# Genome-Wide SNP Analysis Reveals Distinct Origins of Trypanosoma evansi and Trypanosoma equiperdum.

# Genome-Wide SNP Analysis Reveals Distinct Origins of *Trypanosoma evansi* and *Trypanosoma equiperdum*

Bart Cuypers[1,2], Frederik Van den Broeck[1], Nick Van Reet[1], Conor J. Meehan[1], Julien Cauchard[3], Jonathan M. Wilkes[4], Filip Claes[5], Bruno Goddeeris[6], Hadush Birhanu[7], Jean-Claude Dujardin[1,8], Kris Laukens[2], Philippe Büscher[1], and Stijn Deborggraeve[1,*]

[1]Biomedical Sciences Department, Institute of Tropical Medicine, Antwerp, Belgium

[2]Department of Mathematics and Computer Sciences, University of Antwerp, Belgium

[3]Anses Dozulé Laboratory for Equine Diseases, Goustranville, France

[4]Wellcome Trust Centre of Molecular Parasitology, University of Glasgow, United Kingdom

[5]Food and Agriculture Organization of the United Nations (FAO), Regional Office for Asia and the Pacific, Bangkok, Thailand

[6]Faculty of Bioscience Engineering, KU Leuven, Belgium

[7]College of Veterinary Medicine, Mekelle University, Tigray, Ethiopia

[8]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

*Corresponding author: E-mail: sdeborggraeve@itg.be.

Accepted: May 24, 2017

## Abstract

Trypanosomes cause a variety of diseases in man and domestic animals in Africa, Latin America, and Asia. In the *Trypanozoon* subgenus, *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense* cause human African trypanosomiasis, whereas *Trypanosoma brucei brucei*, *Trypanosoma evansi*, and *Trypanosoma equiperdum* are responsible for nagana, surra, and dourine in domestic animals, respectively. The genetic relationships between *T. evansi* and *T. equiperdum* and other *Trypanozoon* species remain unclear because the majority of phylogenetic analyses has been based on only a few genes. In this study, we have conducted a phylogenetic analysis based on genome-wide SNP analysis comprising 56 genomes from the *Trypanozoon* subgenus. Our data reveal that *T. equiperdum* has emerged at least once in Eastern Africa and *T. evansi* at two independent occasions in Western Africa. The genomes within the *T. equiperdum* and *T. evansi* monophyletic clusters show extremely little variation, probably due to the clonal spread linked to the independence from tsetse flies for their transmission.

**Key words:** *Trypanosoma evansi*, *Trypansoma equiperdum*, *Trypanozoon*, whole genome sequencing, SNP analysis, phylogeny.

## Introduction

African trypanosomes are important causes of disease in man and domestic animals in Africa, Latin America, and Asia. The majority of the pathogenic African trypanosomes belongs to the *Nannomonas* (e.g., *Trypanosoma congolense*), *Dutonella* (e.g., *Trypanosoma vivax*), and *Trypanozoon* subgenera, comprising the species *Trypanosoma brucei*, *Trypanosoma evansi*, and *Trypanosoma equiperdum*. *Trypanosoma brucei* has been divided into three subspecies. Whereas *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense* are responsible for human African trypanosomiasis (HAT) in Western/Central and Eastern/Southern Africa, respectively, *Trypanosoma brucei brucei* causes nagana in cattle, horses and small ruminants. *Trypanosoma evansi* is responsible for surra in camels, horses, and other domestic animals and *T. equiperdum* seems to be restricted to horses and other Equidae like donkeys and mules, causing dourine. This classification was established in the early 1900s and is based on morphology, geographical distribution, clinical presentation of the disease, and affected host species (Hoare 1972). In recent decades, *T. b. gambiense* has been further subdivided into group 1 and 2 (Gibson 1986). The former is a monophyletic group comprising most of the *T. b. gambiense* strains (Weir et al. 2015), whereas the latter is a heterogeneous

and rather ambiguous group of trypanosomes that have been isolated from patients in Côte d'Ivoire in the late 1970s (Truc et al. 1997). *Trypanosoma evansi* has also been subdivided in the types A and B based on minicircle kDNA (Borst et al. 1987), and *T. evansi* type A is characterized by the immunodominant RoTat 1.2 variable surface glycoprotein (VSG; Bajyana Songa and Hamers 1988).

The three *T. brucei* subspecies undergo a complex cycle of differentiation and multiplication in the tsetse fly but this is not the case for *T. evansi* and *T. equiperdum*. *Trypanosoma evansi* is not transmitted by tsetse flies but mechanically by a wide range of biting flies, vampire bats and by ingestion of raw meat, whereas *T. equiperdum* is transmitted sexually (Desquesnes et al. 2013a, 2013b; Brun et al. 1998). With transmission independent of tsetse flies, which are present only in sub-Saharan Africa, *T. evansi* and *T. equiperdum* have spread outside Africa and have even lost the ability to develop in the fly. Survival of *T. brucei* inside the tsetse fly requires the expression of numerous mitochondrial genes for oxidative phosphorylation during ATP production (Bringaud et al. 2006). Trypanosomes have a single mitochondrion or 'kinetoplast', which is organized in a complex network of interlocked DNA circles with several thousands of minicircles and a few dozen of maxicircles. *Trypanosoma equiperdum* retained parts of its maxicircle kDNA, whereas *T. evansi* has completely lost the maxicircle DNA (Lai et al. 2008). The kinetoplast DNA encodes proteins that are essential for the survival of the trypanosomes in the blood, such as subunits of respiratory chain complexes, in order to maintain the electrochemical potential across the inner mitochondrial membrane (Neupert et al. 1997; Bertrand et al. 2000). To compensate for their incapability to express functional F1-ATP synthase subunit $\alpha$ from their mitochondrial genome, some of these dyskinetoplastid trypanosomes have acquired compensatory mutations in the F1-ATP synthase subunit $\gamma$ encoded in the nuclear DNA (Schnaufer et al. 2005; Lai et al. 2008). Dean et al. (2013) recently showed that single amino acid changes in the $\gamma$ subunit can fully compensate for the loss of maxicircle kDNA in *T. b. brucei*.

