J Mol Evol (2003) 56:377–386
DOI: 10.1007/s00239-002-2408-z

JOURNAL OF MOLECULAR EVOLUTION

The Transferrin Receptor Genes of *Trypanosoma equiperdum* Are Less Diverse in Their Transferrin Binding Site than Those of the Broad-Host Range *Trypanosoma brucei*

Tomoko Isobe,[1] Edward C. Holmes,[2] Gloria Rudenko[1]

[1] The Peter Medawar Building for Pathogen Research, University of Oxford, South Parks Road, Oxford OX1 3SY, UK
[2] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received: 23 July 2002 / Accepted: 24 September 2002

Abstract. *Trypanosoma brucei* and *T. equiperdum* infect the mammalian bloodstream and tissues. *T. brucei* is transmitted by tsetse flies between an extremely large range of mammals in sub-Saharan Africa. In contrast, *T. equiperdum* is restricted to equines, where it is transmitted as a venereal disease. Both species evade immune destruction by changing their variant surface glycoprotein (VSG), encoded in a telomeric *VSG* expression site. *T. brucei* has about 20 *VSG* expression sites, and it has been proposed that their genetic diversity plays a role in host adaptation. Two expression site-associated genes *ESAG6* and *ESAG7*, encode variable transferrin receptor subunits allowing trypanosomes to internalize polymorphic transferrin molecules from different mammals. We investigated if there was a correlation between the size of the trypanosome host range and the degree of *ESAG6* genetic diversity. Both *T. equiperdum* and *T. brucei* appear to have approximately similar numbers of *ESAG6*, however, the genetic diversity of the *ESAG6* family varies in the two species. We sequenced 114 *T. equiperdum ESAG6* genomic clones, resulting in the isolation of 10 *T. equiperdum ESAG6* variants. The *T. equiperdum ESAG6* genes were less genetically diverse than those of *T. brucei* in regions known to play a role in transferrin binding. This indicates that *ESAG6* genetic diversity playing a role in host adaptation could have been lost in the absence of selection pressure. There was also evidence of positive selection ($d_N/d_S = \sim 5$) acting on other *ESAG6* regions not involved in transferrin binding, perhaps due to antigenic variation of these surface molecules.

Key words: Host adaptation — *Trypanosoma equiperdum* — *Trypanosoma brucei* — Transferrin receptor — Expression site — Expression site-associated gene — Antigenic variation — Positive selection

Introduction

Trypanosoma brucei is a protozoan parasite which causes African sleeping sickness in humans and nagana in cattle, both endemic to sub-Saharan Africa. *T. brucei* infects an extraordinarily large range of mammals, varying from waterbuck and hartebeest to lions and hippopotamus (see Table 1), Transmission occurs primarily via tsetse flies. In contrast, *T. equiperdum* naturally infects only equines (horses, donkeys, and mules), causing a venereal disease called dourine (Alemu et al. 1997). *T. equiperdum* is closely related to *T. brucei* but has lost the ability to be transmitted by tsetse flies.

T. equiperdum differs from *T. brucei*, in that it is present primarily in host tissue rather than the

Nucleotide sequence data reported in this paper are available in the EMBL GenBank, and DDJB databases under accession numbers AY152680–AY152692.

Correspondence to: Gloria Rudenko; *email:* gloria.rudenko@medawar.ox.ac.uk

Table 1. Comparison of the host range of *Trypanosoma brucei* and *Trypanosoma equiperdum*[a]

Species	Mammalian host	Reference
Trypanosoma brucei	Human	Fèvre et al. (2001) *Lancet* **358:**625–628
	Cattle	Robson et al. (1972) *Bull WHO* **46:**765–770
	Bushbuck	Ibid
	Reedbuck	Ibid
	Waterbuck	Ibid; Geigy et al. (1971) *Acta Trop* **28:**211–220
	Goat	Ibid
	Sheep	Ibid
	Dog	Ibid
	Spotted hyena	Ibid
	Lion	Ibid
	Warthog	Ibid
	Coke's hartebeest	Ibid
	Eland	Ashcroft (1959) *East Afr Med* **36:**289–291
	Buffalo	Ibid
	Equines (horse, donkey)	Faye et al. (2001) *Vet Parasio* **101:**101–114
	Pig	Killick-Kendrick and Godfrey (1963) *Ann Trop Med Parasitol* **57:**215-231
	Giraffe	Awan (1979) *Acta Trop* **36:**343–347
	Impala	Mulla and Ricknian (1988) *Trans Roy Soc Trop Med Hyg* **82:**718
	Zebra	Ibid
	Hippopotamus	Dillman and Awan (1972) *Trop Anim Health Prod* **4:**135–137
Trypanosoma equiperdum	Equines (horse, donkey, mule)	Alemu et al. (1997)

[a] This table lists some of the documented mammalian species that *T. brucei* and *T. equiperdum* can naturally infect. Some of these mammals are only sporadically infected by *T. brucei* sp. Experimentally infected animals (including laboratory rodents) have been omitted.

bloodstream (Lun et al. 1992b). However, both parasites evade the mammalian immune system using a highly sophisticated strategy of antigenic variation of a variant surface glycoprotein (VSG) coat. *T. brucei* has approximately 1000 *VSG* genes (Van der Ploeg et al. 1982) and *T. equiperdum* has minimally 100 *VSG* genes as observed in chronic infections in rabbits (Capbern et al. 1977). However, only one *VSG* gene is expressed at a time in a telomeric *VSG* expression site (ES). Switching *VSG* genes can involve DNA rearrangements moving a silent *VSG* into the active ES. Alternatively, the trypanosome can switch between multiple *VSG* ESs. Antigenic variation is reviewed by Pays and Nolan (1998), Borst et al. (1998), Cross et al. (1998), Vanhamme et al. (2001), and Barry and McCulloch (2001).

In *T. brucei*, there are approximately 20 *VSG* ESs (Navarro and Cross 1996), which can vary in size from about 30 to 60 kb. Expression sites are polycistronic transcription units including variable numbers of expression site-associated genes (*ESAG*s) in addition to the telomeric *VSG* gene (Pays et al. 1989; Lips et al. 1993; Xong et al. 1998; Kooter et al. 1987; Berriman et al. 2002). The function of most *ESAG*s is unclear. *ESAG4* encodes an adenylate cyclase (Ross et al. 1991) and *ESAG6* and *ESAG7* encode the subunits of a heterodimeric transferrin receptor (Schell et al. 1991; Salmon et al. 1994; Ligtenberg et al. 1994; Steverding et al. 1995). Recently, a serum resistance-associated (*SRA*) gene has been identified as a new *ESAG* conferring resistance to lysis in human serum (Xong et al. 1998).

