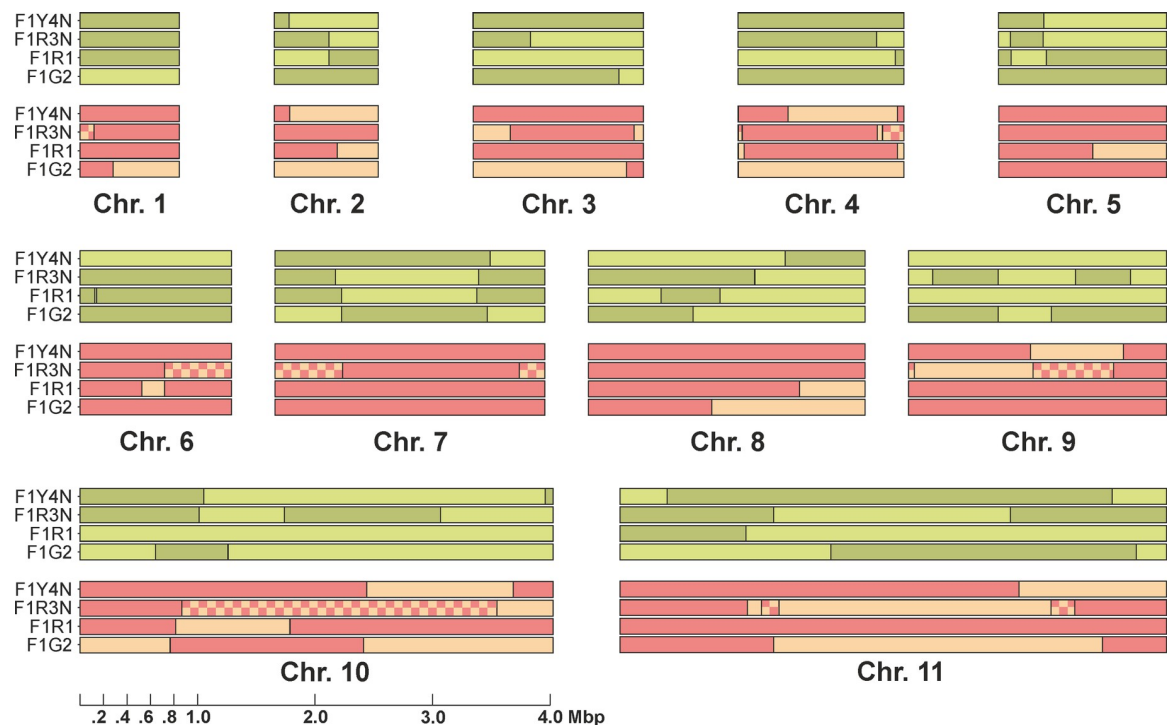


Fig 2. Pattern of inheritance of parental SNPs in four hybrid progeny clones. The core regions of the 11 chromosomes containing housekeeping genes are shown to scale. The two homologues from parent *T. b. brucei* 1738 are depicted in two shades of green, while those from J10 are red/orange. There is evidence of at least one crossover for most chromosomes as portions of both parental homologues are present in the chromosomes of the hybrids. The chequered blocks show regions where SNPs from both parental homologues were present. More detailed introgression maps are shown in S2 Fig.

<https://doi.org/10.1371/journal.ppat.1010300.g002>

crossing over for eight of the 11 chromosomes, which rules out the possibility that a vegetative diploid J10 cell fused with a haploid 1738 gamete. The fact that at least seven of the 11 chromosomes are present in two copies makes it unlikely that multiple, independent non-disjunction events occurred during meiosis in the J10 parent. Instead, the probable scenario is that the 1738 gamete fused either with a J10 cell in which meiosis was incomplete, or with two independent J10 gametes. We recently proposed a model of trypanosome meiosis in which gametes are produced sequentially and some intermediate stages have multiple nuclei [40], hence it is easy to envisage a scenario where fusion of a 1738 gamete with a J10 meiotic intermediate would produce the triploid genome observed in hybrid F1R3N. Moreover, we showed that meiotic intermediates, as well as gametes, expressed the membrane fusogenic protein HAP2 [40], potentially increasing the likelihood that meiotic intermediates, as well as gametes, fuse. Fusion of more than two gametes is also a possibility, although our observations of live gametes have mostly been of interacting pairs [3,41].

For the tetraploid clone F1Y4N, the chromosomal SNP patterns are indistinguishable from the diploid clone F1G2. This would not be the case if two diploid vegetative cells had fused, as both homologues from each parent would be present in the hybrid, and there would be no evidence of crossovers. The possibility that four independent haploid gametes fused is also ruled out by the absence of any regions of overlap (chequered blocks) in all 11 chromosomes (Fig 1). Fusion of meiotic intermediates is also ruled out, because again this would yield regions of overlap and moreover would need to have occurred in both parents. The remaining possibility is that endoreplication of parental chromosomes occurred after zygote formation, a parsimonious explanation since chromosomes from both parents would be duplicated in a single event.

Inheritance of VSG repertoire in hybrid clones

A total of 3120 VSG genes and fragments were identified from the assembled parental and hybrid genomes. Of these, 2256 greater than 200 amino acids in length were clustered with CD-HIT [42] to identify VSGs with >99.5% sequence identity; 617 clusters were identified together with 296 singletons that were present in only one genome. Considerably more VSG genes were recovered from the genomes of the hybrid clones than the parents (Fig 3), reflecting the higher quality of genome data obtained from the hybrid clones compared to the parents, coupled with our stringent search criteria for identifying VSG genes. The total number of VSGs per genome found here falls far short of the typical 1000 VSG estimate derived from DNA hybridization densitometry or genome assembly [39,43], since these totals include closely similar genes, which would be clustered here, and pseudogenes, whereas only open reading frames were counted here. Overall 573 VSGs were recovered from the parental genomes, but only 14 VSGs were found in both parents, most of the VSG repertoire being strain-specific. This allowed the parental origin of individual VSGs to be assigned in the hybrid progeny clones (Fig 3). Each clone had inherited approximately half its VSGs from each parent, with a bias to inheritance from J10 (paired t test, $p = 0.0335$), perhaps because J10 had more VSGs than 1738, or that the assembly of 1738 was less contiguous/complete.

Some VSG genes were found together on contigs as VSG gene arrays, while others occurred as isolated genes. Contigs containing more than two VSG genes were identified in hybrid progeny only, reflecting the higher quality of these genomes. There were a total of 41 arrays of 3–8 VSG genes, which probably represent the subtelomeric VSG arrays found on each arm of the 11 diploid chromosomes [33,39], but may also originate from smaller chromosomes. Considering the parental origin of these 41 arrays, there is a skew towards inheritance from J10: 25 were unambiguously from J10 and 12 from 1738, with 4 unassigned. Inheritance of these arrays in individual hybrid clones follows the same pattern, with roughly twice as many arrays

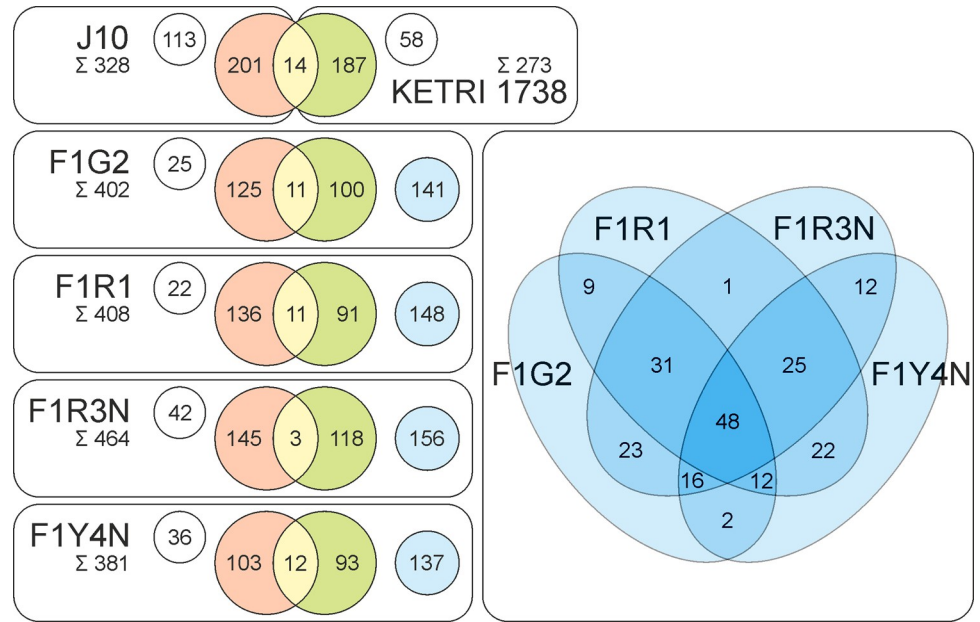


Fig 3. Inheritance of parental VSGs in hybrid progeny clones. Top: 587 VSGs were found in parents J10 (328 VSGs) and 1738 (273 VSGs) with 14 shared; those also found in one or more progeny clones are in red and green respectively, while strain-specific VSGs not found in progeny clones are shown in white circles. Bottom: VSGs inherited by the four hybrid progeny clones; parental origin is indicated by red (J10) and green (1738). Additional VSGs not found in either parent are shown in white circles if clone-specific, or blue circles if shared by one or more hybrid clones; the Venn diagram (right) expands the blue circle total to show the number of VSGs shared between individual hybrid clones.

