ELSEVIER

Molecular & Biochemical Parasitology 122 (2002) 131–140

MOLECULAR & BIOCHEMICAL PARASITOLOGY

www.parasitology-online.com

The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*☆

Matthew Berriman[a], Neil Hall[a], Karen Sheader[b], Frédéric Bringaud[c], Bela Tiwari[d], Tomoko Isobe[b], Sharen Bowman[a], Craig Corton[a], Louise Clark[a], George A.M. Cross[e], Maarten Hoek[e,1], Tyiesha Zanders[e], Magali Berberof[f], Piet Borst[f], Gloria Rudenko[b,*]

[a] *The Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK*
[b] *The Peter Medawar Building for Pathogen Research, University of Oxford, Oxford OX1 3SY, UK*
[c] *Laboratoire de Parasitologie Moléculaire, Université Victor Segalen Bordeaux II, Bordeaux, France*
[d] *Oxford University Bioinformatics Centre, Oxford, UK*
[e] *The Rockefeller University, New York, NY 10021, USA*
[f] *The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066CX, The Netherlands*

Received 30 January 2002; accepted in revised form 23 April 2002

Abstract

Trypanosoma brucei evades the immune system by switching between Variant Surface Glycoprotein (VSG) genes. The active *VSG* gene is transcribed in one of approximately 20 telomeric expression sites (ESs). It has been postulated that ES polymorphism plays a role in host adaptation. To gain more insight into ES architecture, we have determined the complete sequence of Bacterial Artificial Chromosomes (BACs) containing DNA from three ESs and their flanking regions. There was variation in the order and number of ES-associated genes (*ESAG*s). *ESAG*s 6 and 7, encoding transferrin receptor subunits, are the only *ESAG*s with functional copies in every ES that has been sequenced until now. A BAC clone containing the VO2 ES sequences comprised approximately half of a 330 kb 'intermediate' chromosome. The extensive similarity between this intermediate chromosome and the left telomere of *T. brucei* 927 chromosome I, suggests that this previously uncharacterised intermediate size class of chromosomes could have arisen from breakage of megabase chromosomes. Unexpected conservation of sequences, including pseudogenes, indicates that the multiple ESs could have arisen through a relatively recent amplification of a single ES.

Keywords: Antigenic variation; Expression site sequence; Genome project; VSG; Variant surface glycoprotein genes; Telomere; *Trypanosoma brucei*

1. Introduction

Trypanosoma brucei effectively evades the immune response of the mammals that it infects by continuously changing a homogeneous Variant Surface Glycoprotein (VSG) coat. *T. brucei* has hundreds of *VSG* genes and pseudogenes, but only one *VSG* is expressed at a time, from one of several telomeric transcription units known as *VSG* expression sites (ESs). Changing the active *VSG* frequently involves gene conversion, whereby a copy of a silent *VSG* is transposed into the active ES, displacing the existing *VSG*. Alternatively, *VSG* switching can be

Abbreviations: ES, expression site; ESAG, expression-site associated gene; LRRP, leucine-rich repetitive protein; ORF, open reading frame; RHS, retrotransposon hot spot; SRA, serum resistance associated; VSG, variant surface glycoprotein.

☆ *Note:* Nucleotide sequence data reported in this paper are available in the EMBL, GenBank™ and DDJB databases under the accession numbers: AL671259, AL671256, AL670322.

* Corresponding author. Tel.: +44-1865-281-548; fax: +44-1865-281-894

E-mail address: gloria.rudenko@medawar.ox.ac.uk (G. Rudenko).

[1] Present address: Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA.

0166-6851/02/$ - see front matter
PII: S0166-6851(02)00092-0

achieved by switching from one ES to another (reviewed in: [1–4]).

VSG ESs are large polycistronic transcription units varying in size from about 30 to 60 kb [5–7]. In addition to the telomeric *VSG*, each ES contains several classes of ES-associated genes (*ESAG*s) (reviewed in [2,8]). The function of only a few *ESAG*s is known. *ESAG6* and *ESAG7* encode the subunits of a heterodimeric transferrin receptor, allowing the trypanosome to obtain iron in a form that has been sequestered by the host [9,10]. *ESAG4* encodes an adenylate cyclase, which can rescue adenylate cyclase deficient mutants in yeast [11]. *ESAG10* is homologous to the BT1 biopterin transporter of *Leishmania* [12]. The serum resistance associated (*SRA*) gene, which confers human infectivity to *T. brucei* through an unknown mechanism, is also ES-associated in the one strain in which it has been characterised [5].