It was initially puzzling why *T. brucei* would need 20 *VSG* ESs as this redundancy necessitates a counting mechanism ensuring mutually exclusive *VSG* expression (Chaves et al. 1999). However, an appealing explanation is that the multiple *VSG* ESs, each containing similar variants of different *ESAG* genes, could play a role in host adaptation (Bitter et al. 1998). While infecting different mammalian species, *T. brucei* takes up polymorphic host molecules such as transferrin. The two subunits of the heterodimeric trypanosome transferrin receptor are encoded by *ESAG6* and *ESAG7* (see Fig. 1A). Having approximately 20 pairs of the *ESAG6* and *ESAG7* genes could allow the trypanosome to switch between different transferrin receptors by switching between *VSG* ESs. Each pair of transferrin receptors appears to have a different affinity for the transferrin molecules from different mammalian hosts (Bitter et al. 1998; Gerrits et al. 2002).

We attempted to determine if there was a correlation between the size of the trypanosome host range and the extent of genetic diversity in transferrin receptor genes. As *T. equiperdum* infects exclusively equines, it should need only a single optimized pair of transferrin receptor genes (discussed by Borst et al. [1997]). *T. equiperdum* has been shown to be very closely related to *T. brucei* (Haag et al. 1998; Stevens et al. 1999). It has been postulated that *T. equiperdum* evolved from a *T. brucei* that lost the ability to be transmitted through insects (Lun et al. 1992a, b; Brun et al. 1998). However, this appears to have happened relatively recently, as genetic differences between

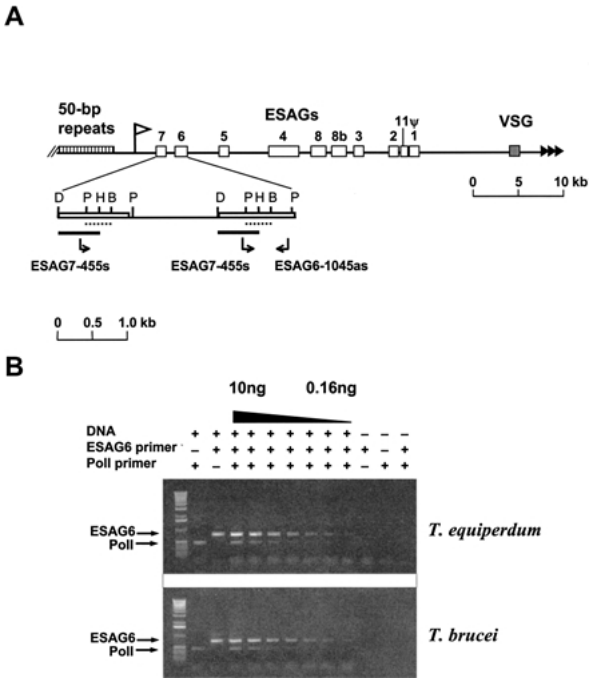

Fig. 1. A Schematic representation of the location of the *ESAG6* and *ESAG7* genes in *T. brucei VSG* expression sites. The *top* diagram is a schematic representation of the AnTat 1.3A *VSG* expression site, according to Revelard et al. (1990), Lips et al. (1993), and Redpath et al. (2000). The *flag* indicates the promoter, and the numbered *open boxes* indicate the *ESAG*s. The region around *ESAG6* and *ESAG7* is expanded *below*. The Southern blot probe (*Dra*I/*Hind*III fragment) is indicated by a *black bar*. The *ESAG6 Pst*I/*Bam*HI fragment used for cloning is shown by a *dotted line*. Abbreviations for restriction endonuclease sites are as follows: D, *Dra*I; P, *Pst*I; H, *Hind*III; B, *Bam*HI. *Arrows* underneath the genes indicate PCR primers: ESAG7-455s and ESAG6-1045as. **B** Equivalent copy number of *ESAG6* in *T. brucei* and *T. equiperdum*. *ESAG6* was PCR-amplified from *T. equiperdum* and *T. brucei* genomic DNA. PCR amplification was performed using two sets of primers: one set specific for *ESAG6* (see A) and one set specific for the single-copy gene encoding the large subunit of RNA polymerase I (*Pol*I) as an internal control. A range of DNA template concentrations was used (10, 5, 2.5, 1.25, 0.63, 0.31, and 0.16 ng) as indicated at the *top*. The presence (+) or absence (−) of template DNA or primers in each reaction is indicated.

these species are minor. These differences have primarily been characterized using repetitive sequence and microsatellite probes and enzyme characterization (Lun et al. 1992a; Brun et al. 1998; North et al. 1983; Biteau et al. 2000). In addition, compared with *T. brucei*, *T. equiperdum* has reduced diversity in its kinetoplast DNA (kDNA) minicircles and a deletion of the maxicircle kDNA (Riou and Saucier 1979; Frasch et al. 1980; Shu and Stuart 1994; Lun et al. 1992b).

T. equiperdum has been shown to have multiple *VSG* ESs between which it can switch (Longacre et al. 1983; Raibaud et al. 1986), but the exact number and structure are unknown. *ESAG9* has a copy number of six to eight in both *T. brucei* and *T. equiperdum*, although not all of these genes are telomere linked (Florent et al. 1991). *ESAG4* and *ESAG8* were also found associated with *ESAG9*, although copy numbers have not been determined (Florent et al. 1991; Smiley et al. 1990). Although *T. equiperdum* is closely related to *T. brucei*, one would expect that in the absence of selection pressure for maintenance of many telomeric *VSG* ESs, the copy number, degree of genetic diversity, and strength of selection pressure on these telomeric sequences would be reduced. We isolated and sequenced different members of the *T. equiperdum ESAG6* gene family and compared them with *T. brucei ESAG6* sequences to determine if the degree of genetic diversity and evolutionary processes acting in this gene family could be correlated with the size of the host range.

Materials and Methods

Trypanosome DNA Analysis and Genomic Clone Isolation

The trypanosome strains used were *Trypanosoma brucei* strain 427 variant VSG221a (Bernards et al. 1984) and *T. equiperdum* STIB818 isolated from a horse in China in 1979 (kind gift of W. Gibson, University of Bristol) (Lun et al. 1992a, b)

For *ESAG6* quantitation, *ESAG6* was amplified by polymerase chain reaction (PCR) using the primers ESAG7-455s (5′-CATTCCAGCAGGAGTTGGAGG-3′) and ESAG6-1045as (5′-TTGTTCACTCACTCTCTTTGACAG-3′) with *Taq* DNA polymerase (Invitrogen) for 20 cycles. PCR was also performed on the single-copy gene encoding the large subunit of RNA polymerase I (*PolI*) as an internal control for the amount of DNA using PolI-484 (5′CAGGAGGATCGTTCGGCACCTTGGC-3′) and PolI-913 (5′-CATGCGCCTGTGGTTCAGCATAGC-3′). PCR products were cloned into pBluescript SK(−) (Stratagene) and sequenced on both strands to check identity. *T. brucei* and *T. equiperdum* have identical sequences in the area of the *PolI* gene analyzed (results not shown). The PCR yield was quantified using the Gel Doc 2000 Gel Documentation System with Quantity One software (BioRad).