<https://doi.org/10.1371/journal.ppat.1010300.g003>

from J10 as 1738 (mean no. of arrays from J10 and 1738 respectively = 17.50 and 6.75; paired t test, $p = 0.0074$; **Table 2**). This was true for diploid (F1G2) as well as polyploid (F1R3N, F1Y4N) hybrid clones, and does not therefore result from inheritance of extra J10 chromosomes.

Inheritance of metacyclic VSG expression sites

Metacyclic VSG genes (*MVSGs*) are found in short expression sites (ES) at the ends of chromosomes and are distinguished from bloodstream form expression sites (BESs) by the lack of ES-associated genes (*ESAGs*) and the presence of an upstream *MVSG* promoter [44–49]. Screening genomic contigs for the *MVSG* promoter produced 47 contigs, which could be resolved to 13 clusters with sequence identity >99.5%. The longest contig in each cluster terminated downstream of the promoter and VSG in telomeric repeats (TTAGGG) (with the exception of cluster 6), with any *ESAGs* identified located upstream of the promoter, giving confidence that these loci represent genuine metacyclic expression sites (MES; [44–49]). Expression of these *MVSGs* has also been observed in transcriptomic analysis of individual trypanosomes isolated

Table 2. Comparison of numbers of VSG gene arrays inherited by hybrid progeny from parents J10 and 1738.

Trypanosome clone	DNA content	Origin of VSG gene array			Total arrays
		J10	1738	Unassigned	
F1G2	2C	18	5	3	26
F1R1	2C	20	6	4	30
F1R3N	3C	19	10	1	30
F1Y4N	4C	13	6	4	23

<https://doi.org/10.1371/journal.ppat.1010300.t002>



Fig 4. Metacyclic VSG expression sites in parents and hybrid progeny. Contigs containing *mVSG* promoters were clustered by neighbour-joining into 13 distinct loci, represented by coloured circles, where parental origin is indicated by colour (red = J10, green = 1738). The parents J10 and 1738 share only locus 7 in common (within 4% identity over the aligned region). Each hybrid has inherited loci from each parent as shown. For locus 7, both parental copies were present in F1G2, F1Y4N and F1R3N. F1R3N has all but two of the parental loci.

<https://doi.org/10.1371/journal.ppat.1010300.g004>

from the tsetse salivary glands [50]. Eight MES were found in J10 and six in 1738, with cluster 7 shared between them (~96% sequence identity over aligned region). All putative MES identified in the hybrid progeny were unequivocally of parental origin, with a mixture inherited from each parent (Fig 4); while most hybrid progeny inherited 7–9 MES, the triploid F1R3N had 12 MES, suggesting increased diversity of its subtelomeric ends.

Inheritance of kinetoplast DNA in hybrid clones

The kinetoplast DNA (kDNA) of *T. brucei* consists of ~25 kb maxicircles and ~1 kb minicircles. Entire maxicircle coding regions were obtained from the parental and hybrid genomes. Assembly of the parental maxicircles revealed 50 SNPs and 15 single base pair indels distinguishing the maxicircles of J10 and 1738. Reads were aligned to the assembled sequences and no heteroplasmy was observed. Alignment of the maxicircle coding regions from the four hybrid clones showed that F1G2 and F1Y4N had maxicircles identical to those of J10, while F1R1 and F1R3N had maxicircles identical to those of 1738; no additional variation or recombination was observed. These results confirm our earlier findings based on presence of a single *HinfI* site in the cytochrome oxidase gene (Table 1), except that F1R1 was reported to have maxicircles of both parental types [2]. As this prior result was obtained during early passage of clone F1R1, we assume that maxicircles since became homogeneous through sequential vegetative divisions [21,51]. Again, the presence of only one maxicircle genotype in F1R1 confirms that it is not a mixture of two different strains, but a mixture of the original clone and a selfed population.

Sequences of 308 unique minicircles were recovered from the parental and hybrid genomes. Of these, 117 were found in parent J10 and 98 in parent 1738, while the hybrid clones had 213 (F1G2), 230 (F1R1), 219 (F1R3N) and 130 (F1Y4N) (Table 3). We assume that more minicircles were recovered from the hybrid genomes due to the higher quality of sequence data. The hybrid clones had inherited approximately equal numbers of minicircles from each parent (mean no. of minicircles from J10 and 1738 was respectively 72 and 62; paired t test, $p = 0.0668$; Table 3). A total of 67 minicircles were shared across the four hybrid clones, of which 28 were found in J10, 19 in 1738 and 20 in neither parent. It is noteworthy that the triploid and tetraploid hybrids had no excess of minicircles compared to the diploid hybrids.

Table 3. Parental origin of minicircles in hybrid clones.

Parental origin of minicircle	F1G2	F1R1	F1R3N	F1Y4N
1738	59	76	73	40
J10	80	83	78	48
Subtotal	139	159	151	88
Unassigned	74	71	68	42
Total	213	230	219	130

<https://doi.org/10.1371/journal.ppat.1010300.t003>

In summary, we have confirmed uniparental inheritance of the maxicircle component of kDNA in hybrid progeny, but biparental inheritance of the minicircles in *T. brucei* [19–21].

Comparison of field isolates

Comparison of the inheritance of nuclear and kinetoplast DNAs in laboratory crosses has shown that hybrids inherit a mixture of VSGs and minicircles from both parents, but inherit their maxicircle type from just one parent. Therefore, we investigated whether such patterns of inheritance could be identified among field-collected isolates to detect signatures of genetic exchange. Genome data from 33 isolates of *T. brucei* (16 *T. b. brucei*, 14 *T. b. rhodesiense*, 3 *T. b. gambiense*) and four isolates of *T. evansi* was kindly provided by Adalgisa Caccone (Yale University) [30]; details of isolates and assembled genomes can be found in **S1 Table** with k-mer and SNP analysis shown in **S3 Fig**. VSG genes and minicircles were recovered from each genome. For the *T. brucei* isolates, kinetoplast DNA maxicircle coding regions obtained previously [52] were used to derive a phylogenetic tree (**Fig 5A**). While the majority of maxicircle sequences were similar and formed a Pan-African group, three discrete clades were evident: Sindo, Kiboko and Lister 427 group. These groupings were used to order *T. brucei* isolates in the data matrix (**Fig 5B**), which displays the numbers of shared VSGs and minicircles between each pair of isolates; only VSG data is shown for *T. evansi*, because minicircles are largely homogeneous in this species [53].

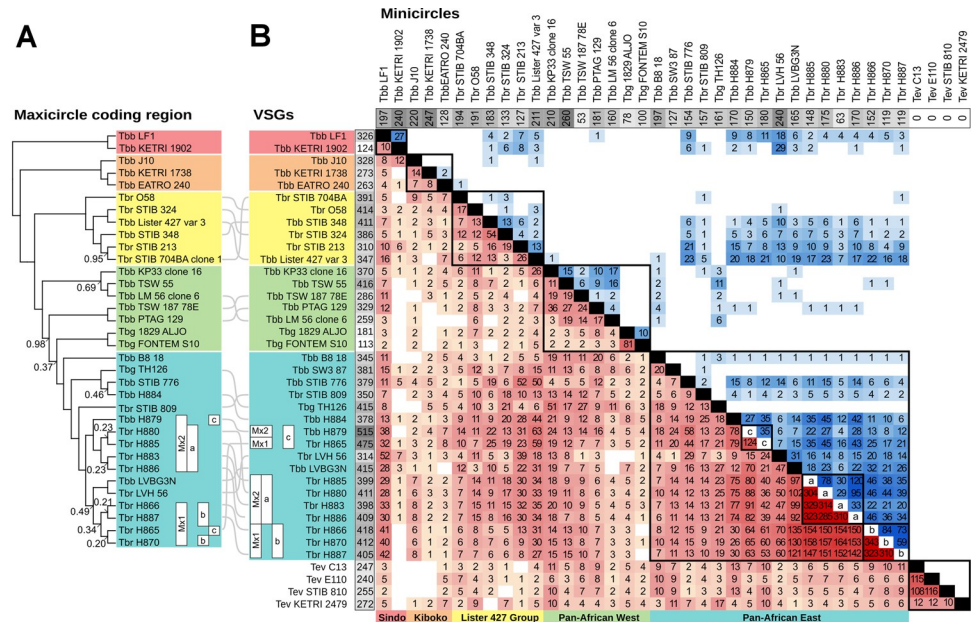


Fig 5. Shared VSGs and kinetoplast DNA minicircles among field-collected isolates of subgenus *Trypanozoom*. A. Phylogenetic tree of the maxicircle coding region of *T. brucei* isolates, excluding SW3/87, which produced a fragmented maxicircle on assembly. Node values show posterior probabilities <1. Isolates are grouped by colour: red, Sindo; orange, Kiboko; yellow, Lister 427 group; green, Pan-African West; cyan, Pan-African East. In this last group, two clades are evident among Ugandan isolates (Mx1 and 2), and the VSG cluster (a, b, c) is also shown. B. Data matrices of shared VSGs (red) and minicircles (blue) from 39 isolates of subgenus *Trypanozoom*. Higher values are indicated by darker colour. The grey boxes in the horizontal row and vertical column show total numbers of minicircles and VSGs recovered, respectively. Isolates are grouped by colour largely as in A, with SW3/87 now included in the cyan group and other small changes within group to highlight shared VSGs. As *T. evansi* isolates have only one major minicircle type, minicircle data is voided. For the Ugandan isolates, the VSG cluster (a, b, c) is indicated on the diagonal.