Sequence polymorphisms in *ESAG6* and *7* affect the affinities of the transferrin receptors for the transferrin molecules from different mammalian hosts [13]. As *T. brucei* can infect many mammalian species, this could provide a reason for the existence of multiple ESs, which then requires a mechanism to ensure mutually exclusive VSG expression [14]. The role of the *SRA* gene in human infectivity [5] supports the idea that *ESAG*s could play a role in host adaptation. However, the function of *ESAG*s other than 4, 6, 7 and 10 is more speculative, and based on recognisable protein motifs. It is also unclear which *ESAG*s are essential ES components. Some *ESAG*s (1, 3 and 4, for example) are members of large gene families that are also present in non-ES locations [15–17]. *ESAG8* appears to be exclusively ES-located, but does not appear to be an essential gene under the laboratory conditions tested [5,18]. If the host adaptation hypothesis is correct, it is possible that some *ESAG*s will be essential or advantageous only in some host environments.

The sequence of the *T. brucei* 927 genome is currently being determined. However, ES sequences are highly underrepresented in standard large-insert libraries. Determining the sequence of telomeric ESs will require specific cloning efforts. Little is known about the extent of ES polymorphism. In order to get more insight into this variability, we have determined the contiguous DNA sequences of three BAC clones containing sequences from three *T. brucei* 427 bloodstream-form ESs. These sequences included flanking regions extending for up to one hundred kilobases upstream of the ES promoters. These data allowed us to evaluate the overall architecture of six *T. brucei* ESs, four of which are complete. There is an overall conservation of ES architecture, but individual ESs may contain different numbers of functional *ESAG*s and pseudogenes.

2. Materials and methods

2.1. Bacterial artificial chromosome ES clones

ES clones were isolated from BAC libraries (P. de Jong, Children's Hospital Oakland Research Institute: http://www.chori.org/bacpac/) made from clones of *T. brucei* strain 427, variant 221a [19,20] into which specific ES tags had been introduced. BAC H25N7 (containing the 221 *VSG* ES on a 3.2 Mb *T. brucei* chromosome-VIa [20]) and BAC N19B2 (containing part of the VO2 *VSG* ES on a 330 kb chromosome [21]) were isolated from BAC library RPCI-97, which was made in the vector pBACe3.6 [22] with partial *Eco*RI-digested genomic DNA of *T. brucei* transformant HNI, which contains a hygromycin resistance gene downstream of the promoter of the 221 ES and a neomycin resistance gene downstream of the promoter of the VO2 ES [21]. Four independent BACs containing the 221 ES and five independent BACs containing the VO2 ES were isolated, using the hygromycin or neomycin resistance genes as probes. BAC 13J3 was isolated from library RPCI-102, which was made in the vector pTARBAC1 [23] from partial *Mbo*I-digested DNA from a *T. brucei* 221 cell line containing genes for hygromycin, neomycin and bleomycin resistance downstream of ES promoters on chromosomes VIa, IVa and on a 300-kb 'intermediate' chromosome, respectively (Zeng et al., manuscript in preparation).

2.2. Sequence determination and assembly

Three BAC ES clones were fully sequenced using a two-stage strategy involving random sequencing of subcloned DNA followed by directed sequencing to resolve problem areas [24]. In the first stage, DNA from prepared BAC clones was shattered by sonification and fragments of 1.4–2 kb were cloned into pUC18. The DNAs from randomly selected clones were sequenced with dye-terminator chemistry and analysed on automatic sequencers. Each BAC was sequenced to a depth of 7-fold sequence coverage. Contiguous sequences were assembled using the PHRAP software (Phil Green, University of Washington) [25]. Manual base calling and finishing was carried out using Gap4 software (http://www.mrc-lmb.cam.ac.uk/pubseq/manual/gap4_unix_1.html). Gaps and low quality regions of the sequence were resolved by techniques such as primer walking, PCR and re-sequencing clones under conditions giving increased read lengths. Once the inserts had been resolved into large contiguous sequences, the assemblies were verified against restriction maps.