The genetic relationships between *T. evansi*, *T. equiperdum* and the different pathogenic African trypanosomes are unclear since the majority of phylogenetic analyses has been based on only a few genes (Balmer et al. 2011). By analyzing microsatellites and the genes coding for cytochrome oxidase 1 (COX1), dihydrolipamide (LipDH), and F1-ATP $\gamma$-subunit, Carnes et al. showed that *T. evansi* and *T. equiperdum* originated at least at four independent occasions from *T. brucei* (2015). Recently, the first genome of *T. equiperdum* (reference strain OVI) was sequenced and annotated (Hebert et al. 2017). In the present study, we have sequenced three additional *T. equiperdum* and three *T. evansi* isolates and conducted a phylogenetic analysis using a global collection of 56 genomes of *T. evansi*, *T. equiperdum*, *T. b. brucei*, *T. b. gambiense*, and *T. b. rhodesiense*. Our data reveal that the *T. equiperdum*, *T. evansi* type A, and *T. evansi* type B genomes form distinct but highly homogenous clusters, whereby *T. equiperdum* is more related to *T. brucei* from Eastern Africa and *T. evansi* to *T. brucei* from Western Africa.

## Materials and Methods

### Sequence Data

Sequence data of three *T. b. gambiense* group 1, three *T. b. gambiense* group 2, 17 *T. b. rhodesiense*, 21 *T. b. brucei*, eight *T. evansi*, and four *T. equiperdum* strains were either downloaded from publicly available sources (Sistrom et al. 2014; Berriman et al. 2005; Carnes et al. 2015; Goodhead et al. 2013; Jackson et al. 2010; Hebert et al. 2017), or generated in this study. The complete list of 56 strains with information on the country and year of isolation is presented in supplementary table S1, Supplementary Material online. In this study, genomes were sequenced of the *T. b. gambiense* group 2 strain ABBA, *T. evansi* type A strains RoTat 1.2 and MCAM/ET/2013/MU/09 (MU09), *T. evansi* type B strain MCAM/ET/2013/MU/10 (MU10) and the *T. equiperdum* strains BoTat, Dodola 943, and TeAp-N/D1 (Hide et al. 1990, Bajyana Songa & Hamers 1988; Birhanu et al. 2016; Claes et al. 2004, Hagos et al. 2010; Perrone et al. 2009; Sánchez et al. 2015b; Sánchez et al. 2015a; Capbern et al. 1977; Mehlitz et al. 1982). The study protocol for in vivo expansion of trypanosomes was approved by the animal experimentation ethics committee of ITM (BM2013-7). This protocol follows the European Commission Recommendation on guidelines for the accommodation and care of animals used for experimental and other scientific purposes (June 18, 2007, 2007/526/EG) and the Belgian National law on the protection of animals under experiment. Cryostabilates containing bloodstream forms of the strains sequenced in this study were thawed at 37 °C, diluted in phosphate buffered saline (PSG, pH 8.0), containing 1% glucose, and $2 \times 10^5$ trypanosomes were injected intraperitoneal in 25–30 g female OF-1 mice. Parasitemia was measured daily using the matching method (Herbert and Lumsden 1976). At the first peak of the parasitemia, usually after 3–4 days postinfection, the animal was anesthetized and blood was collected with a heparinized needle. Parasites were purified from the blood using anion exchange chromatography (Lanham and Godfrey 1970), washed twice in PSG and sedimented as pellets of pure trypanosomes through centrifugation at $1{,}500 \times g$ for 10 min. Trypanosome pellets were stored at −80 °C until DNA extraction using the QIAamp DNA mini kit (Qiagen, Hilden, Germany). Sequencing libraries of the *T. equiperdum* strains Dodola 943, TeAp-N/D1, BoTat, OVI and *T. evansi* strain RoTat 1.2 were prepared with the NEBNext® DNA Library Prep Reagent Set. Clean-up of the library was performed with Qiagen columns and size selection

(~350 bp) by excision from a gel. The libraries were multiplexed, pooled and 2 × 76 base pair (bp) reads were generated on two lanes with the Illumina GAIIX. Sequencing libraries of the Ethiopian *T. evansi* strains MU09 and MU10 and the *T. b. gambiense* type 2 strain ABBA were prepared with Illumina's Nextera XT sample preparation kit and 2 × 300 bp reads were generated with the Illumina MiSeq (ABBA) or 2 × 100 bp reads with the HiSeq 1500 (MU09 and MU10). Quality scored sequence reads have been submitted to the NCBI Sequence Read Archive (SRA) with accession numbers SRP100981 and SRP100990.

### Genome-Wide SNP Analysis

All 56 genomes were aligned to the *T. b. brucei* TREU927 reference genome with Bowtie2 using default alignment parameters (Langmead and Salzberg 2012). The resulting SAM files were converted to bam and sorted by position using samtools (Li et al. 2009). Variants were called with Samtools' mpileup and the resulting vcf file was filtered to retain only SNPs with read depths of at least 5 in all strains, a SNP quality of at least 50, a mapping quality of at least 40 and no significant strand bias ($P > 0.05$). The high mapping quality cut-off prevents repetitive regions or unreliably mapped reads from causing false positive SNPs.