<https://doi.org/10.1371/journal.ppat.1010300.g005>

About a quarter of isolates, including J10 and 1738, had fewer than 25 VSGs in common with any other isolate (KETRI 1902, J10, 1738, EATRO 240, 058, STIB 704, LM 56, B8/18, SW3, KETRI 2479; **Fig 5B**), reflecting the immense diversity of VSG repertoires among trypanosomes separated by vast distances and collected at different times. Although isolates of *T. b. gambiense* Type 1 (*Tbg1*) and *T. evansi* Type A (*TeA*) were also collected separately in space and time, *Tbg1* Aljo and Fontem shared 81 VSGs, and *TeA* C13, E110 and STIB 810 shared 108–116 VSGs, despite originating from three different continents. However, the *TeA* isolates shared only 12–14 VSGs with KETRI 2479, a Type B *T. evansi*, confirming the distinct nature of this subgroup [54–57].

The most overlapping VSG repertoires were found among *T. b. brucei* (*Tbb*) and *T. b. rhodesiense* (*Tbr*) isolates originating from the HAT focus in southeast Uganda, collected over a 20 year period from 1990 to 2010. Some of these isolates shared >300 VSG genes, a large part of their identified VSG repertoire, indicating that they are very closely related by descent, either as clonal lineages or through interbreeding. Among the ten isolates analysed here, three VSG clusters can be distinguished: (a) H880, H883, H885, H886, sharing 285–329 VSGs; (b) H866, H870, H887, sharing 310–343 VSGs; (c) H865, H879, sharing 124 VSGs (**Fig 5B**). H884, a bovine *Tbb* isolate from 2003, is an outlier, sharing 74–79 VSGs with isolates in clusters a and c, and only 26–30 VSGs with isolates in cluster b. Genomes of these ten isolates have been analysed previously by SNP analysis excluding VSGs [30] and clusters a–c correspond to clusters 5, 6 and 9 respectively, with the outlier H884 also placed in cluster 5. Similar relationships were evident from phylogenetic analysis of the SNP data, with isolates in clusters a–c distributed in three clades and H884 placed with cluster a isolates [58]. Concordance between the VSG and SNP data is expected, as both datasets relate to information on the 11 pairs of large chromosomes, though some VSGs are also found on smaller chromosomes.

Minicircles were rarely found in common between isolates, even those that shared maxicircle type (**Fig 5B**); for example, the three Kiboko isolates (J10, 1738, EATRO 240) shared 0–2 minicircles, *Tbg1* isolates Aljo and Fontem shared only 10 minicircles, and the Lister 427 group shared a maximum of 13 minicircles. The largest numbers of shared minicircles occurred among the Ugandan isolates that also had extensively overlapping VSG repertoires. The question is whether these shared VSG and minicircle repertoires signify genetic exchange or simply result from clonal descent.

Evidence for genetic exchange among Ugandan isolates

The ten Ugandan isolates all originate from SE Uganda, an epidemic that started in the 1970s and spread northwards [59–63]. Clusters a and b contain *Tbr* isolates collected in 2003–2010 and 1990–1992 respectively (**Table 4**) and represent two different *Tbr* genotypes circulating in the epidemic. A third *Tbr* genotype is represented by H865, isolated from a patient in 1990. H865 was found in cluster c, together with a bovine *Tbb* isolate, H879; as *Tbb* H879 was isolated nearly two decades after *Tbr* H865, the shared VSG genes could indicate clonal descent of H879 from H865, with loss of the *SRA* gene, which confers human infectivity [64], or past genetic exchange between *Tbb* and *Tbr*.

How can we distinguish clonal descent from genetic exchange? During clonal descent, the kinetoplast DNA maxicircles and minicircles will be inherited conservatively, but, as we have seen from the laboratory cross, hybrids inherit maxicircles from either parent and minicircles from both parents. Hence hybrids are predicted to show discordance between their genetic relatedness based on maxicircle genotype compared to minicircles and VSGs. Maxicircle SNPs divide the Ugandan isolates into two clades Mx1 and Mx2 (**Fig 5A**), corresponding broadly to VSG clusters a and b, with the two isolates in cluster c subsumed into either Mx1 (H865) or

Table 4. VSG and kinetoplast DNA profiles of ten Ugandan isolates.

Isolate	Subspecies	Host	Isolation year	VSG cluster	Maxicircle clade	Minicircles maximum share
H865	Tbr	H	1990	c	Mx1	H880 45/180 (25%)
H866	Tbr	H	1990	b	Mx1	H886 46/152 (30%)
H870	Tbr	H	1990	b	Mx1	H866 84/119 (71%)
H879	Tbb	Bv	2009	c	Mx2	H865 35/150 (23%)
H880	Tbr	H	2003	a	Mx2	H886 95/175 (54%)
H883	Tbr	Dog	2005	a	Mx2	H886 33/63 (52%)
H884	Tbb	Bv	2003	-	-	H880 45/170 (26%)
H885	Tbr	H	2010	a	Mx2	H886 120/148 (81%)
H886	Tbr	H	2010	a	Mx2	H885 120/170 (71%)
H887	Tbr	H	1992	b	Mx1	H866 73/119 (61%)

Subspecies designation is based on presence of the SRA gene [57]; Tbb = *T. brucei brucei*, Tbr = *T. b. rhodesiense*. Host: H = human, Bv = bovine. VSG cluster: according to numbers of VSGs shared between isolates (see Fig 5B); H884 shared <80 VSGs with any of these isolates. Maxicircle clade: see Fig 5A. Minicircles maximum share: maximum proportion of identified minicircles shared with another isolate (see Fig 5B).

<https://doi.org/10.1371/journal.ppat.1010300.t004>

Mx2 (H879), though statistical support for the exact position of H865 in Mx1 is low. In effect, maxicircle genotype has partitioned the isolates according to date of isolation, with clade Mx1 corresponding to 1990–1992 and Mx2 to 2003–2010 (Table 4). Thus, H865 (isolated 1990) and H879 (isolated 2009) have the common maxicircle genotype circulating during their separate eras, but do not share the expected VSG repertoire, a for Mx2 or b for Mx1 (Fig 5A), indicating that both are products of genetic exchange. In contrast, isolates from clusters a and b share both VSG repertoires and maxicircle genotypes (Table 4), as well as SNP profiles [30,58], and are probably linked by clonal descent. For clusters a and b, this conclusion is reinforced by the minicircle inheritance results, which show that a large proportion of minicircles are shared as well as VSGs (Table 4). However, for the two isolates in cluster c (H865, H879), minicircle repertoires are more heterogeneous in origin and do not show concordance with maxicircle genotype, both signatures of kDNA hybridization.

In summary, the evidence suggests that *Tbr* H865 and *Tbb* H879 represent lineages that have undergone genetic exchange, based on discordance between maxicircle genotype and VSG and minicircle repertoires. We cannot rule out the possibility that genetic exchange has also occurred among cluster a and b isolates, but clonal descent adequately explains the congruence between their kinetoplast DNA and VSG repertoires.

Evidence for genetic exchange among distantly-related lineages

Mating between distantly related trypanosomes will yield hybrid progeny with melded VSG and minicircle repertoires, but maxicircles of either parental type. As these related clonal lineages diverge, the proportion of shared VSGs and minicircles will gradually diminish, while the maxicircles will also diverge through accumulated mutations. Hence signatures of old hybridization events might still be present in apparently unrelated isolates.

Fig 6 highlights the relationships between VSG repertoire and maxicircle genotype among the 39 diverse isolates of subgenus *Trypanozoon* studied here. Within maxicircle genotype, isolates that share a large proportion of their VSG repertoire are connected by broad blue ribbons: Ugandan clusters a, b and c; *Tbg1* Aljo and Fontem; *TeA* C13, E110, STIB 810. But connections between isolates of different maxicircle genotype are also evident. Most remarkable is the linkage between Sindo *Tbb* LF1 and isolates with the Pan-African East maxicircle genotype from the same geographical region surrounding Lake Victoria. While LF1 shares only 10 VSGs with