2.3. Sequence comparisons

BAC sequences were annotated using Artemis sequence analysis software (http://www.sanger.ac.uk/Software/Artemis/) [26], and sequence comparisons were performed using Artemis Comparison Tool (ACT) (http://www.sanger.ac.uk/Software/ACT/). Results presented were the results of BLASTN comparisons processed by MSP crunch. The figures shown were made using the default setting, meaning that all matches are shown. Sequence comparisons were performed after masking various *T. brucei* repetitive sequences *RHS* (pseudo)genes [27], *ingi* retroelements [28,29], ribosomal mobile elements (*RIME*) [30] and the 50-bp repeats [31]. Protein sequence motifs were determined using the PFAM (http://www.sanger.ac.uk/Software/Pfam/) and SMART (http://smart.embl-heidelberg.de/) databases.

2.4. ES analysis

Pulsed field gels were run using a CHEF DRIII (BioRad) electrophoresis system. Separations of the VO2 chromosome were performed in 1% agarose gels run at 6 V cm^{-1}, using a 25 s switching time for 20 h in 0.5 × TBE buffer at 14 °C [32]. Separations of the chromosome containing the 221 ES were performed according to [33] using a ramp of 1400–700 s for 144 h at 2.5 V cm^{-1}. Southern blots were performed according to [32], and washed at a stringency of 0.1 × SSC. Probe LRR is a 697-bp fragment from the leucine-rich repetitive protein (LRRP) gene in BAC H25N7, which was PCR-amplified using 5′-ATGTTGAAAAGGCTTTGTCTCAG-3′ and 5′-CTCCACGAGTGTAACAATGCTG-3′ as sense and antisense primers, respectively. Probe DES12 is the *Dra*I-*Hind*III fragment, which includes *ESAG7*, from the DES promoter region indicated in Fig. 1 of Ref. [34].

3. Results

BACs provide an efficient means of cloning DNA inserts of up to 300 kilobases [35,36]. Since it is difficult to distinguish between different ESs, BAC libraries were made from *T. brucei* lines in which single-copy drug-resistance genes had been inserted immediately downstream of the promoters of specific ESs, and BAC clones were isolated using the marker genes as probes. ES BACs were about ten-fold underrepresented compared with BACs containing chromosome-internal genes.

The 55-kb 221 ES is the largest ES described, so far, in *T. brucei* [6,7]. A schematic interpretation of the sequence shows that this ES has undergone a duplication of *ESAG3* and *ESAG4* and a triplication of *ESAG8* (Fig. 1). Directly upstream of the ES promoters are long regions of 50-bp repeats [31]. In the 221 and VO2 ES, these repeats extend for 44 and 49 kb, respectively [21]. The 50-bp repeat arrays are smaller in the BAC clones than in the genome (mapping results not shown), presumably due to slippage during replication in the *E. coli* DH10B bacteria used for DNA amplification. With the exception of simple repeat collapse, no other rearrangements or deletions were detected in the ES BACs. The Bn-2 ES contains a duplicated promoter, an organisation present in approximately half of the ESs of *T. brucei* 427 [37,38].

The 221 ES contains an *ESAG5* pseudogene: an 'extra' G in a stretch of 7 Gs causes a frameshift. *Escherichia coli* can have difficulty replicating homopolymeric G-tracts resulting in slippage [39], but this does not appear to be the source of the *ESAG5* frameshift. Analysis of 14 *ESAG5* sequences cloned by reverse-transcriptase PCR from VSG 221 cells showed that half of the sequences corresponded to the frameshifted 221 *ESAG5* (data not shown). Two other *ESAG5* genes were also represented in the mRNA population. A truncated ES, containing only *ESAG*s 5, 6, and 7, *SRA* and *VSG* has been described [5]. The occurrence of a frameshift in *ESAG5* suggests that an ES containing only *ESAG*s 6 and 7 might be sufficient for survival in the bloodstream. However, because of their proximity to a 'leaky' promoter, *ESAG*s 6 and 7 are also transcribed, at a reduced rate, from multiple 'silent' ESs [40,41], and this might also be the case for *ESAG5*. Alternatively, functional copies of *ESAG5* could be located outside of ESs [42]. In addition to the *ESAG* duplications and triplications present in the 221 ES, the VO2 ES has two copies of *ESAG7*.