For Southern blot analysis, 1 μg of genomic DNA was digested with restriction endonucleases, separated on an 0.8% agarose gel, and transferred to a nylon membrane according to Sambrook et al. (1989). The blot was hybridized with the DES 12 probe (*Dra*I/*Hind*III fragment indicated in Fig. 1A of Rudenko et al. [1994]), which hybridizes with both *ESAG7* and *ESAG6*. The blot was washed to an end stringency of 0.3× SSC at 65°C for 30 min. As a positive control, and for quantitation of the genomic DNA, the blot was stripped and reprobed with a *PolI* gene probe. The positive bands were visualized and quantified using a Personal Molecular Imager FX System with Quantity One software (BioRad).

To determine the genetic diversity of *ESAG6*, *ESAG6* genes were cloned from *T. equiperdum* and *T. brucei* using size-selected genomic libraries. Genomic DNA was digested with *Pst*I and *Bam*HI, and fragments of approximately 350–450 bp were isolated from agarose gels and cloned into pBluescript SK(−) digested with *Pst*I/*Bam*HI. Bacterial clones were screened for the presence of either *ESAG6* or *ESAG7* by colony hybridization with a 392-bp *Pst*I/*Bam*HI fragment from *ESAG6* of the 221 expression site. Clones containing a single insert were isolated and sequenced on both strands using T7 and T3 primers. Sequences were assembled using a Staden package or Contig Express (Informax Inc.) and aligned at BCM Search Launcher: Multiple Sequence Alignments Server with ClustalW 1.8.

The sequences TI-1e, TI-3e, TI-5e, TI-6e, and TI-8e to TI-13e were isolated from *T. equiperdum*, and the sequences TI-1b to TI-6b from *T. brucei*, in this study (accession numbers: AY152680–AY152692). TI-2b is the same as AJ007022 and AJ007027; TI-4b is the same as AF068703 and *ESAG6* from the 221 ES = AL671259; TI-6b is the same as AF068702, M33720 (Pays et al. 1989), and BS1.2 from Hobbs and Boothroyd (1990) = M34520; and *ESAG6* from the VO2 ES = AL671256. The Bn-2 ES *ESAG6* = AL670322; TBR7029 = AJ007029, and IsTat1.7 T5 = L07805 (Lodes et al. 1993). TBR7028 = AJ007028; AF068704 is the same as AJ010094 = Etat1.2CR (Xong et al. 1998). TBR7024 = AJ007024, TBR7026 = AJ007026, TBR7021 = AJ007021, TBR7023 = AJ007023, and the 10.1 ES = AC087700 (LaCount et al. 2001). TBR7020 = AJ007020, and TRBESAG6G = M62631 (Dai Do Thi et al. 1991); TBR7025 is AJ007025. AJ007020–AJ007029 are all described by Ansorge et al. (1999). AF068702–AF068705 are described by Alarcon et al. (1999). AL671259, AL671256, and AL670322 are all described by Berriman et al. (2002).

Phylogenetic analyses

Maximum likelihood (ML) phylogenetic trees were estimated using the PAUP* package (Swofford 2002). The GTR + I Γ model of nucleotide substitution was employed, with the values for the relative substitution rates among nucleotides (GTR), proportion of invariable sites (I), Γ distribution of among-site rate variation (Γ), and base composition estimated from the data (parameter values available from the authors on request). To assess phylogenetic robustness, bootstrap resampling was undertaken using 1000 replicate neighbor-joining trees reconstructed under the ML substitution model.

The CODEML program from the PAML package was employed to analyze selection pressures (Yang 1997). This implements a ML method that compares various models of codon evolution which differ in how they treat rates of synonymous (d_S) and nonsynonymous (d_N) substitution (d_N/d_S ratio) among codons or along lineages using likelihood ratio tests (Yang et al. 2000). Some models (denoted M2, M3, and M8) allow for positive selection in that they can incorporate classes of codon where $d_N > d_S$, while other models (denoted M0, M1, and M7) specify neutral evolutions as d_N is constrained to be less than d_S at all codons. If a model that allows positive selection is significantly favored over one that does not, and there is a class of codons $d_N/d_S > 1$, then we may infer that positive selection has occurred. If positive selection is found in any comparison, Bayesian methods can be used to identify the individual codons that have been subjected to this process. For simplicity, we compared two models: M7, which assumes that codons can be placed into 10 d_N/d_S categories estimated from the data following a β distribution, but where all d_N/d_S values are constrained to be < 1, thereby describing a strictly neutral evolutionary process; and M8, which differs from M7 in that an eleventh category of sites is added at which d_N/d_S may exceed 1, thereby allowing positive selection.

Results

The heterodimeric transferrin receptors of African trypanosomes have two subunits encoded by the *ESAG6* and *ESAG7* genes located in *VSG* ESs (see Fig. 1A). *ESAG6* and *ESAG7* are highly similar, but *ESAG7* appears to be truncated compared with *ESAG6*, which contains 3′ sequences encoding a signal for glycophosphatidyl inositol (GPI) anchorage to the cell membrane (Schell et al. 1991; Salmon et al. 1994; Ligtenberg et al. 1994). There appear to be approximately 20 pairs of transferrin receptor subunit encoding genes in *T. brucei* 427 (Ansorge et al. 1999), which fits with an estimated 20 *VSG* ESs (Navarro and Cross 1996).

We compared the relative number of *ESAG6* genes in *T. brucei* 427 variant 221a (Bernards et al. 1984) with a *T. equiperdum* STIB818 strain isolated in China in 1979 (Lun et al. 1992a; b) (Fig. 1B). We amplified *ESAG6* in both species by PCR using a primer pair which amplifies specifically *ESAG6*. We amplified a 740-bp fragment in both *T. brucei* and *T. equiperdum*. The PCR products were cloned, and their identity was confirmed by sequencing both ends of 10 clones from both species. By comparing PCR yields, and accounting for the fact that *T. brucei* 427 has about 20 *ESAG6* copies (Navarro and Cross 1996; Ansorge et al. 1999), we established that *T. equiperdum* STIB818 has approximately 16 ± 2 copies of *ESAG6* (mean and standard deviation from three independent experiments). The approximately equivalent number of *ESAG6* genes in both trypanosome species was confirmed in Southern blots of restriction enzyme digests of equal amounts of genomic DNA from both species (see Fig. 2).