For each strain, the number of heterozygous SNPs was determined and SNPs were concatenated to a FASTA sequence. Invariant sites were removed from this SNP alignment and RaxML version 8.1.17 (Stamatakis 2014) was employed to construct a maximum likelihood tree with 1,000 bootstrap replicate support. The phylogenetic tree was created using a GTR CAT model of substitution with Felsenstein correction for the ascertainment bias inherent in SNP alignments, and visualized using FigTree version 1.4.2. In order to detect conflicting phylogenetic signals, a split network was constructed using SplitsTree (Huson and Bryant 2006). Finally, we generated phased haplotype data using SHAPEIT v2 (Delaneau et al. 2013) using the following parameters: 10 burn steps, 10 prune steps, 50 main steps, 200 conditioning states, and window size of 100 kb; for the remainder of the parameters we kept the default values. The phased SNP data were used to infer population genetic structure with fineSTRUCTURE v2 (Lawson et al. 2012) using default parameters.

### Human Serum Resistance Genes, Diagnostic VSG's, and F1-ATP Synthase Subunit Gamma

For all genomes, we have investigated the presence and sequence of the trypanosomes' human resistance genes, coding for the *T. b. rhodesiense* specific Serum Resistance Associated (SRA) protein (Xong et al. 1998), the *T. b. gambiense* specific glycoprotein (TgsGP; Uzureau et al. 2013) and haptoglobin–hemoglobin (TbHpHb) receptor (Vanhollebeke et al. 2008). FASTA sequences of the reference genes (Genbank accession number Z37159.2, FN555988, AF317914, and AY007705.1, AY007706.1 and AJ870487, respectively) were downloaded and the reads from the 56 sequenced genomes were mapped against these references sequences. Bowtie2 was used in local mode to cover both the beginning and the end of the genes, except for the VSG's that required the full read sequence alignment given their repetitive nature. Following alignment, the frequency of each nucleobase per position was calculated and nucleobases with a frequency of minimum 0.25 for a certain position were retained and represented as ambiguity bases, where necessary. Positions with a coverage <10 were represented as "N". Presence or absence of genes was manually verified in the alignments using Integrative Genomics Viewer (IGV). Genes were considered present if there were mapping reads for each nucleobase position in the gene. Sequences of the F1-ATPase subunit $\gamma$ genes, as well as the corresponding amino acids, were aligned using Clustal Omega (Sievers et al. 2011).

## Results

### Genome-Wide SNP Analysis

All genomes had an average coverage of at least 29.6×, except DAL972 which had 5.7× coverage and was the only genome in the data set sequenced with the Sanger method (supplementary table S2, Supplementary Material online). In total, 890,170 SNPs were called in the genomes of the 56 *Trypanosoma* strains and 194,566 passed our filtering criteria. The filtered SNPs were used to construct a Neighbournet network (fig. 1), a haplotype-based clustering analysis (fig. 2), and a RAxML Maximum Likelihood tree (supplementary fig. S1, Supplementary Material online).

All three analyses revealed a similar evolutionary history for *T. evansi* and *T. equiperdum*. The 6 *T. evansi* type A strains form a monophyletic cluster and exhibit only minor SNP variation over time and space although they were isolated from different animal species in Kenya, Ethiopia, Brazil, Indonesia, and China between 1980 and 2013. Within this cluster, the largest genomic difference was found between STIB810 and E110 with a total of only 2,534 SNP differences (homozygous and heterozygous). The African strains C13 and MU09 showed the lowest genomic difference (375 SNPs) and were more closely related to the Brazilian strain E110 than to the Asian strains STIB805, STIB810, and RoTat 1.2. The two *T. evansi* type B strains KETRI2479 and MU010 also form a monophyletic cluster, which has emerged separately from the ancestor of Western/Central African trypanosomes.

The *T. equiperdum* strains are genetically most related to the Eastern African *T. brucei* strains. The *T. equiperdum* strains Dodola 943, TeAp-N/D1, and OVI form a monophyletic cluster, closely related to the Kiboko *T. b. brucei* strains TREU927 and KETRI1738 and the *T. b. rhodesiense* strain EATRO 240. We observed only 27 SNP differences (homozygous and heterozygous) between the *T. equiperdum* genomes of Dodola