There are extensive tracts of repetitive elements, including the retroposons *RIME* [30] and *ingi* [28,29] upstream of the 221 and VO2 ESs, as has been found upstream of the truncated ES on the left telomere of *T. brucei* 927 chromosome I [43] and upstream of the *VSG* 10.1 ES [44]. In addition to *RIME* and *ingi*, the 221 and VO2 ESs are flanked by extensive arrays of a recently described multigene family called *RHS* (Retrotransposon Hot Spot) [27] (see Fig. 1). RHS coding sequences have been divided into six sub-families according to the divergent C-terminal domain of their gene product. This highly repetitive multigene family is composed of about 280 copies per diploid genome, about two-thirds of them are non-functional pseudogenes. The RHS (pseudo)genes appear to be frequently located in the subtelomeres adjacent to *VSG* ESs, including the size-polymorphic telomeric repetitive regions described in chromosome I [27].

In addition to RHS (pseudo)genes, a new *LRRP* gene was found upstream of both the 221 and VO2 ESs. Three copies of *LRRP* are also found upstream of the truncated ES on the left telomere of *T. brucei* 927 chromosome I (EMBL accession number AL359782; manuscript in preparation). In BLAST searches, LRRP

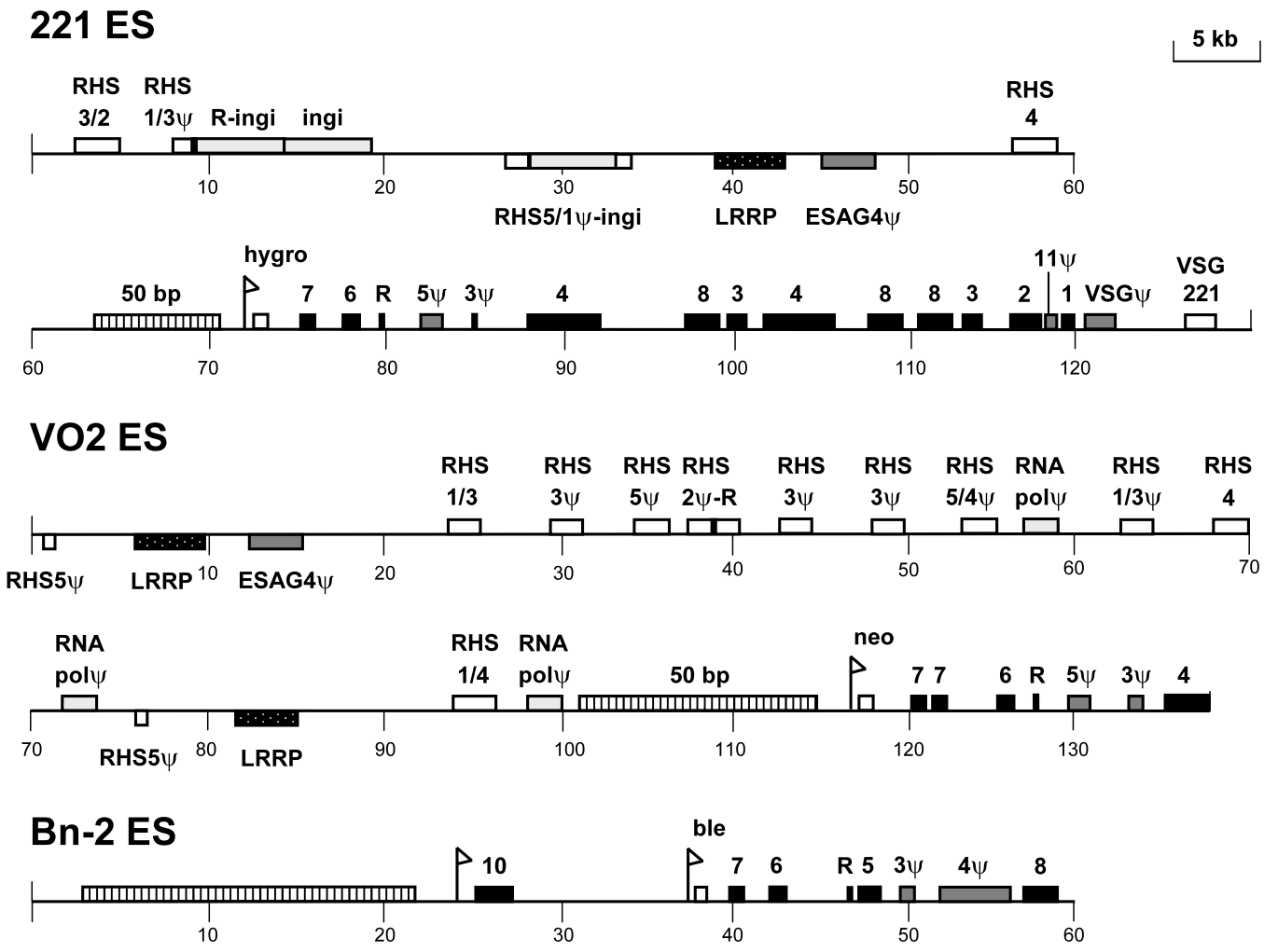

Fig. 1. Schematic of the 221 ES and flanking sequences, and BACs containing part of the VO2 ES and Bn-2 ES plus upstream sequences. The ES promoters are indicated with white flags, and ORFs with boxes: *ESAG*s with black boxes, pseudogenes (ψ) with dark grey boxes, hygromycin (hygro), neomycin (neo) and bleomycin (ble) resistance genes are indicated with white boxes. Directly upstream of ESs are arrays of 50-bp repeats (50-bp) indicated with vertically striped boxes. Upstream of ESs are various repetitive elements including *RHS* genes and pseudogenes (white boxes) which are numbered according to [27]. *Ingi* repetitive elements are indicated with light grey boxes. *RIME* elements are indicated with (R) and a black box. Members of a novel *LRRP* gene family