We next determined the genetic diversity of these *T. equiperdum ESAG6* genes. PCR amplification of genes in highly similar multicopy gene families frequently leads to the formation of chimeric products (Henri van Luenen and Piet Borst, personal communication; our observations; Tanabe et al. 2002). To avoid this complication, we isolated genomic clones of *ESAG6* from size-enriched genomic libraries from both *T. equiperdum* and *T. brucei*. Size fractions of genomic DNA including a 392-bp *Pst*I/*Bam*HI fragment in *ESAG6* were isolated from both species and cloned into pBluescript SK(−). Bacterial clones were screened by colony hybridization for the presence of either *ESAG6* or *ESAG7*. Clones with a single insert were sequenced on both strands.

In *T. equiperdum*, of 151 positive clones, 114 had *ESAG6* sequences (see Fig. 3), 32 had *ESAG7* sequences (not shown), and 5 had unknown sequences. We compared the 114 *T. equiperdum ESAG6* sequences and found 10 sequence variants. We conducted the same, but smaller-scale, analysis on *T. brucei* as a positive control and pooled the data (six variants from 10 *ESAG6* sequences) together with 23 *T. brucei ESAG6* sequences available in the database. Although the *T. brucei ESAG6* sequences present in the database were derived from multiple *T. brucei* strains, a large number of the *ESAG6* sequences were derived from the *T. brucei* 427 strain (Ansorge et al. 1999). In addition, a comparable degree of genetic diversity was found in a large set of genomic *ESAG6* sequences isolated exclusively from

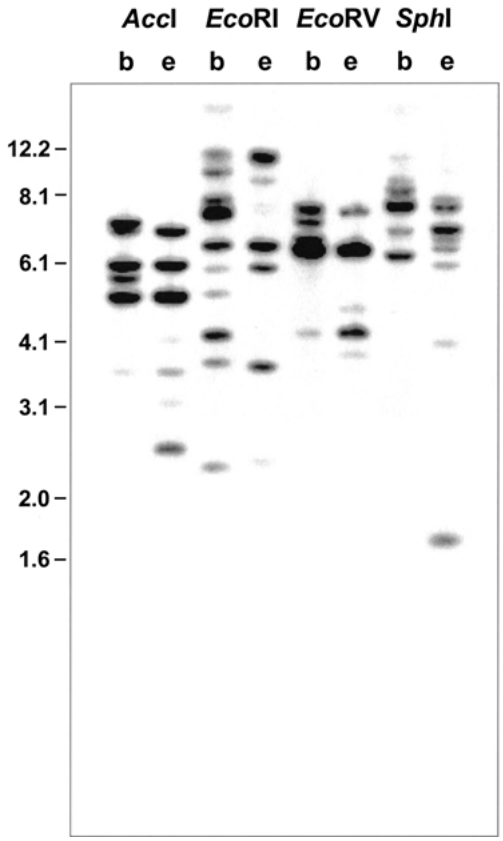

Fig. 2. *ESAG6* and *ESAG7* distribution in *T. equiperdum* and *T. brucei*. Equal amounts of genomic DNA from *T. brucei* (b) and *T. equiperdum* (e) as quantitated on ethidium bromide-stained gels were digested with the restriction enzymes indicated *above* the panel. DNA was separated in an agarose gel and transferred to a membrane via Southern blotting, and the blot was hybridized with the DES12 probe (*Dra*I/*Hind*III fragment indicated in Fig. 1A). This probe hybridizes with both *ESAG7* and *ESAG6*. The blot was washed to an end stringency of 0.3× SSC at 65°C for 30 min. Sizes of a DNA marker are indicated in kilobases at the *left*. *ESAG6* quantitation was performed using a PhosphorImager (BioRad).

the *T. brucei* 427 strain (van Luenen and Borst, unpublished results). For *T. brucei* we found a total of 20 *ESAG6* variants in the 33 sequences that we compared. Only one sequence was found in both *T. brucei* and *T. equiperdum* (TI-1b and TI-1e). Considerable sequence variation was found in the hypervariable region (box labeled HV in Fig. 3) (Zomerdijk et al. 1990) in both species. This hypervariable region has been shown not to be involved in transferrin binding (Salmon et al. 1997) but has been proposed to play a role in antigenic variation of a surface receptor (Borst 1991). The hypervariable domain in a proposed putative transferrin receptor structure is not as exposed on the surface as the transferrin binding site (Salmon et al. 1997). However, the precise protein structure of this molecule has not yet been determined. This explanation would fit with the variability of the hypervariable region in *ESAG6* genes from both *T. brucei* and *T. equiperdum*. Antigenic variation, rather than adaptation to multiple hosts, would be a selection pressure operating on both parasites.

Two areas of the ESAG6 protein involved in transferrin binding, as established by Salmon et al. (1997), are indicated in Fig. 3 as Box I and Box II. *T. brucei*, which infects a wide range of mammalian species with different polymorphic transferrins, shows considerable genetic diversity in these regions of the transferrin receptor genes. In contrast, in *T. equiperdum* the degree of genetic diversity in Box I and Box II is greatly reduced, as over 90% of the isolated clones contain the same sequence in these regions. This indicates a correlation between genetic diversity in these regions and size of trypanosome host range.

To determine whether these differences in genetic diversity were caused by underlying differences in selection pressure, we undertook a ML analysis of d_N/d_S ratios in the *ESAG6* sequences from *T. equiperdum* and *T. brucei*. In both species there was strong evidence of positive selection, as model M8 (see Materials and Methods) was significantly favored over model M7 in all cases ($p < 0.001$) and contained a class of codons with $d_N/d_S > 1$. Similar results were seen for the other models of codon evolution (results not shown; available from the authors on request). Further, all models gave similar values for the strength of positive selection: $d_N/d_S = 5.5$ for *T. equiperdum* and $d_N/d_S = 4.7$ for *T. brucei*. In the case of *T. brucei*, the Bayesian approach identified sites 16, 17, 18, 22, 51, 66, and 94 as having greater than >95% chance of falling into the positively selected class under model M8. Four of these sites (16, 17, 18, and 22) fell in the hypervariable region, and, significantly, one (site 94) fell in the transferrin binding region (Box II) that is conserved in *T. equiperdum*. This confirms that the selective regime differs between these species. Unfortunately, the small number of sequences available for *T. equiperdum* meant that we were unable to use Bayesian methods to identify individual positively selected sites accurately in this case. Despite sequencing 114 *T. equiperdum ESAG6* genomic clones, we only found 10 of the expected total of 16 sequence variants estimated by PCR. It is likely that multiple *T. equiperdum ESAG6* genes are identical in this area of the gene.