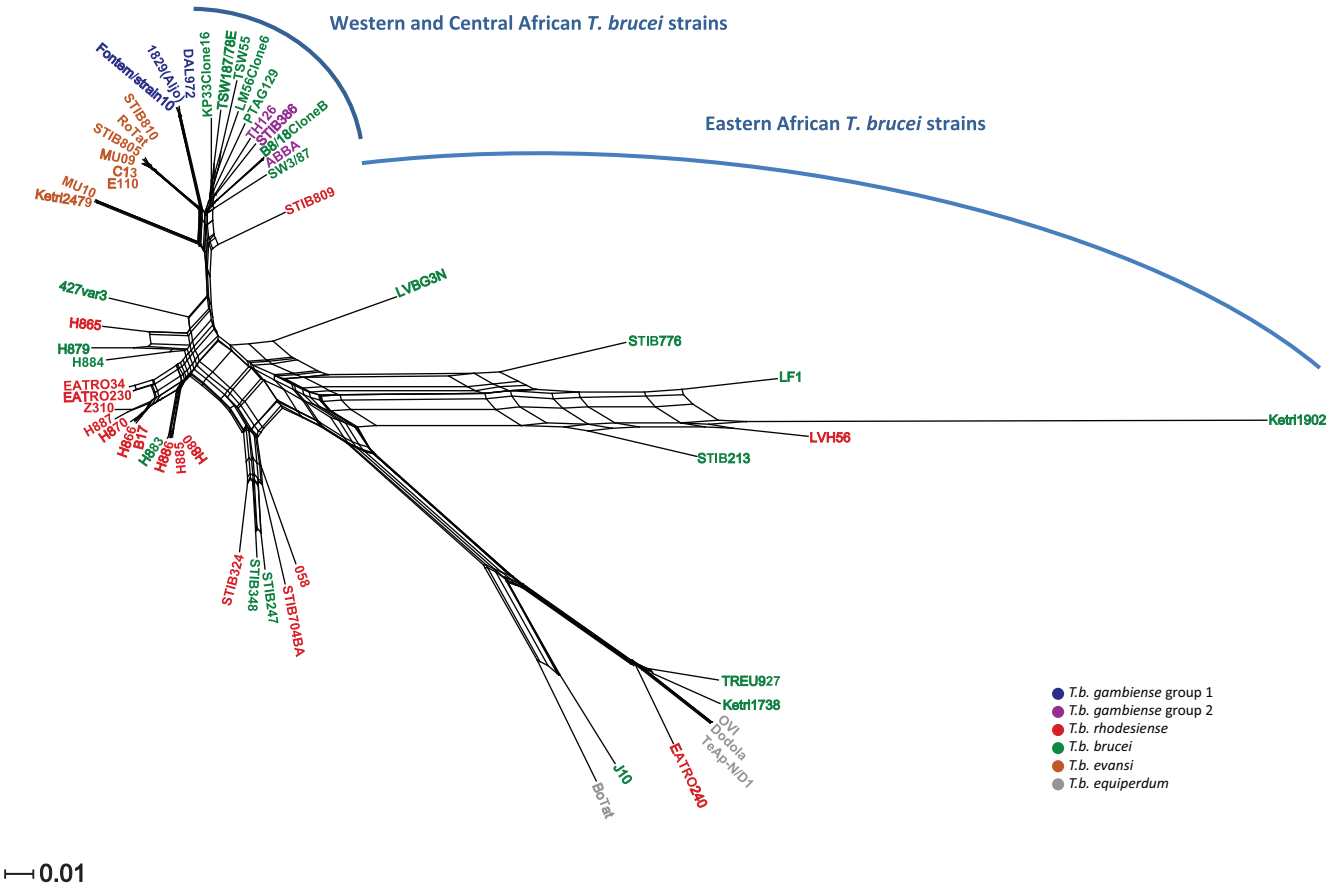


**Fig. 1.**—NeighborNet network based on 194,566 genome-wide SNP loci in 3 *Trypanosoma brucei gambiense* group 1, 3 *T. b. gambiense* group 2, 17 *Trypanosoma brucei rhodesiense*, 21 *Trypanosoma brucei brucei*, 8 *Trypanosoma evansi*, and 4 *Trypanosoma equiperdum* strains.

943 and TeAp-N/D1, 27 SNP differences between OVI and TeAp-N/D1, and 24 SNP differences between Dodola 943 and OVI. The *T. equiperdum* strain BoTat, isolated from a horse in Morocco is distinct from this monophyletic cluster and its genome is closely related to the *T. b. brucei* strain J10 isolated from a hyena in Zambia. Both BoTat and J10 show an uncertain ancestry and share haplotypes with *T. b. rhodesiense* EATRO 240, *T. b. brucei* TRUE972 and KETRI1738, and *T. equiperdum* Dodola 943, TeAp-N/D1 and OVI (fig. 2). To a lower extent, they also share haplotypes with Eastern and Western African *T. brucei* strains.

## Subgroup-Specific SNPs

SNPs unique for *T. evansi* type A, *T. evansi* type B, *T. equiperdum* BoTat and the *T. equiperdum* Dodola 943, TeAp-N/D1, and OVI monophyletic cluster were identified. We only included mutations that differed from the homozygous reference state (compared with the reference genome TREU927) by being homozygous for the alternative allele. The complete list of SNPs for each studied subgroup is presented in supplementary table S3, Supplementary Material online. We identified 354 SNPs that are unique to the monophyletic *T. equiperdum* cluster with Dodola 943, TeAp-N/D1 and OVI, and that did not occur in any other of the 53 strains of this study. Of the 354 SNPs, 224 were in coding regions of which 109 were nonsynonymous substitutions. In the *T. equiperdum* BoTat strain, 1,425 unique SNPs were observed, of which 850 in coding regions and 429 were nonsynonymous substitutions. Only five unique SNPs were shared by all *T. equiperdum* genomes, including the distinct BoTat genome. For *T. evansi* type B we detected 701 unique SNPs of which 454 in coding regions and 238 were nonsynonymous substitutions. An overview of the subgroup specific SNPs is presented in supplementary table S2, Supplementary Material online.

## Human Serum Resistance Genes, Diagnostic VSG's, and F1-ATP Synthase Subunit $\gamma$

The TgsGP gene was detected in all *T. brucei gambiense* group 1 strains and in none of the other trypanosomes. The *T. b. gambiense* group 1 specific S210 codon in the TbHpHbR gene was also unique for all *T. b. gambiense group 1* strains, whereas the other strains in this study coded for L210 in the TbHpHbR gene. All *T. b. rhodesiense* genomes contained *SRA*, except EATRO240. Surprisingly, the *T. b. rhodesiense*