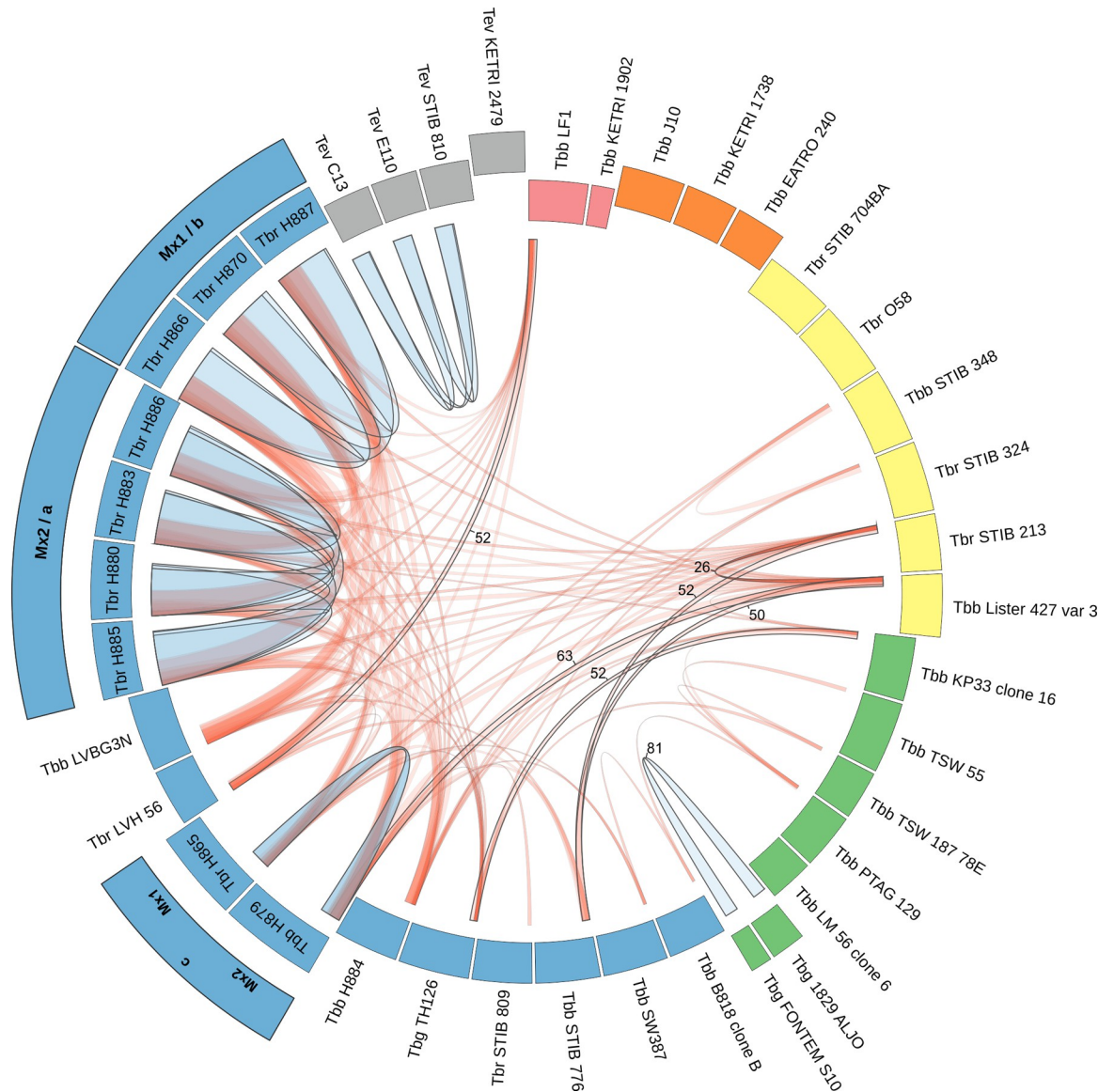


Fig 6. Relationships between VSG repertoire and maxicircle type. Data from the VSG matrix (Fig 5B) was filtered to show only linkages between isolates with more than 20 VSGs in common. The 35 *T. brucei* isolates are arranged and colour-coded according to the maxicircle phylogeny (Fig 5A), with the *T. evansi* isolates included as grey blocks. Maxicircle clade (Mx1 or 2) and VSG cluster (a, b, c) are shown for related Ugandan isolates. The block corresponding to each isolate, and the ribbons connecting them, are scaled to the number of VSGs; Ugandan isolates in VSG clusters a and b share the majority of their VSGs, as ribbons are almost as wide as the blocks. Ribbons are coloured red to help visualise more distant links, which are likely to reflect genetic exchange, from within-group links (blue); the numbers give the actual number of shared VSGs for selected links.

<https://doi.org/10.1371/journal.ppat.1010300.g006>

the other Sindo isolate, *Tbb* 1902, it shares 52 VSGs with *Tbr* LVH 56 from Lambwe Valley, Kenya, and 40–42 with Ugandan cluster b isolates (H866, H870, H887) (Figs 5B and 6). Both LF1 and LVH 56, together with LVBG3N, originate from an early 1980's HAT outbreak in the Lambwe Valley [65], while *Tbb* 1902 was collected earlier (1971) from a waterbuck. Despite the divergence of maxicircle genotypes, the Sindo isolates LF1 and 1902 also share minicircles with LVH 56 (18 and 29 respectively, Fig 5B). Discordant maxicircle genotype combined with overlapping VSG and minicircle repertoires points to genetic exchange among trypanosomes in the Lambwe Valley, with further links to trypanosomes circulating in south-east Uganda.

The Lister 427 group also shows the hallmarks of genetic exchange, with strong linkages of 50 or more VSGs between STIB 213 and Lister 427 to Pan-African East isolates STIB 776, H879 and H865 (Fig 6); around 20 minicircles are also shared (Fig 5B).

In contrast, the Kiboko maxicircle genotype shows no linkages, either within the group (<14 VSGs in common) or to different maxicircle genotypes (Fig 6), and the majority of their identified VSGs (~80%) are unique to the isolate (S2 Table). As both parents in the experimental cross (J10 and 1738) belong to the Kiboko group, there is no question that isolates within this group are fully capable of mating, both within the Kiboko group [2] and with strains of different maxicircle genotype [20,66,67], demonstrating that this is not an asexual maxicircle clade. Analysis of more isolates, compared to the limited sampling in this study, may reveal stronger connections between maxicircle clades.

Discussion

Comparison of the genomes of parents and hybrid progeny from an experimental cross of *T. brucei* has provided support for a number of hypotheses about the process of sexual reproduction in trypanosomes. There was already experimental evidence of a meiotic division in *T. brucei* [4] and SNP analysis has confirmed that crossovers occurred on all 11 pairs of megabase chromosomes when hybrid progeny were formed. SNP analysis also demonstrated the presence of two sets of post-meiotic chromosomes from one parent, together with one set from the other parent, in a hybrid with a 3C DNA content. This triploid hybrid was therefore most likely formed by fusion of a gamete with a meiotic intermediate rather than a vegetative diploid cell. Similarly, the demonstration of numerous crossovers in the chromosomes of a hybrid with a 4C DNA content showed that it did not arise by fusion of diploid cells, but probably by endoreplication of chromosomes post zygote formation.

Surprisingly, one of the two diploid hybrid progeny included in this analysis turned out not to be a pure clonal population, but a mixture of the original clone and a selfed population. This was evident from the k-mer analysis, but not suspected from inheritance of VSGs or minicircles. Tracing the derivation of the cryopreserved stock used for genome sequencing, revealed that it had been tsetse-transmitted without subsequent re-cloning. Our previous studies have shown that meiosis and production of haploid gametes occur in clonal populations during fly transmission [3,4], opening the possibility of selfing, though this is thought to be rare [34–36]. This assumption needs to be reconsidered in the light of the evidence of selfing revealed here.

Hybrid progeny inherited substantial numbers of VSG genes from both parents, demonstrating that sexual reproduction mixes up VSG genes and thereby generates novel VSG repertoires. This confirms previous results demonstrating immunologically that experimental hybrids had recombinant variable antigen type repertoires [68]. Besides the inheritance of chromosome-internal VSG gene arrays, we also showed that hybrid progeny inherit a mixture of metacyclic expression sites from both parents, each containing a metacyclic VSG gene; every hybrid is therefore equipped with a new combination of metacyclic VSGs, which may be a competitive advantage in successfully establishing infection in the mammalian host. In addition to exchange of nuclear DNA, trypanosomes recombine their mitochondrial (kinetoplast) DNA during mating. Though kinetoplast DNA (kDNA) maxicircles showed uniparental inheritance with no detectable evidence of heteroplasmy, the minicircles were inherited from both parents, giving rise to a hybrid kDNA network. Hybrid trypanosomes thus share minicircles with their siblings as well as their parents. Fig 7 summarizes the different modes of inheritance of nuclear and mitochondrial genomes.

We used these observations to search for evidence of hybridization among trypanosome isolates from endemic areas. Isolates collected from geographically distant locations at