As it is probable that *T. equiperdum* has evolved from *T. brucei* after losing the ability to be passaged through tsetse flies (Brun et al. 1998), it is likely that the genetic diversity present in an earlier *T. equiperdum* ancestor was lost. Phylogenetic analysis comparing the *T. equiperdum ESAG6* sequences with those of *T. brucei* shows that they cluster within a subset of the *T. brucei* sequences, although they do not form a monophyletic group (Fig. 4). This indicates that in the absence of the selection pressure operating on a parasite needing to infect multiple

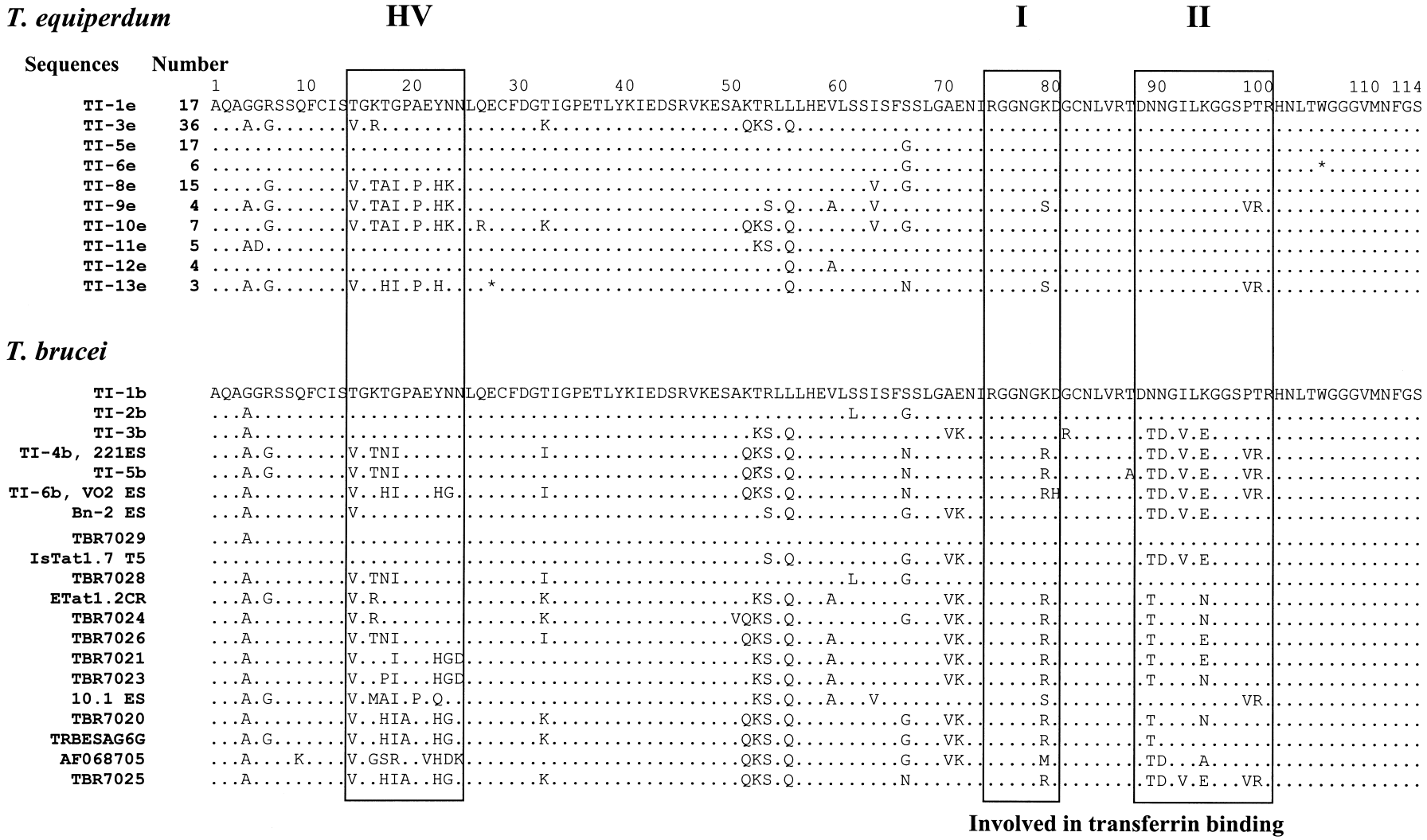

Fig. 3. Amino acid sequence alignment of ESAG6 in *T. equiperdum* and *T. brucei*. We isolated 114 *T. equiperdum ESAG6* genomic clones which had 10 different sequences (TI-1e to TI-13e). The number of identical clones sequenced is indicated next to the name of the sequence. The *T. equiperdum ESAG6* amino acid alignments were compared with those from 20 different *T. brucei ESAG6* sequences within a stretch of 114 amino acids. Only one *ESAG6* sequence was found in both *T. brucei* and *T. equiperdum* (TI-1b and TI-1e). Residues identical to TI-1b and TI-1e are represented by *dots*. *Asterisks* represent stop codons in TI-6e and TI-13e. The hypervariable region in *T. brucei ESAG6* (Zomerdijk et al. 1990) not involved in transferrin binding (Salmon et al. 1997) is indicated by the *box* labeled HV. Two regions of the ESAG6 protein that have been shown to be part of the transferrin binding site (Box I and Box II from Fig. 2 of Salmon et al. 1997) are indicated by *boxes*. The accession numbers of the *T. brucei* and *T. equiperdum ESAG6* sequences shown here are listed in Materials and Methods. Accession numbers for additional identical sequences are also listed in Materials and Methods.