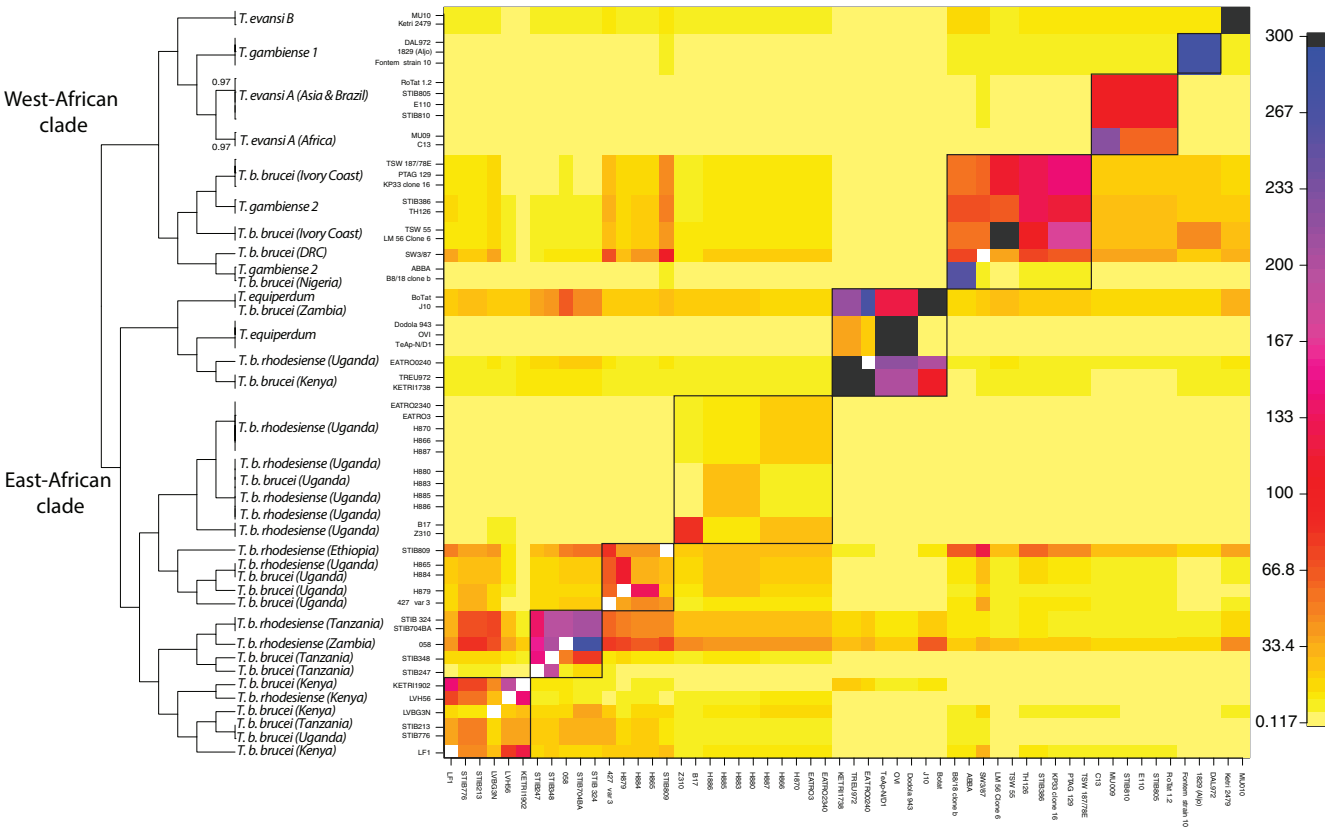


**Fig. 2.**—Coancestry matrix based on phased haplotype data. Heatmap summarizes the number of haplotype segments (color key on the right) that a given parasite received (rows) from any another parasite (columns). Individuals are ordered along each axis according to the tree (left) inferred from the fineSTRUCTURE run.

specific SRA gene was also detected in the *T. b. brucei* strains H883 and STIB213 isolated respectively from a dog in Uganda and a hyena in Tanzania. RoTat 1.2 was found in all *T. evansi* type A strains and not in any other strain. In contrast, our data show that the VSG JN 2118HU, considered to be unique to *T. evansi* type B, is also present in the *T. b. gambiense* type 2 strains ABBA, TH126 and STIB 386, and the *T. b. brucei* strains B8/18 Clone B, KP33 Clone 16 and TSW187/78E. The sequences for JN 2118HU were not identical in all genomes but there was not a single SNP identified that is unique for *T. evansi* type B. F1-ATP synthase subunit $\gamma$ DNA and amino acid sequences were aligned for all genomes included in this study (supplementary figs. S2 and S3, Supplementary Material online). The nonsynonymous heterozygous substitution C142C/T (R48R/G) and heterozygous deletion GCT841del (A281del) are unique for all *T. evansi* type A strains, and the heterozygous A844A/T (M282M/L) for all *T. evansi* type B strains. The nonsynonymous homozygous substitution G817C (A273P) is unique for the *T. equiperdum* BoTat strain.

## Discussion

We have undertaken a phylogenetic analysis of 56 genomes from the *Trypanozoon* subgenus. This study provides new insights into the origin of *T. evansi* and *T. equiperdum* and their relation with the different *T. brucei* subspecies. We show that *T. evansi* emerged at least on two independent occasions from a West African *T. brucei* ancestor, whereas *T. equiperdum* emerged from East-African *T. brucei*.

The genomes of the *T. evansi* type A and type B strains are grouped in two distinct but highly monophyletic clusters, although the strains within each type span different continents and >30 years. Our phylogenetic analysis indicates that both emerged from the same pool of *T. brucei* strains isolated in Western Africa, but on two different occasions. This observation is surprising because *T. evansi* type B has only been reported in Eastern Africa. However, given the lack of reliable diagnostic tools, *T. evansi* type B is probably present but not detected in Western and Northern Africa. The *T. evansi* type B strains KETRI 2479 and MU010 lack the RoTat 1.2 gene which is primarily used to identify *T. evansi* type A isolates. Our observations are in line with the hypothesis of Lun et al. who suggested in 1992 that *T. evansi* type A has emerged from a single *T. brucei* strain from Western Africa. Our data further indicate that the host and route of transmission shared by *T. evansi* type A and type B are due to parallel but independent evolutions, and that *T. evansi* is in fact a paraphyletic group.