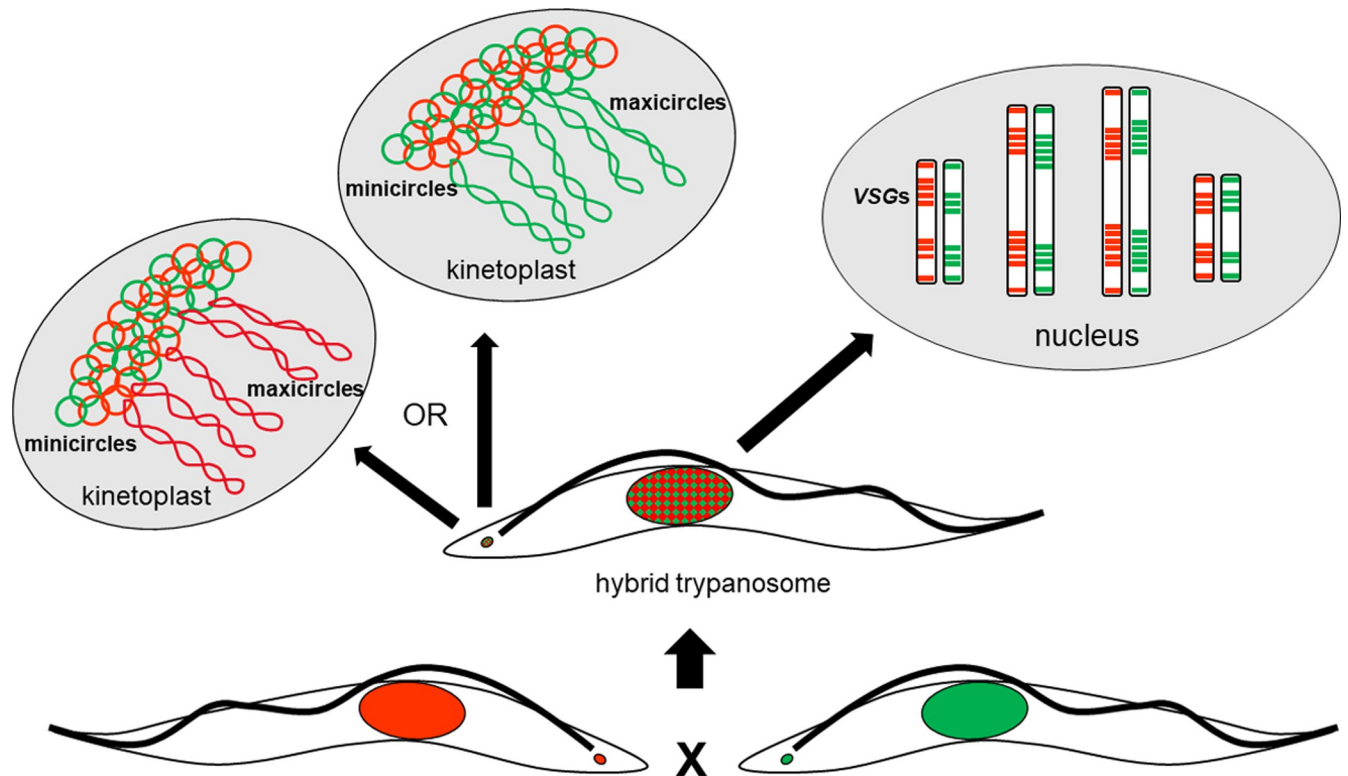


Fig 7. Modes of inheritance of nuclear and kinetoplast DNA. Diagram summarizing the different modes of inheritance of nuclear and mitochondrial genomes.

<https://doi.org/10.1371/journal.ppat.1010300.g007>

different times generally had non-overlapping *VSG* and *kDNA* minicircle repertoires, with some exceptions such as the clonally propagated genotypes *Tbg1* and *TeA*, reaffirming the consensus that these are widespread clonal lineages [28,57,58,69–72]. However, for nine Ugandan isolates, more closely associated in space and time, a large proportion of *VSGs* and minicircles were shared, demonstrating close genetic relatedness brought about by clonal descent or genetic exchange. While shared maxicircle genotype would be expected for clonal descent, divergent maxicircle and nuclear genotypes provided unequivocal evidence of hybridization. Just this limited sampling of isolates allowed us to build up a complex picture of interrelatedness, some by direct, clonal descent, and some by genetic exchange.

Across the broader range of isolates sampled, further instances of discordance between maxicircle genotype and *VSG* and minicircle repertoires were identified, suggesting widespread genetic exchange among *T. brucei* isolates and that many lineages are hybrid. This has resolved some previous discrepancies where isolates have been assigned to different phylogenetic groups according to maxicircle or nuclear DNA genotyping [73].

Significantly, there was evidence of hybridization between human infective and non-infective trypanosomes, providing a mechanism for the generation of new, potentially more pathogenic, strains of *Tbr*. *VSG* repertoires overlapped extensively among *Tbr* and *Tbb* isolates from East African HAT foci; for example, *VSG* cluster *c* from Uganda contains both *Tbr* and *Tbb* isolates with 124 *VSGs* in common, while *Tbb* LVBG3N from Lambwe Valley, Kenya, shared 121–135 *VSGs* with *Tbr* isolates from Uganda. Similarly in West Africa, *Tbg2* TH126 from the Daloa/Vavoua/Bouaflé HAT area in Côte d'Ivoire shared 52 *VSGs* with *Tbb* KP33 isolated from a tsetse fly.

Methods

Trypanosomes

Hybrid clones originated from an experimental cross of *T. b. brucei* J10 and 1738 [2] and comprised two diploid progeny F1R1 and F1G2 (originally called SG22 clone 16 and SG22 clone 7 respectively), which were subsequently used in F1 crosses [41], and two hybrid progeny with raised DNA contents presumed to be triploid (F1R3N = SG1 clone 18) or tetraploid (F1Y4N = SG1 clone 4; **Table 1**). The four hybrid clones were grown as procyclics in Cunningham's medium (CM) [74] supplemented with 10 µg/ml gentamycin, 5 µg/ml hemin and 15% v/v heat-inactivated foetal calf serum (FCS) at 27°C. High molecular weight DNA for genome sequencing was purified from approximately 5×10^8 trypanosomes using a Blood and cell culture kit (Qiagen) and a modification of the manufacturer's yeast cell protocol. Briefly, cells were pelleted by centrifugation, washed once with PBS and resuspended in 5 ml lysis buffer containing proteinase and RNAase as per the manufacturer's protocol. Following 1 hour incubation at 50°C, lysates were centrifuged at 5000 rpm for 5 minutes at room temperature in a microfuge to pellet debris before the supernatant was applied to a Genomic-tip 100/G column (Qiagen). Subsequent processing followed the manufacturer's protocol; after isopropanol precipitation, DNA was resuspended in 200 µl 10 mM Tris, 0.1 mM EDTA, pH 8 and stored at 4°C.

Genome sequencing and assembly

DNAs from the four hybrid clones (F1R1, F1G2, F1R3N, F1Y4N) were sequenced by the Earlham Institute, Norwich, UK, using Illumina NovaSeq with 150 bp paired end reads (data available from <https://www.ncbi.nlm.nih.gov/sra> Project no. PRJNA795331). Illumina sequence data (75 bp reads) from J10 and 1738, together with 37 other subgenus *Trypanozoon* isolates, was kindly provided by Adalgisa Caccone, Yale, USA [30]. Reads were assembled using SPAdes v3.13.1 [75]. K-mer frequency analysis was performed using Fastp [76], and histograms produced in KAT [32]. SNPs were identified by read alignment using BWA, processed with samtools and bcftools [77] before SNPs were called using Freebayes. RTG Tools and Tabix [78] were used to produce numerical reports and a Tablet v1.19.09.03 [79] and Artemis [80] were used to visualise SNP distribution. Further details of genomes and bioinformatic analyses are provided in **S1 Table** and **S1 Text**. The data are available from the Dryad repository: <https://doi.org/10.5061/dryad.xd2547djb> [81].

VSG gene analysis

VSG open reading frames were predicted from the contig pool using Transdecoder [82] and VSGs identified by Phmmer [83] using the Pfam HMM for the VSG C-terminal domain. For cluster analysis, sequences of greater than 200 amino acids in length with a Phmmer score $> 1e-6$ were clustered with CD-HIT [42]. Clusters were defined by proteins with regions overlapping 95% of total length sharing 99.5% sequence identity. Further details of bioinformatic analysis are provided in **S1 Text** and the data are available from the Dryad repository: <https://doi.org/10.5061/dryad.xd2547djb> [81].

Metacyclic expression sites

An HMM was derived from known metacyclic promoter sequences [47,49] and nhmmer [84] was then used to screen genomic contigs with this model. Sequences were clustered by alignment using MAFFT and tree construction via neighbour-joining. These groups were then confirmed by all vs all BLAST; the great variation in contig length and position of overlap made

this method preferable over CD-HIT, though largely similar results were obtained. Further details of bioinformatic analysis are provided in **S1 Text** and the data are available from the Dryad repository: <https://doi.org/10.5061/dryad.xd2547djb> [81].

Kinetoplast DNA analysis

Minicircles were identified from the contig SPAdes pool using nhmmer [84] with HMM derived from *T. brucei* minicircle sequences from the public Entrez database, using an e-value threshold of 1e-6. Reads aligning to these sequences were extracted from the read pool using Magic-BLAST [85] and a specialised sub-assembly of the reads performed with SPAdes using a kmer value of 55 and the plasmid mode. Circular assembled molecules were filtered by size and HMM. Sequences were oriented uniformly by identifying the minicircle conserved sequence blocks by BLAST v2.2.31+ [86]. Clustering was performed using CD-HIT [42] with thresholds of >98% identity for sequences within 90% of total length. Maxicircles were identified from the contig pool using BLAST v2.2.31+ [86]. Sequences were oriented and aligned using MAFFT v7.427 [87] and trimmed to the coding region. Site differences were identified using SNP-sites [88]. Heteroplasmy was assessed by aligning the reads to assembled contigs using BWA MEM v0.7.17 [89] and visualising the alignment in Tablet v1.19.09.03 [79]. A phylogeny was produced from the alignment using IQ-Tree ModelFinder [84] to compare base substitution models and parameters, and BEAST [90] to perform the final phylogeny. Trees were sampled every 1000 iterations over a chain length of 10,000,000; run results were visualised in Tracer [91] and consensus tree by Treeannotator v1.10.4 [90]. Further details of bioinformatic analysis are provided in **S1 Text** and the data are available from the Dryad repository: <https://doi.org/10.5061/dryad.xd2547djb> [81].

Dryad DOI

<https://doi.org/10.5061/dryad.xd2547djb>

Supporting information

S1 Fig. Smudgeplot analysis of hybrid and parent read data. Read pools were quality filtered with Fastp and k-mer histogram tables were then analysed with Smudgeplot (Ranallo-Benavidez TR et al 2020 doi.org/10.1038/s41467-020-14998-3). Proposed ploidy is shown under isolate name, with the probability of other karyotypes shown on the right. Parental strains 1738 and J10, together with hybrid progeny clones F1G2 and F1Y4N, all fit to diploid, while F1R3N fits best to triploid. The anomalous hybrid clone F1R1 has an unusual intermediate pattern inconsistent with a pure diploid population.