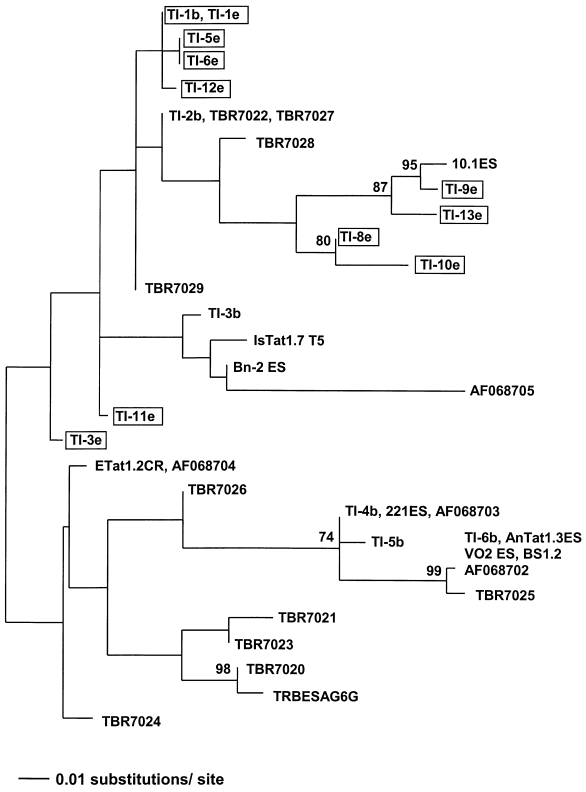

Fig. 4. Maximum likelihood phylogenetic tree of *ESAG6* sequences from *T. equiperdum* and *T. brucei*. The sequences are as indicated in Fig. 3. The names of the *T. equiperdum* sequences are *boxed*. Some different names for the same sequence are also indicated. The tree is midpoint rooted for purposes of clarity only, and all horizontal branch lengths are drawn to scale. Bootstrap support values (>70% in 1000 replications) are shown for key nodes only.

mammalian hosts, genetic diversity in genes involved in host adaptation appears to be reduced.

Discussion

We attempted to address whether there is a correlation between the genetic diversity of genes shown to be involved in parasite–host adaptation and the size of the host range. We show that although *T. equiperdum* and *T. brucei* have approximately similar numbers of *ESAG6* transferrin receptor genes, the *T. equiperdum ESAG6* sequences are related to a subset of the *T. brucei ESAG6* sequences, indicating that they are less diverse. Our ML analysis of selection pressures acting on *ESAG6* also found significant evidence of positive selection acting on both *T. equiperdum* and *T. brucei*. However, the *T. equiperdum ESAG6* genes appear to be less diverse than those of *T. brucei* in the *ESAG6* Box I and Box II shown to be directly involved in transferrin binding (Salmon et al. 1994), the latter of which shows evidence of positive selection in *T. brucei*.

As *T. equiperdum* naturally infects only equines rather than the large range of mammals which can be infected by *T. brucei*, this is consistent with genetic diversity being lost in the absence of selection pressure. This finding supports the proposal that the multiplicity of polymorphic *VSG* ESs confers an advantage to *T. brucei* in adapting to multiple hosts (Bitter et al. 1998; Gerrits et al. 2002). Although *T. equiperdum* and *T. brucei* primarily inhabit different tissues, this is unlikely to explain the observed differences in the degree of genetic diversity of the transferrin receptor genes within the two species. Uptake of transferrin within different tissues of one host should not require different sets of parasite transferrin receptors.

In general in microorganisms, genetic diversity appears to get lost in the absence of positive selection pressure (Moran 2002). We assume that this is also the case in *T. brucei* and *T. equiperdum*. As *T. equiperdum* is evidently closely related to *T. brucei*, this loss of genetic diversity in some regions of their ES genes must have occurred relatively recently. There is indirect evidence that homogenization could be continuously occurring in *VSG* ES sequences, which would speed up the loss of diversity in the absence of selection pressure to maintain it. For example, comparison of all known sequenced *T. brucei VSG* ESs shows the conservation of an *ESAG3* pseudogene downstream of *ESAG5* in five of the six ESs (Berriman et al. 2002). In addition, an *ESAG4* pseudogene is present upstream of multiple *T. brucei VSG* ESs (Berriman et al. 2002). It is unclear why these nonfunctional ES sequences have been maintained. One possibility is that *T. brucei VSG* ESs have arisen relatively recently from a single precursor. Alternatively, gene conversion events between *VSG* ES-containing telomeres could result in their homogenization in the absence of a selective pressure maintaining genetic diversity. Gene conversion events between telomeric *VSG* ESs appear to be a relatively frequent event in *T. brucei*, as they play a role in *VSG* switching (McCulloch et al. 1997; Robinson et al. 1999). Gene conversion plays a major role in homogenization of other multicopy gene families including the multiple ribosomal DNA transcription units in yeast and *Drosophila* (Schlotterer and Tautz 1994; Gangloff et al. 1996).

In contrast to the results with the *ESAG6* transferrin binding site, there is clear sequence variation in the hypervariable domain of *ESAG6* in both *T. equiperdum* and *T. brucei* (Zomerdijk et al. 1990). It has been proposed that variability in this region could play a role in antigenic variation of a surface receptor (Borst 1991). Mutations in this region (mutations 7 and 8 in Table I of Salmon et al. [1997]) did not significantly affect transferrin binding. The transferrin receptor appears to be very similar in structure to a VSG (Salmon et al. 1997). Superposition of the transferrin receptor amino acid sequence on the

known VSG structure indicates that the residues of the transferrin binding site are present in loops that are particularly exposed to the extracellular environment (Salmon et al. 1997). Although the hypervariable domain is also near the tip of the molecule, it is unclear whether it is antigenically exposed. Antigenic variation of a surface receptor therefore remains possible. *T. equiperdum*, although naturally infecting only a very restricted number of species, would be exposed to the same immune selection pressures as *T. brucei* when present in the bloodstream.

Why would *T. equiperdum* have so many *VSG* ESs when, in theory, it only needs one? One possibility is that *T. equiperdum* has diverged from *T. brucei* relatively recently, so that there has been insufficient time for a drastic loss of genetic diversity. However, the presence of multiple *VSG* ESs presumably also confers additional options to a trypanosome switching *VSG* genes during antigenic variation. One mechanism mediating a *VSG* switch is an *in situ* switch between *VSG* ESs (Bernards et al. 1984). Another mechanism involved in *VSG* switching is telomere conversion, which appears to be particularly frequent between *VSG* ES-containing telomeres (McCulloch et al. 1997; Robinson et al. 1999). A large number of expression sites would facilitate switching via either of these mechanisms.

T. equiperdum can infect mammals in addition to equines and can be maintained in laboratory mammals such as mice, rats, and rabbits in experimental infections. However, this is an artificial situation and does not reproduce the selection pressures operating in the wild. For example, *T. brucei* variants with different active *VSG* expression sites can be effectively grown in laboratory rodents or in fetal calf serum, even when the transferrin receptor expressed has suboptimal affinity for the transferrin source in the growth medium (Bitter et al. 1998; Salmon et al. 1997; Gerrits et al. 2002). However, in mixed infections present in the field, trypanosomes expressing suboptimal transferrin receptor subunits are presumably competed out.