The *T. equiperdum* strains TeAp-N/D1 (Venezuela), OVI (South Africa), and Dodola 943 (Ethiopia) also form a highly monophyletic cluster, although the strains have been collected in different continents and over a period of 30 years. This *T. equiperdum* cluster is most closely related to *T. b. brucei* Kiboko strains KETRI 1738, TREU927 and *T. b. brucei* EATRO240, confirming previous findings based on microsatellite markers (Carnes 2015). Because of their close genetic distance from the Kiboko strains, they share identical COX1, LipDH and F1-ATP $\gamma$-subunit genes, making it impossible to distinguish them based on these genes only (Carnes 2015). The *T. equiperdum* BoTat strain, isolated in 1924 in Morocco, emerged from an event distinct from the other *T. equiperdum* strains included in this study. Our haplotype coancestry analysis indicates that BoTat may have undergone recombination with the *T. b. brucei* strain J10.

We hypothesize that the small genetic variation within *T. evansi* and *T. equiperdum* is due to the restricted host range (camelids and Equidae, respectively) and the fact that *T. evansi* and *T. equiperdum* do not have cyclical transmission through the tsetse fly, where genetic recombination between trypanosomes occurs (Gibson 2015). The number of heterozygous SNPs in *T. evansi* and *T. equiperdum* were amongst the lowest in our data set and the splitstree network shows no reticulate events for the *T. evansi* and *T. equiperdum* strains, except BoTat, and they cluster tightly at the tips, indicating absence of recombination. The *T. evansi* and *T. equiperdum* clusters are situated at the two extreme ends of the *Trypanozoon* phylogenetic network. This is particularly interesting since this means that dyskinetoplasty, and thus tsetse fly independent transmission, has evolved independently at least four times. This efficient transmission strategy has enabled *T. evansi* and *T. equiperdum* to spread across the globe, as previously postulated by Lun et al. (1995) and Jensen et al. (2008). On the basis of our observations, the classification of *T. evansi* and *T. equiperdum* as species does not reflect their evolutionary background. Indeed, the genetic distance between Western and Eastern African *T. b. brucei* strains is larger than the differences between many *T. b. brucei* and *T. evansi* or *T. equiperdum.* On the basis of this genomic information, *T. evansi* and *T. equiperdum* could be regarded as subspecies of *T. brucei* (*T. b. evansi* and *T. b. equiperdum*), as previously suggested by Claes et al. (2005), Lai et al. (2008). However, *T. b. evansi* and *T. b. equiperdum* would both remain polyphyletic species. Hence, *T. b. evansi* type A and *T. b. evansi* type B, *T. b. equiperdum* type OVI and *T. b. equiperdum* type BoTat would be a more correct classification.

For *T. equiperdum* there are currently no genetic markers for identification. *Trypanosoma evansi* type B is identified using PCR with primers targeting either the nuclear encoded VSG JN 2118HU gene or the EVAB gene in the minicircles. However, Birhanu et al. (2016) showed that this VSG JN 2118HU also occurs in *T. b. gambiense* type 2 ABBA and *T. brucei* AnTat 1.1. Our genome data show additional presence of *VSG JN 2278HU* in several other West-African *T. b. brucei* indicating that this gene is not a specific marker for *T. evansi* type B. Also, *T. evansi* type B diagnosis with the EVAB PCR, which targets a *T. evansi* B specific minicircle sequence, is questionable. Since *T. evansi* lacks maxicircles, there might be no known remaining function and therefore no selective pressure for minicircles left. Indeed it has been shown that akinetoplasty occurs in *T. evansi* frequently in vivo and in vitro (Schnaufer et al. 2002). Whether akinetoplasty also occurs naturally in *T. evansi* type B is presently not known, and consequently using nuclear markers for genotyping would be a more secure option. In our work, we have identified 702 homozygous SNPs, of which 239 nonsynonymous nuclear coded ones are unique for *T. evansi* type B and could therefore be promising candidate markers for molecular identification. Our genome-wide SNP analysis also identified 354 homozygous SNPs that are all unique for the monophyletic *T. equiperdum* OVI, Dodola 943, and TeAp-N/D1 strains and may be promising candidate genetic markers for typing and diagnosis.

Dean et al. have reported specific single amino acid mutations in the nuclearly encoded F1-ATPase subunit $\gamma$ of *T. equiperdum* and *T. evansi* that compensate for loss of their kinetoplast DNA (2013). They showed that the homozygous A273P substitution in *T. equiperdum* BoTat is sufficient for normal growth in the absence of kDNA. The heterozygous mutations A281del and M282L, were reported to be secondary adaptations and not sufficient for survival (Dean et al. 2013). These amino acid mutations were also observed in our genome data: A273P was only present in *T. equiperdum* strain BoTat, A281del in all *T. evansi* type A strains and M282M/L in all *T. evansi* type B strains.

The 21 *T. b. brucei* strains are very heterogeneous and scattered around the network. In addition, the 17 *T. b. rhodesiense* strains are present within all major clades of East African *T. b. brucei*. *Trypanosoma brucei brucei* strains from Western Africa showed to be most related to *T. b. gambiense* group 1, group 2 and *T. evansi*. Further downstream the phylogenetic network, the *T. b. brucei* and *T. b. rhodesiense* strains are spread geographically from central Africa (Uganda) to Eastern Africa (Kenya, Tanzania and Zambia). This is in line with the observations of Sistrom et al. (2014) who showed that geography significantly accounts for cluster assignment in *T. brucei*. Most of the *T. brucei* genomes sequenced by Sistrom et al. (2014) are also part of our analysis. In addition, our network shows that the *T. b. brucei* genomes are less heterogeneous in Western and Central Africa compared with East Africa, suggesting an Eastern African origin. This polyphyletic character of *T. b. brucei* and *T. b. rhodesiense* suggests multiple origins of *T. b. rhodesiense* and supports the hypothesis suggested by Gibson et al. (2002) that the SRA gene can be exchanged between *T. b. brucei* and *T. b. rhodesiense* through recombination. A reticulated network (fig. 1) and the fragmented haplotypic ancestry of most parasites (fig. 2) indicate that recombination events are frequent.