(TIF)

S2 Fig. Introgression maps of hybrid clones. Introgression maps were constructed from patterns of inherited heterozygous SNPs from the parental strains. There were ~20,000 SNPs in each hybrid and the figure illustrates their density and distribution.

(TIF)

S3 Fig. K-mer analysis of genomic reads from *T. brucei* and *T. evansi* isolates. K-mer analysis for the 37 additional field isolates. Although k-mer peaks are unresolved for some isolates, the range of peak shapes suggests that these populations have varying levels of heterozygosity.

(TIF)

S1 Table. Details of trypanosome isolates and assembled genomes.

(XLSX)

S2 Table. VSG repertoire by proportion shared by maxicircle clade.
(TIF)

S1 Text. Supplementary bioinformatics methods and data.
(DOCX)

Acknowledgments

Many thanks to Adalgisa Caccone for supplying raw read sequence data from field isolates.

Author Contributions

Conceptualization: Christopher Kay, Tom A. Williams, Wendy Gibson.

Data curation: Christopher Kay.

Formal analysis: Christopher Kay, Lori Peacock.

Funding acquisition: Tom A. Williams, Wendy Gibson.

Investigation: Christopher Kay, Lori Peacock.

Methodology: Christopher Kay, Tom A. Williams.

Project administration: Tom A. Williams, Wendy Gibson.

Resources: Lori Peacock, Wendy Gibson.

Supervision: Tom A. Williams, Wendy Gibson.

Visualization: Christopher Kay, Wendy Gibson.

Writing – original draft: Wendy Gibson.

Writing – review & editing: Christopher Kay, Lori Peacock, Tom A. Williams.

References

1. Jenni L, Marti S, Schweizer J, Betschart B, Lepage RWF, Wells JM, et al. Hybrid formation between African trypanosomes during cyclical transmission. *Nature*. 1986; 322:173–5. <https://doi.org/10.1038/322173a0> PMID: 3724860
2. Gibson W, Peacock L, Ferris V, Williams K, Bailey M. The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. *Parasit Vectors*. 2008; 1:4. <https://doi.org/10.1186/1756-3305-1-4> PMID: 18298832
3. Peacock L, Bailey M, Carrington M, Gibson W. Meiosis and haploid gametes in the pathogen *Trypanosoma brucei*. *Curr Biol*. 2014; 24:1–6. <https://doi.org/10.1016/j.cub.2013.11.017> PMID: 24332542
4. Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, Carrington M, et al. Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. *Proc Natl Acad Sci USA*. 2011; 108(9):3671–6. <https://doi.org/10.1073/pnas.1019423108> PMID: 21321215
5. Hope M, MacLeod A, Leech V, Melville S, Sasse J, Tait A, et al. Analysis of ploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange. *Mol Biochem Parasitol*. 1999; 104(1):1–9. ISI:000083821100001. [https://doi.org/10.1016/S0166-6851\(99\)00103-6](https://doi.org/10.1016/S0166-6851(99)00103-6) PMID: 10589977
6. MacLeod A, Tweedie A, McLellan S, Taylor S, Cooper A, Sweeney L, et al. Allelic segregation and independent assortment in *Trypanosoma brucei* crosses: Proof that the genetic system is Mendelian and involves meiosis. *Mol Biochem Parasitol*. 2005; 143(1):12–9. <https://doi.org/10.1016/j.molbiopara.2005.04.009> PMID: 15941603
7. Gibson W, Garside L, Bailey M. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Mol Biochem Parasitol*. 1992; 52:189–200. [https://doi.org/10.1016/0166-6851\(92\)90069-v](https://doi.org/10.1016/0166-6851(92)90069-v) PMID: 1349422
8. Paindavoine P, Zampetti-Bosseler F, Pays E, Schweizer J, Guyaux M, Jenni L, et al. Trypanosome hybrids generated in tsetse flies by nuclear fusion. *EMBO J*. 1986; 5:3631–6. PMID: 3830130

9. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, Taylor MC, et al. Mechanism of genetic exchange in American trypanosomes. *Nature*. 2003; 421(6926):936–9. doi: [10.1038/nature01438](https://doi.org/10.1038/nature01438). ISI:000181186900047. PMID: [12606999](https://pubmed.ncbi.nlm.nih.gov/12606999/)
10. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, et al. Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science*. 2009; 324(5924):265–8. doi: [10.1126/science.1169464](https://doi.org/10.1126/science.1169464). WOS:000265024400050. PMID: [19359589](https://pubmed.ncbi.nlm.nih.gov/19359589/)
11. Inbar E, Akopyants NS, Charmoy M, Romano A, Lawyer P, Elnaiem D-EA, et al. The mating competence of geographically diverse *Leishmania major* strains in their natural and unnatural sand fly vectors. *Plos Genetics*. 2013; 9(7). WOS:000322321100058. <https://doi.org/10.1371/journal.pgen.1003672> PMID: [23935521](https://pubmed.ncbi.nlm.nih.gov/23935521/)
12. Inbar E, Shaik J, Iantorno SA, Romano A, Nzelu CO, Owens K, et al. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in *Leishmania*. *Plos Genetics*. 2019; 15(5):e1008042. doi: [10.1371/journal.pgen.1008042](https://doi.org/10.1371/journal.pgen.1008042). WOS:000470208000009. PMID: [31091230](https://pubmed.ncbi.nlm.nih.gov/31091230/)
13. Romano A, Inbar E, Debrabant A, Charmoy M, Lawyer P, Ribeiro-Gomes F, et al. Cross-species genetic exchange between visceral and cutaneous strains of *Leishmania* in the sand fly vector. *Proc Natl Acad Sci USA*. 2014; 111:16808–13. <https://doi.org/10.1073/pnas.1415109111> PMID: [25385616](https://pubmed.ncbi.nlm.nih.gov/25385616/)
14. Calvo-Álvarez E, Álvarez-Velilla R, Jiménez M, Molina R, Pérez-Pertejo Y, Balaña-Fouce R, et al. First evidence of intraclonal genetic exchange in trypanosomatids using two *Leishmania infantum* fluorescent transgenic clones. *PLoS NTD*. 2014; 8 e3075. <https://doi.org/10.1371/journal.pntd.0003075> PMID: [25188587](https://pubmed.ncbi.nlm.nih.gov/25188587/)
15. Louradour I, Ferreira TR, Ghosh K, Shaik J, Sacks D. In vitro generation of *Leishmania* hybrids. *Cell Reports*. 2020; 31(2):107507. doi: [10.1016/j.celrep.2020.03.071](https://doi.org/10.1016/j.celrep.2020.03.071). WOS:000526938500017. PMID: [32294444](https://pubmed.ncbi.nlm.nih.gov/32294444/)
16. Lukes J, Guilbride DL, Votycka J, Zikova A, Benne R, Englund PT. Kinetoplast DNA network: Evolution of an improbable structure. *Euk Cell*. 2002; 1(4):495–502. doi: [10.1128/EC.1.4.495-502.2002](https://doi.org/10.1128/EC.1.4.495-502.2002). ISI:000178729400001. PMID: [12455998](https://pubmed.ncbi.nlm.nih.gov/12455998/)
17. Jensen RE, Englund PT. Network News: The Replication of Kinetoplast DNA. *Annu Rev Microbiol*. 2012; 66:473–91. doi: [10.1146/annurev-micro-092611-150057](https://doi.org/10.1146/annurev-micro-092611-150057). WOS:000310143500024. PMID: [22994497](https://pubmed.ncbi.nlm.nih.gov/22994497/)
18. Stuart KD, Schnauer A, Ernst NL, Panigrahi AK. Complex management: RNA editing in trypanosomes. *Trends in Biochemical Sciences*. 2005; 30(2):97–105. <https://doi.org/10.1016/j.tibs.2004.12.006> ISSN 0968-0004. WOS:000227299300007. PMID: [15691655](https://pubmed.ncbi.nlm.nih.gov/15691655/)
19. Gibson W, Crow M, Kearns J. Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of *Trypanosoma brucei*. *Parasitol Res*. 1997; 83:483–8. <https://doi.org/10.1007/s004360050284> PMID: [9197397](https://pubmed.ncbi.nlm.nih.gov/9197397/)
20. Gibson W, Garside L. Kinetoplast DNA mini-circles are inherited from both parents in genetic hybrids of *Trypanosoma brucei*. *Mol Biochem Parasitol*. 1990; 42:45–54. [https://doi.org/10.1016/0166-6851\(90\)90111-x](https://doi.org/10.1016/0166-6851(90)90111-x) PMID: [2233899](https://pubmed.ncbi.nlm.nih.gov/2233899/)
21. Turner CMR, Hide G, Buchanan N, Tait A. *Trypanosoma brucei*—inheritance of kinetoplast DNA maxicircles in a genetic cross and their segregation during vegetative growth. *Expt Parasitol*. 1995; 80:234–41. <https://doi.org/10.1006/expr.1995.1029> PMID: [7895834](https://pubmed.ncbi.nlm.nih.gov/7895834/)
22. Van den Broeck F, Savill NJ, Imamura H, Sanders M, Maes I, Cooper S, et al. Ecological divergence and hybridization of Neotropical *Leishmania* parasites. *Proc Natl Acad Sci USA*. 2020; 117(40):25159–68. doi: [10.1073/pnas.1920136117](https://doi.org/10.1073/pnas.1920136117). WOS:000579045200021. PMID: [32958676](https://pubmed.ncbi.nlm.nih.gov/32958676/)
23. Shapiro TA, Englund PT. The structure and replication of kinetoplast DNA. *Annu Rev Microbiol*. 1995; 49:117–43. ISI:A1995TB31900006. <https://doi.org/10.1146/annurev.mi.49.100195.001001> PMID: [8561456](https://pubmed.ncbi.nlm.nih.gov/8561456/)
24. Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: The population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas* and *Trypanosoma* and their medical and taxonomical consequences. *Proc Natl Acad Sci USA*. 1990; 87:2414–8. <https://doi.org/10.1073/pnas.87.7.2414> PMID: [2320563](https://pubmed.ncbi.nlm.nih.gov/2320563/)
25. Mathieu-Daude F, Stevens JR, Welsh J, Tibayrenc M, McClelland M. Genetic diversity and population structure of *Trypanosoma brucei*: clonality versus sexuality. *Mol Biochem Parasitol*. 1995; 72:89–101. [https://doi.org/10.1016/0166-6851\(95\)00083-d](https://doi.org/10.1016/0166-6851(95)00083-d) PMID: [8538703](https://pubmed.ncbi.nlm.nih.gov/8538703/)
26. Tibayrenc M, Ayala FJ. How clonal are *Trypanosoma* and *Leishmania*? *Trends Parasitol*. 2013; 29(6):264–9. WOS:000320477600002. <https://doi.org/10.1016/j.pt.2013.03.007> PMID: [23602631](https://pubmed.ncbi.nlm.nih.gov/23602631/)
27. Gibson W, Peacock L, Ferris V, Fischer K, Livingstone J, Thomas J, et al. Genetic recombination between human and animal parasites creates novel strains of human pathogen. *PLoS NTD*. 2015; 9(3):e0003665. <https://doi.org/10.1371/journal.pntd.0003665> PMID: [25816228](https://pubmed.ncbi.nlm.nih.gov/25816228/)