It is plausible that the extensive polymorphisms found in the large gene families present in different *T. brucei VSG* expression sites play a role in host adaptation (Bitter et al. 1998; Pays et al. 2001). The challenge comes in determining the selective advantage that this genetic diversity confers.

Acknowledgments. We are grateful to Wendy Gibson (University of Bristol) for providing *T. equiperdum* and *T. brucei* strains and DNA. We thank Henri van Luenen and Piet Borst (The Netherlands Cancer Institute) for sharing unpublished data and discussing the problem of chimeric products generated during PCR amplification of large families composed of similar genes. We thank Wendy Gibson, Piet Borst, Henri van Luenen, Niall Aitcheson, Katie Hughes, Karen Sheader, and Daniëlle te Vruchte for comments on the manuscript and Daniel Farchy for assistance with this project. This work was funded by the Wellcome Trust. G.R. is a Wellcome Senior Fellow in the Basic Biomedical Sciences.

References

Alarcon CM, Pedram M, Donelson J (1999) Leaky transcription of variant surface glycoprotein gene expression sites in bloodstream African trypanosomes. J Biol Chem 274:16884–16893

Alemu T, Luckins AG, Phipps LP, Reid SW, Holmes PH (1997) The use of enzyme linked immunosorbent assays to investigate the prevalence of *Trypanosoma equiperdum* in Ethiopian horses. Vet Parasitol 71:239–250

Ansorge I, Steverding D, Melville S, Hartmann C, Clayton C (1999) Transcription of 'inactive' expression sites in African trypanosomes leads to expression of multiple transferrin receptor RNAs in bloodstream forms. Mol Biochem Parasitol 101: 81–94

Barry JD, McCulloch R (2001) Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite. Adv Parasitol 49:1–70

Bernards A, de Lange T, Michels PA, Liu AY, Huisman MJ, Borst P (1984) Two modes of activation of a single surface antigen gene of *Trypanosoma brucei*. Cell 36:163–170

Berriman M, Hall N, Sheader K, Bringaud F, Tiwari B, Isobe T, Bowman S, Corton C, Clark L, Cross GAM, Hoek M, Zanders T, Berberof M, Borst P, Rudenko G (2002) The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. Mol Biochem Parasitol 122:131–140

Biteau N, Bringaud F, Gibson W, True P, Baltz T (2000) Characterization of *Trypanozoon* isolates using a repeated coding sequence and microsatellite markers. Mol Biochem Parasitol 105:185–201

Bitter W, Gerrits H, Kieft R, Borst P (1998) The role of transferrin-receptor variation in the host range of *Trypanosoma brucei*. Nature 391:499–502

Borst P (1991) Transferrin receptor, antigenic variation and the prospect of a trypanosome vaccine. Trends Genet 7:307–309.

Borst P, Bitter W, Blundell P, Cross M, McCulloch R, Rudenko G, Taylor M, van Leeuwen F (1997) The expression sites for variant surface glycoproteins of *Trypanosoma brucei*. In: Hide G, Mottram JC, Coombs GH, Holmes PH (eds) Trypanosomiasis and leishmaniasis. Oxford, UK, CAB International, pp 109–131

Borst P, Bitter W, Blundell PA, Chaves I, Cross M, Gerrits H, van Leeuwen F, McCulloch R, Taylor M, Rudenko G (1998) Control of *VSG* gene expression sites in *Trypanosoma brucei*. Mol Biochem Parasitol 91:67–76

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

Capbern A, Giroud C, Baltz T, Mattern P (1977) *Trypanosoma equiperdum*: Antigenic variation in experimental trypanosomiasis of rabbits. Exp Parasitol 42:6–13

Chaves I, Rudenko G, Dirks-Mulder A, Cross M, Borst P (1999) Control of variant surface glycoprotein gene-expression sites in *Trypanosoma brucei*. EMBO J 18:4846–4855

Cross GA, Wirtz LE, Navarro M (1998) Regulation of *vsg* expression site transcription and switching in *Trypanosoma brucei*. Mol Biochem Parasitol 91:77–91

Dai Do Thi C, Aerts D, Steinert M, Pays E (1991) High homology between variant surface glycoprotein gene expression sites of *Trypanosoma brucei* and *Trypanosoma gambiense*. Mol Biochem Parasitol 48:199–210

Florent IC, Raibaud A, Eisen H (1991) A family of genes related to a new expression site-associated gene in *Trypanosoma equiperdum*. Mol Cell Biol 11:2180–2188

Frasch ACC, Hajduk SL, Hoeijmakers JHJ, Borst P, Brunel F, Davison J (1980) The kinetoplast DNA of *Trvpanosoma equiperdum*. Biochim Biophys Acta 607:397–410

Gangloff S, Zou H, Rothstein R (1996) Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. EMBO J 15:1715–1725

Gerrits H, MuBmann R, Bitter W, Kieft R, Borst P (2002) The physiological significance of transferrin receptor variations in *Trypanosoma brucei*. Mol Biochem Parasitol 119:237–247

Haag J, O'Huigin C, Overath P (1998) The molecular phylogeny of trypanosomes: Evidence for an early divergence of the Salivaria. Mol Biochem Parasitol 91:37–49

Hobbs MR, Boothroyd JC (1990) The expression-site-associated gene family of trypanosomes is expressed in vivo and shows homology to a variant surface glycoprotein gene. Mol Biochem Parasitol 43:1–16

Kooter JM, van der Spek HJ, Wagter R, d'Oliveira CE, van der Hoeven F, Johnson PJ, Borst P (1987) The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei*. Cell 51:261–272

LaCount DJ, El-Sayed NM, Kaul S, Wanless D, Turner CM, Donelson JE (2001) Analysis of a donor gene region for a variant surface glycoprotein and its expression site in African trypanosomes. Nucleic Acids Res 29:2012–2019

Ligtenberg MJ, Bitter W, Kieft R, Steverding D, Janssen H, Calafat J, Borst P (1994) Reconstitution of a surface transferrin binding complex in insect form *Trypanosoma brucei*. EMBO J 13:2565–2573

Lips S, Revelard P, Pays E (1993) Identification of a new expression site-associated gene in the complete 30.5 kb sequence from the AnTat 1.3A variant surface protein gene expression site of *Trypanosoma brucei*. Mol Biochem Parasitol 62:135–137

Lodes MJ, Smiley BL, Stadnyk AW, Bennett JL, Myler PJ, Stuart K (1993) Expression of a retroposon-like sequence upstream of the putative *Trypanosoma brucei* variant surface glycoprotein gene expression site promoter. Mol Cell Biol 13:7036–7044