This complex phylogenetic background makes the status of *T. b. rhodesiense* confusing, since its current classification based on the presence of the SRA gene does not mimic its population genomic structure (Sistrom et al. 2016). It was therefore no surprise that we observed several misclassifications in our strain set. *SRA* was present in the genomes of the *T. b. brucei* strains H883 and STIB213, showing that these strains have probably been misclassified as *T. b. brucei* and are in fact *T. b. rhodesiense*. For H883, this correlates with its place in the phylogenetic network, since it was the only outlier in a group of 10 Ugandan *T. b. rhodesienses* strains (H866, H870, B17, Z310, H887, EATRO2340, EATRO3, H885, H880). On the other hand, we found no evidence of *SRA* in EATRO240. This strain was isolated from a patient in Uganda in 1961 and, apart from the possibility of strain contamination, might thus have been an atypical human infection with *T. b. brucei*. Human infections with animal trypanosomes have been linked to dysfunctional APOL1 in the patient's serum (Vanhollebeke et al. 2006; Cuypers et al. 2016).

In summary, we have presented the first whole genome SNP analysis of *T. evansi* and *T. equiperdum* and provided new insights in the origin of both species and their relation with the different *T. brucei* subspecies. That *T. evansi* type A, *T. evansi* type B and *T. equiperdum* have emerged independently from each other was already observed by Carnes et al. (2015), using four molecular markers (microsatellites, COX1, LipDH, and F1-ATP $\gamma$-subunit) and with only one isolate for all groups except *T. evansi* type A. Our study was based on genome-wide SNP analysis of 56 *Trypanozoon* genomes, including eight *T. evansi* and four *T. equiperdum*, providing new insights in the geographical origins of *T. evansi* and *T. equiperdum*. The *T. evansi* types A and B genomes are related to the *T. brucei* genomes from Western Africa whereas the *T. equiperdum* genomes relate to the *T. brucei* genomes from Eastern Africa.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

Funding was received from the Research Foundation Flanders (FWO, grants 1501413N and 1101614N) and the European DG Health and Food Safety (SANTE). We thank the Center of Medical Genetics at the University of Antwerp for hosting the NGS facility. The computational resources used for this work were provided by the VSC (Flemish Supercomputer Center) at the University of Antwerp. We thank Dr Mark Sistrom, Dr Joshua Richardson, and Dr Adalgisa Caccone (Yale University) for providing genome reads from 19 *T. b. brucei*, 13 *T. b. rhodesiense*, 3 *T.b gambiense* and 4 *T. evansi* strains. We thank Dr Dave Barry for critical reading of the manuscript. The *T. equiperdum* strain TeAp-N/D1 was kindly donated by Dr Maria Isabel Gonzatti (Universidad Simón Bolívar, Venezuela).

## Literature Cited

Bajyana Songa E, Hamers R. 1988. A card agglutination test (CATT) for veterinary use based on an early VAT RoTat 1/2 of *Trypanosoma evansi*. Ann Soc Belg Med Trop. 68:233–240.

Balmer O, Beadell JS, Gibson W, Caccone A. 2011. Phylogeography and taxonomy of *Trypanosoma brucei*. PLoS Negl Trop Dis. 5:e961.

Berriman M, et al. 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416–422.

Bertrand KI, Hajduk SL. 2000. Import of a constitutively expressed protein into mitochondria from procyclic and bloodstream forms of *Trypanosoma brucei*. Mol Biochem Parasitol. 106:249–260.

Birhanu HA, Gebrehiwot T, Goddeeris BM, Büscher P, Van Reet N. 2016. New *Trypanosoma evansi* type B isolates from Ethiopian dromedary camels. PLoS Negl Trop Dis. 10:e0004556.

Borst P, Fase-Fowler F, Gibson WC. 1987. Kinetoplast DNA of *Trypanosoma evansi*. Mol Biochem Parasitol. 23:31–38.

Bringaud F, Riviere L, Coustou V. 2006. Energy metabolism of trypanosomatids: adaptation to available carbon sources. Mol Biochem Parasitol. 149:1–9.

Brun R, Hecker H, Lun ZR. 1998. *Trypanosoma evansi* and *T. equiperdum*: distribution, biology, treatment and phylogenetic relationship. Vet Parasitol. 79:95–107.

Capbern A, Giroud C, Baltz T, Mattern P. 1977. Trypanosoma equiperdum: étude des variations antigéniques au cours de la trypanosomose experimentale du lapin. Exp Parasitol. 42:6–13.

Carnes J, et al. 2015. Genome and phylogenetic analyses of *Trypanosoma evansi* reveal extensive similarity to *T. brucei* and multiple independent origins for dyskinetoplasty. PLoS Negl Trop Dis. 9:e3404.

Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N. 2009. Importing mitochondrial proteins: machineries and mechanisms. Cell 138:628–644.

Claes F, Büscher P, Touratier L, Goddeeris BM. 2005. *Trypanosoma equiperdum*: master of disguise or historical mistake? Trends Parasitol. 21 316–321.

Claes F, Radwanska M, Urakawa T, Majiwa PA, Goddeeris B, Büscher P. 2004. Variable Surface Glycoprotein RoTat 1.2 PCR as a specific diagnostic tool for the detection of Trypanosoma evansi infections. Kinetoplastid Biology and Disease 3:3.

Cuypers B, et al. 2016. Apolipoprotein L1 variant associated with increased susceptibility to trypanosome infection. MBio 7:e02198–e02215.

Dean S, Gould MK, Dewar CE, Schnaufer AC. 2013. Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes. Proc Natl Acad Sci U S A. 110:14741–14746.

Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. 2013. Haplotype estimation using sequencing reads. Am J Hum Genet. 93:687–696.

Desquesnes M, et al. 2013a. *Trypanosoma evansi* and surra: a review and perspectives on transmission, epidemiology and control, impact, and zoonotic aspects. Biomed Res Int. 2013: 321237.