28. Balmer O, Beadell JS, Gibson W, Caccone A. Phylogeography and taxonomy of *Trypanosoma brucei*. PLoS NTD. 2011; 5(2):e961. <https://doi.org/10.1371/journal.pntd.0000961> PMID: 21347445
29. Goodhead I, Capewell P, Bailey JW, Beament T, Chance M, Kay S, et al. Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence. MBio. 2013; 4(4):e00197–213. doi: [10.1128/mBio.00197-13](https://doi.org/10.1128/mBio.00197-13). WOS:000326881100001. PMID: 23963174
30. Sistrom M, Evans B, Bjornson R, Gibson W, Balmer O, Mäser P, et al. Comparative genomics reveals multiple genetic backgrounds of human pathogenicity in the *Trypanosoma brucei* complex. Genome Biol Evol. 2014; 6(10):2811–9. <https://doi.org/10.1093/gbe/evu222> PMID: 25287146
31. Simpson JT. Exploring genome characteristics and sequence quality without a reference. Bioinformatics. 2014; 30:1228–35. <https://doi.org/10.1093/bioinformatics/btu023> PMID: 24443382
32. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: A K-Mer analysis toolkit to quality control NGS datasets and genome assemblies Bioinformatics. 2017; 33:574–6. <https://doi.org/10.1093/bioinformatics/btw663> PMID: 27797770
33. Müller LSM, Cosentino RO, Förstner KU, Guizzetti J, Wedel C, Kaplan N, et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature. 2018; 563:121–5. <https://doi.org/10.1038/s41586-018-0619-8> PMID: 30333624
34. Tait A, Buchanan N, Hide G, Turner M. Self-fertilisation in *Trypanosoma brucei*. Mol Biochem Parasitol. 1996; 76:31–42. [https://doi.org/10.1016/0166-6851\(95\)02528-6](https://doi.org/10.1016/0166-6851(95)02528-6) PMID: 8919993
35. Peacock L, Ferris V, Bailey M, Gibson W. Intraclonal mating occurs during tsetse transmission of *Trypanosoma brucei*. Parasit Vectors. 2009; 2(1):43. doi: [10.1186/1756-3305-2-43](https://doi.org/10.1186/1756-3305-2-43). MEDLINE:PMID: 19772562.
36. Gibson W, Winters K, Mizen G, Kearns J, Bailey M. Intraclonal mating in *Trypanosoma brucei* is associated with out-crossing. Microbiology. 1997; 143:909–20. <https://doi.org/10.1099/00221287-143-3-909> PMID: 9084175
37. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Comms. 2020; 11:1432. <https://doi.org/10.1038/s41467-020-14998-3> PMID: 32188846
38. Tihon E, Imamura H, Dujardin JC, Van Den Abbeele J, Van den Broeck F. Discovery and genomic analyses of hybridization between divergent lineages of *Trypanosoma congolense*, causative agent of Animal African Trypanosomiasis. Mol Ecol. 2017; 26: 6524–6538. <https://doi.org/10.1111/mec.14271> PMID: 28752916
39. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science. 2005; 309(5733):416–22. ISI:000230574900036. <https://doi.org/10.1126/science.1112642> PMID: 16020726
40. Peacock L, Kay C, Farren C, Bailey M, Carrington M, Gibson W. Sequential production of gametes during meiosis in trypanosomes. Communications Biology. 2021; 4:555. <https://doi.org/10.1038/s42003-021-02058-5> PMID: 33976359
41. Peacock L, Ferris V, Bailey M, Gibson W. Mating compatibility in the parasitic protist *Trypanosoma brucei*. Parasit Vectors. 2014; 7:78. <https://doi.org/10.1186/1756-3305-7-78> PMID: 24559099
42. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics. 2001; 17:282–3. <https://doi.org/10.1093/bioinformatics/17.3.282> PMID: 11294794
43. Borst P, Cross GAM. Molecular basis for trypanosome antigenic variation. Cell. 1982; 29:291–303. [https://doi.org/10.1016/0092-8674\(82\)90146-5](https://doi.org/10.1016/0092-8674(82)90146-5) PMID: 6811137
44. Graham SV, Terry S, Barry JD. A structural and transcription pattern for variant surface glycoprotein gene expression sites used in metacyclic stage *Trypanosoma brucei*. Mol Biochem Parasitol. 1999; 103(2):141–54. ISI:000083264600002. [https://doi.org/10.1016/s0166-6851\(99\)00128-0](https://doi.org/10.1016/s0166-6851(99)00128-0) PMID: 10551359
45. Bringaud F, Biteau N, Donelson JE, Baltz T. Conservation of metacyclic variant surface glycoprotein expression sites among different trypanosome isolates. Mol Biochem Parasitol. 2001; 113(1):67–78. ISI:000167753200007. [https://doi.org/10.1016/s0166-6851\(00\)00381-9](https://doi.org/10.1016/s0166-6851(00)00381-9) PMID: 11254955
46. Ginger ML, Blundell PA, Lewis AM, Browitt A, Gunzl A, Barry JD. Ex vivo and in vitro identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. Euk Cell. 2002; 1(6):1000–9. doi: [10.1128/EC.1.6.1000-1009.2002](https://doi.org/10.1128/EC.1.6.1000-1009.2002). ISI:000179722200016. PMID: 12477800
47. Cross GAM, Kim HS, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. Mol Biochem Parasitol. 2014; 195(1):59–73. doi: [10.1016/j.molbiopara.2014.06.004](https://doi.org/10.1016/j.molbiopara.2014.06.004). WOS:249920427100010.