Longacre S, Hibner U, Raibaud A, Eisen H, Baltz T, Giroud C, Baltz T (1983) DNA rearrangements and antigenic variation in *Trypanosoma equiperdum*: Multiple expression-linked sites in independent isolates of trypanosomes expressing the same antigen. Mol Cell Biol 3:399–409

Lun ZR, Allingham R, Brun R, Lanham SM (1992a) The isoenzyme characteristics of *Trypanosoma evansi* and *Trypanosoma equiperdum* isolated from domestic stocks in China. Ann Trop Med Parasitol 86:333–340

Lun ZR, Brun R, Gibson W (1992b) Kinetoplast DNA and molecular karyotypes of *Trypanosoma evansi* and *Trypanosoma equiperdum* from China. Mol Biochem Parasitol 50:189–196

McCulloch R, Rudenko G, Borst P (1997) Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. Mol Cell Biol 17:833-843

Moran N (2002) Microbial minimalism: Genome reduction in bacterial pathogens. Cell 108:583–586

Navarro M, Cross GA (1996) DNA rearrangements associated with multiple consecutive directed antigenic switches in *Trypanosoma brucei*. Mol Cell Biol 16:3615–3625

North MJ, Coombs GH, Barry JD (1983) A comparative study of the proteolytic enzymes of *Trypanosoma brucei*, *T. equiperdum*, *T. evansi*, *T. vivax*, *Leishmania tarentolae* and *Crithidia fasciculata*. Mol Biochem Parasitol 9:161–180

Pays E, Nolan DP (1998) Expression and function of surface proteins in *Trypanosoma brucei*. Mol Biochem Parasitol 91: 3–36

Pays E, Tebabi P, Pays A, Coquelet H, Revelard P, Salmon D, Steinert M (1989) The genes and transcripts of an antigen gene expression site from *T. brucei*. Cell 57:835–845.

Pays E, Lips S, Nolan D, Vanhamme L, Perez-Morga D (2001) The *VSG* expression sites of *Trypanosoma brucei*: Multipurpose tools for the adaptation of the parasite to mammalian hosts. Mol Biochem Parasitol 114:1–16

Raibaud A, Buck G, Baltz T, Eisen H (1986) Cloning and characterization of a variant surface glycoprotein expression site from *Trypanosoma equiperdum*. Mol Cell Biol 6:2950–2956

Redpath MB, Windle H, Nolan D, Pays E, Voorheis HP, Carrington M (2000) ESAG11, a new *VSG* expression site-associated gene from *Trypanosoma brucei*. Mol Biochem Parasitol 111:223–228

Revelard P, Lips S, Pays E (1990) A gene from the *VSG* expression site of *Trypanosoma brucei* encodes a protein with both leucine-rich repeats and a putative zinc finger. Nucleic Acids Res 18:7299–7303

Riou GF, Saucier J.-M. (1979) Characterization of the molecular components in kinetoplast-mitochondrial DNA of *Trypanosoma equiperdum*. J Cell Biol 82:248–263

Robinson NP, Burman N, Melville S, Barry JD (1999) Predominance of duplicative *VSG* gene conversion in antigenic variation in African trypanosomes. Mol Cell Biol 19:5839–5846

Ross DT, Raibaud A, Florent IC, Sather S, Gross MK, Storm DR, Eisen H (1991) The trypanosome *VSG* expression site encodes adenylate cyclase and a leucine-rich putative regulatory gene. EMBO J 10:2047–2053

Rudenko G, Blundell P, Taylor MC, Kieft R, Borst P (1994) *VSG* gene expression site control in insect form *Trypanosoma brucei*. EMBO J 13:5470–5482

Salmon D, Geuskens M, Hanocq F, Hanocq-Quertier J, Nolan D, Ruben L, Pays E (1994) A novel heterodimeric transferrin receptor encoded by a pair of *VSG* expression site-associated genes in *T. brucei*. Cell 78:75–86

Salmon D, Hanocq-Quertier J, Paturiaux-Hanocq F, Pays A, Tebabi P, Nolan DP, Michel A, Pays E (1997) Characterization of the ligand-binding site of the transferrin receptor in *Trypanosoma brucei* demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. EMBO J 16:7272–7278

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schell D, Evers R, Preis D, Ziegelbauer K, Kiefer H, Lottspeich F, Cornelissen AW, Overath P (1991) A transferrin-binding protein of *Trypanosoma brucei* is encoded by one of the genes in the variant surface glycoprotein gene expression site. EMBO J 10:1061–1066

Schlotterer C, Tautz D (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4:777–783

Shu HH, Stuart K (1994) Mitochondrial transcripts are processed but are not edited normally in *Trypanosoma equiperdum* (ATCC 30019) which has kDNA sequence deletion and duplication. Nucleic Acids Res 22:1696–1700

Smiley BL, Stadnyk AW, Myler PJ, Stuart K (1990) The trypanosome leucine repeat gene in the variant surface glycoprotein expression site encodes a putative metal-binding domain and a region resembling protein-binding domains of yeast, *Drosophila*, and mammalian proteins. Mol Cell Biol 10:6436–6444

Stevens JR, Noyes HA, Dover GA, Gibson WC (1999) The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. Parasitology 118:107–116

Steverding D, Stierhof YD, Fuchs H, Tauber R, Overath P (1995) Transferrin-binding protein complex is the receptor for transferrin uptake in *Trypanosoma brucei*. J Cell Biol 131:1173–1182

Swofford DL (2002). PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA

Tanabe K, Sakihama N, Färnert A, Rooth I, Björkman A, Walliker D, Ranford-Cartwright L (2002) In vitro recombination during PCR of *Plasmodium falciparum* DNA: A potential pitfall in molecular population genetic analysis. Mol Biochem Parasitol 122:211–216

Van der Ploeg LHT, Bernards A, Rijsewijk FAM, Borst P (1982) Characterization of the DNA duplication-transposition that controls the expression of two genes for variant surface glycoproteins in *Trypanosoma brucei*. Nucleic Acids Res 10:593–609

Vanhamme L, Pays E, McCulloch R, Barry JD (2001) An update on antigenic variation in African trypanosomes. Trends Parasitol 17:338–343

Xong HV, Vanhamme L, Chamekh M, Chimfwembe CE, van den Abbeele J, Pays A, van Meirvenne N, Hamers R, de Baetselier P, Pays E (1998) A *VSG* expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. Cell 95:839–846

Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13:555–556

Yang Z, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zomerdijk JC, Ouellette M, ten Asbroek AL, Kieft R, Bommer AM, Clayton CE, Borst P (1990) The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J 9:2791–2801