Desquesnes M, et al. 2013b. *Trypanosoma evansi* and surra: a review and perspectives on origin, history, distribution, taxonomy, morphology, hosts, and pathogenic effects. Biomed Res Int. 2013: 2013:194176.

Gibson W. 2015. Liaisons dangereuses: sexual recombination among pathogenic trypanosomes. Res Microbiol. 166:459–466.

Gibson W, Backhouse T, Griffiths A. 2002. The human serum resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infect Genet Evol. 1 207–214.

Gibson WC. 1986. Will the real *Trypanosoma b. gambiense* please stand up? Parasitol Today. 2 255–257.

Goodhead I, et al. 2013. Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence. MBio 4:1–8.
Hagos A, Goddeeris BM, Yilkal K, Alemu T, Fikru R, Yacob HT, Feseha G, Claes F. 2010. Efficacy of cymelarsan and diminasan against Trypanosoma equiperduminfections in mice and horses. Vet Parasitol. 171:200–206.
Hebert L, et al. 2017. First draft genome sequence of the dourine causative agent: *Trypanosoma equiperdum* strain OVI. J Genomics. 5:1–3.
Herbert WJ, Lumsden WHR. 1976. *Trypanosoma brucei:* a rapid "matching" method for estimating the host's parasitemia. Exp Parasitol. 40:427–431.
Hide G, Cattand P, Le Ray D, Barry JD, Tait A. 1990. The identification of Trypanosoma brucei subspecies using repetitive DNA sequences. Molecular and Biochemical Parasitology. 39:213–226.
Hoare CA. 1972. The trypanosomes of mammals. Oxford: Blackwell Scientific Publications.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.
Jackson AP, et al. 2010. The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human African trypanosomiasis. PLoS Negl Trop Dis. 4:e658.
Jensen RE, Simpson L. Englund PT. 2008. What happens when *Trypanosoma brucei* leaves Africa. Trends Parasitol. 24:428–431.
Lai DH, Hashimi H, Lun Z-R, Ayala FJ, Lukes J. 2008. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are *petite* mutants of *T. brucei*. Proc Natl Acad Sci U S A. 105:1999–2004.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
Lanham SM, Godfrey DG. 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. Exp Parasitol. 28:521–534.
Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8:e1002453.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.
Lun Z-R, Brun R, Gibson W. 1992. Kinetoplast DNA and molecular karyotypes of *Trypanosoma evansi* and *Trypanosoma equiperdum* from China. Mol Biochem Parasitol. 50:189–196.
Lun ZR, Desser SS. 1995. Is the broad range of hosts and geographical distribution of Trypanosoma evansi attributable to the loss of maxicircle kinetoplast DNA? Parasitol Today. 11:131–133.
Mehlitz D, Zillmann U, Scott CM, Godfrey DG. 1982. Epidemiological studies on the animal reservoir of Gambiense sleeping sickness. Part III. Characterization of Trypanozoon stocks by isoenzymes and sensitivity to human serum. Tropenmedizin und Parasitologie 33:113-118.
Neupert W. 1997. Protein import into mitochondria. Annu Rev Biochem. 66:863–917.
Perrone TM, Gonzatti MI, Villamizar G, Escalante A, Aso PM. 2009. Molecular profiles of Venezuelan isolates of Trypanosoma sp. by random amplified polymorphic DNA method. Veterinary Parasitol. 161:194-200.
Sánchez E, Perrone T, Recchimuzzi G., Cardozo Biteau I, Aso N, Mijares PM, Baltz A, Berthier T, Balzano-Nogueira D, L, et al. 2015a. Erratum to: Molecular characterization and classification of Trypanosoma spp. Venezuelan isolates based on microsatellite markers and kinetoplast maxicircle genes. Parasit Vectors 8:566.
Sánchez E, Perrone T, Recchimuzzi G., Cardozo Biteau I, Aso N, Mijares PM, Baltz A, Berthier T, Balzano-Nogueira D, L, et al. 2015b. Molecular characterization and classification of Trypanosoma spp. Venezuelan isolates based on microsatellite markers and kinetoplast maxicircle genes. Parasites and Vectors 8:536.
Schnaufer A, Clark-Walker GD, Steinberg AG, Stuart K. 2005. The F1-ATP synthase complex in bloodstream stage trypanosomes has an unusual and essential function. EMBO J. 24:4029–4040.
Schnaufer A, Domingo GJ, Stuart K. 2002. Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. Int J Parasitol. 32:1071–1084.
Sievers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539.
Sistrom M, et al. 2016. De novo genome assembly shows genome wide similarity between *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*. PLoS ONE. 11:e0147660.
Sistrom M, et al. 2014. Comparative genomics reveals multiple genetic backgrounds of human pathogenicity in the *Trypanosoma brucei* complex. Genome Biol Evol. 6:2811–2819.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Truc P, Formenty P, Diallo PB, Komoin-Oka C, Lauginie F. 1997. Confirmation of two distinct classes of zymodemes of *Trypanosoma brucei* infecting man and wild mammals in Côte d'Ivoire: suspected difference in pathogenicity. Ann Trop Med Parasitol. 91:951–956.
Uzureau P, et al. 2013. Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. Nature 501:430–434.
Vanhollebeke B, et al. 2006. Human *Trypanosoma evansi* infection linked to a lack of apolipoprotein L-1. N Engl J Med. 355:2752–2756.
Vanhollebeke B, et al. 2008. A haptoglobin–hemoglobin receptor conveys innate immunity to *Trypanosoma brucei* in humans. Science 320:677–681.
Weir W, et al. 2015. Population genomics reveals the origin and asexual evolution of human infective trypanosomes. Elife 5:e11473.
Xong HV, et al. 1998. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. Cell 95:839–846.

**Associate editor:** Geoff McFadden