48. Muller LSM, Cosentino RO, Forstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*. 2018; 563(7729):121–+. doi: [10.1038/s41586-018-0619-8](https://doi.org/10.1038/s41586-018-0619-8). WOS:000448900900057. PMID: [30333624](https://pubmed.ncbi.nlm.nih.gov/30333624/)
49. Kolev NG, Gunzl A, Tschudi C. Metacyclic VSG expression site promoters are recognized by the same general transcription factor that is required for RNA polymerase I transcription of bloodstream expression sites. *Mol Biochem Parasitol*. 2017; 216:52–5. doi: [10.1016/j.molbiopara.2017.07.002](https://doi.org/10.1016/j.molbiopara.2017.07.002). WOS:000411771100009. PMID: [28716719](https://pubmed.ncbi.nlm.nih.gov/28716719/)
50. Howick VM, Peacock L, Kay C, Collett C, Gibson W, Lawniczak MKN. Single-cell transcriptomics reveals expression profiles of *Trypanosoma brucei* sexual stages. *BioRxiv*. 2021. <https://www.biorxiv.org/content/10.1101/2021.10.13.463681v1>.
51. Birky CW. Relaxed cellular controls and organelle heredity. *Science*. 1983; 222:468–75. <https://doi.org/10.1126/science.6353578> PMID: [6353578](https://pubmed.ncbi.nlm.nih.gov/6353578/)
52. Kay C, Williams TA, Gibson W. Mitochondrial DNAs provide insight into trypanosome phylogeny and molecular evolution. *BMC Evolutionary Biology*. 2020; 20:161. <https://doi.org/10.1186/s12862-020-01701-9> PMID: [33297939](https://pubmed.ncbi.nlm.nih.gov/33297939/)
53. Borst P, Fase-Fowler F, Gibson WC. Kinetoplast DNA of *Trypanosoma evansi*. *Mol Biochem Parasitol*. 1987; 23:31–8. [https://doi.org/10.1016/0166-6851\(87\)90184-8](https://doi.org/10.1016/0166-6851(87)90184-8) PMID: [3033499](https://pubmed.ncbi.nlm.nih.gov/3033499/)
54. Gibson WC, Wilson AJ, Moloo SK. Characterisation of *Trypanosoma (Trypanozoon) evansi* from camels in Kenya using isoenzyme electrophoresis. *Res Vet Sci*. 1983; 34:114–8. PMID: [6836172](https://pubmed.ncbi.nlm.nih.gov/6836172/)
55. Ngaira JM, Olembo NK, Njagi ENM, Ngeranwa JJN. The detection of non-RoTat 1.2 *Trypanosoma evansi*. *Expt Parasitol*. 2005; 110(1):30–8. WOS:000228318600004. <https://doi.org/10.1016/j.exppara.2005.01.001> PMID: [15804376](https://pubmed.ncbi.nlm.nih.gov/15804376/)
56. Njiru ZK, Constantine CC, Masiga DK, Reid SA, Thompson RCA, Gibson WC. Characterization of *Trypanosoma evansi* type B. *Infect Genet Evol*. 2006; 6(4):292–300. WOS:000238717400006. <https://doi.org/10.1016/j.meegid.2005.08.002> PMID: [16157514](https://pubmed.ncbi.nlm.nih.gov/16157514/)
57. Cuypers B, Van den Broeck F, Van Reet N, Meehan CJ, Cauchard J, Wilkes JM, et al. Genome-wide SNP analysis reveals distinct origins of *Trypanosoma evansi* and *Trypanosoma equiperdum*. *Genome Biol Evol*. 2017; 9(8):1990–7. <https://doi.org/10.1093/gbe/evx102> PMID: [28541535](https://pubmed.ncbi.nlm.nih.gov/28541535/)
58. Richardson JB, Lee K-Y, Mireji P, Enyaru J, Sistro M, Aksoy S, et al. Genomic analyses of African *Trypanozoon* strains to assess evolutionary relationships and identify markers for strain identification. *PLoS Negl Trop Dis*. 2017; 11:e0005949. <https://doi.org/10.1371/journal.pntd.0005949> PMID: [28961238](https://pubmed.ncbi.nlm.nih.gov/28961238/)
59. Gibson WC, Gashumba JK. Isoenzyme characterisation of some *Trypanozoon* stocks from a recent trypanosomiasis epidemic in Uganda. *Trans Roy Soc Trop Med Hyg*. 1983; 77:114–8. [https://doi.org/10.1016/0035-9203\(83\)90033-0](https://doi.org/10.1016/0035-9203(83)90033-0) PMID: [6679356](https://pubmed.ncbi.nlm.nih.gov/6679356/)
60. Enyaru JC, Odiit M., Gashumba J.K., Carasco J.F., Rwendeire A.J. Characterization by isoenzyme electrophoresis of *Trypanozoon* stocks from sleeping sickness endemic areas of south-east Uganda. *Bull WHO*. 1992; 70:631–6. PMID: [1464150](https://pubmed.ncbi.nlm.nih.gov/1464150/)
61. Enyaru JC, Stevens JR, Odiit M, Okuna NM, Carasco JF. Isoenzyme comparison of *Trypanozoon* isolates from two sleeping sickness areas of south-eastern Uganda. *Acta Trop*. 1993; 55:97–115. [https://doi.org/10.1016/0001-706x\(93\)90072-j](https://doi.org/10.1016/0001-706x(93)90072-j) PMID: [7903841](https://pubmed.ncbi.nlm.nih.gov/7903841/)
62. Hide G, Angus SD, Holmes PH, Maudlin I, Welburn SC. *Trypanosoma brucei*: comparison of circulating strains in an endemic and epidemic area of a sleeping sickness focus. *Expt Parasitol*. 1998; 89:21–9.
63. Echodu R, Sistro M, Bateta R, Murilla G, Okedi L, Aksoy S, et al. Genetic diversity and population structure of *Trypanosoma brucei* in Uganda: Implications for the epidemiology of sleeping sickness and nagana. *PLoS NTD*. 2015; 9(2):e0003353. <https://doi.org/10.1371/journal.pntd.0003353> PMID: [25695634](https://pubmed.ncbi.nlm.nih.gov/25695634/)
64. Xong VH, Vanhamme L, Chamekh M, Chimfwembe CE, Van den Abbeele J, Pays A, et al. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell*. 1998; 95:839–46. [https://doi.org/10.1016/s0092-8674\(00\)81706-7](https://doi.org/10.1016/s0092-8674(00)81706-7) PMID: [9865701](https://pubmed.ncbi.nlm.nih.gov/9865701/)
65. Gibson WC, Wellde BT. Characterisation of *Trypanozoon* stocks from the South Nyanza sleeping sickness focus in Western Kenya. *Trans Roy Soc Trop Med Hyg*. 1985; 79:671–6. [https://doi.org/10.1016/0035-9203\(85\)90187-7](https://doi.org/10.1016/0035-9203(85)90187-7) PMID: [4095748](https://pubmed.ncbi.nlm.nih.gov/4095748/)
66. Gibson WC, Garside LH. Genetic exchange in *Trypanosoma brucei brucei*: variable location of house-keeping genes in different trypanosome stocks. *Mol Biochem Parasitol*. 1991; 45:77–90. [https://doi.org/10.1016/0166-6851\(91\)90029-6](https://doi.org/10.1016/0166-6851(91)90029-6) PMID: [1675762](https://pubmed.ncbi.nlm.nih.gov/1675762/)
67. Peacock L, Ferris V, Bailey M, Gibson W. Fly transmission and mating of *Trypanosoma brucei brucei* strain 427. *Mol Biochem Parasitol*. 2008; 160:100–6. <https://doi.org/10.1016/j.molbiopara.2008.04.009> PMID: [18524395](https://pubmed.ncbi.nlm.nih.gov/18524395/)

68. Turner CM, Aslam N., Smith E., Buchanan N., Tait A. The effects of genetic exchange on variable antigen expression in *Trypanosoma brucei*. *Parasitology*. 103 Pt 31991. p. 379–86. <https://doi.org/10.1017/s0031182000059898> PMID: 1780175
69. Gibson WC. Will the real *Trypanosoma brucei gambiense* please stand up? *Parasitol Today*. 1986; 2:255–7. [https://doi.org/10.1016/0169-4758\(86\)90011-6](https://doi.org/10.1016/0169-4758(86)90011-6) PMID: 15462856
70. Lun ZR, Brun R, Gibson WC. Kinetoplast DNA and molecular karyotypes of *Trypanosoma evansi* and *T. equiperdum* from China. *Mol Biochem Parasitol*. 1992; 50:189–96. [https://doi.org/10.1016/0166-6851\(92\)90215-6](https://doi.org/10.1016/0166-6851(92)90215-6) PMID: 1311051
71. Morrison LJ, Tait A, McCormack G, Sweeney L, Black A, Truc P, et al. *Trypanosoma brucei gambiense* Type 1 populations from human patients are clonal and display geographical genetic differentiation. *Infect Genet Evol*. 2008; 8(6):847–54. doi: [10.1016/j.meegid.2008.08.005](https://doi.org/10.1016/j.meegid.2008.08.005). WOS:000261596000012. PMID: 18790085
72. Koffi M, De Meeus T, Bucheton B, Solano P, Camara M, Kaba D, et al. Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. *Proc Natl Acad Sci USA*. 2009; 106(1):209–14. doi: [10.1073/pnas.0811080106](https://doi.org/10.1073/pnas.0811080106). WOS:000262263900040. PMID: 19106297
73. Godfrey DG, Baker RD, Rickman LR, Mehlitz D. The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. *Adv Parasitol*. 1990; 29:1–74. [https://doi.org/10.1016/s0065-308x\(08\)60104-9](https://doi.org/10.1016/s0065-308x(08)60104-9) PMID: 2181826
74. Cunningham I. New culture medium for maintenance of tsetse tissues and growth of trypanosomatids. *J Protozool*. 1977; 24:325–9. <https://doi.org/10.1111/j.1550-7408.1977.tb00987.x> PMID: 881656
75. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012; 19:455–77. <https://doi.org/10.1089/cmb.2012.0021> PMID: 22506599
76. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018; 34:i884–i90. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086
77. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011; 27:2987–93. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627
78. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files *Bioinformatics*. 2011; 27:718–9. <https://doi.org/10.1093/bioinformatics/btq671> PMID: 21208982
79. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, et al. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*. 2013; 14:193–202. <https://doi.org/10.1093/bib/bbs012> PMID: 22445902
80. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 2012; 28:464–9. <https://doi.org/10.1093/bioinformatics/btr703> PMID: 22199388
81. Kay C, Peacock L, Williams TA, Gibson W (2022) Data from: Signatures of hybridisation in *Trypanosoma brucei*. Dryad Digital Repository. <https://doi.org/https://doi.org/10.5061/dryad.xd2547djb>
82. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*. 2013; 8:1494–512. <https://doi.org/10.1038/nprot.2013.084> PMID: 23845962
83. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011; 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195> PMID: 22039361
84. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 2013; 29:2487–9. <https://doi.org/10.1093/bioinformatics/btt403> PMID: 23842809
85. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics*. 2019; 20:405. <https://doi.org/10.1186/s12859-019-2996-x> PMID: 31345161
86. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
87. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30:772–80. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
88. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*. 2016; 2:e000056. <https://doi.org/10.1099/mgen.0.000056> PMID: 28348851
89. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168

90. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2019; 15: e1006650. <https://doi.org/10.1371/journal.pcbi.1006650> PMID: 30958812
91. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018; 67:901–4. <https://doi.org/10.1093/sysbio/syy032> PMID: 29718447