frequently found upstream of VSG ESs are indicated with dark stippled boxes. Some *RHS* pseudogenes are inactivated by *ingi* (*RHS-ingi*) or *RIME* (*RHS-RIME*) retroelement insertion and and some *RHS* (pseudo)genes are chimaeras between two *RHS* belonging to different subfamilies (*RHS 1/3, 1/4, 3/2, 5/1* and *5/4*). ORFs encoded on the sense strand are indicated above the line, and ORFs encoded on the antisense strand are indicated underneath the line. The schematic of the 221 ES sequence was drawn from the sequence of the 221 BAC (which extends to the EcoRI site immediately upstream of the 221 VSG gene) and [68].

gets its highest score with ESAG8, due to the leucine-rich repeats [45,46], but LRRP proteins lack the RING Zn-finger motif and the nucleolar localisation domains of ESAG8 [18]. LRR repeats can be very degenerate, making them difficult to distinguish [47].

In pulsed field gel separations of *T. brucei* 427 chromosomal DNA, *LRRP* genes appear to be present on most ES-containing chromosomes (Fig. 2). All known *T. brucei* ESs contain *ESAG*s 6 and 7, which do not appear to be found outside ESs [40] (results not shown). Most chromosomes hybridising with a probe for the *ESAG*s 6 and 7 appear to hybridise with an *LRRP* probe, but *LRRP* also hybridises with chromosomes that do not contain an ES.

We compared the 221 and VO2 ES sequences with each other using Artemis Comparison Tool (ACT) after masking the most repetitive sequences: *RHS* (pseudo)genes, *ingi* and *RIME* retroelements and 50-bp repeats (Fig. 3). Sequence similarities are shown in red, with *LRRP* similarities highlighted in yellow. The VO2 ES appears to have undergone large duplications in the area upstream of the 50-bp repeats, including *LRRP* duplication. In addition, a DNA segment including *LRRP* and an *ESAG4* pseudogene is conserved.

We next compared both the 221 and VO2 ES sequences with the left telomere of *T. brucei* 927 chromosome I (EMBL accession number AL359782, manuscript in preparation). The 'left' telomere of *T.*

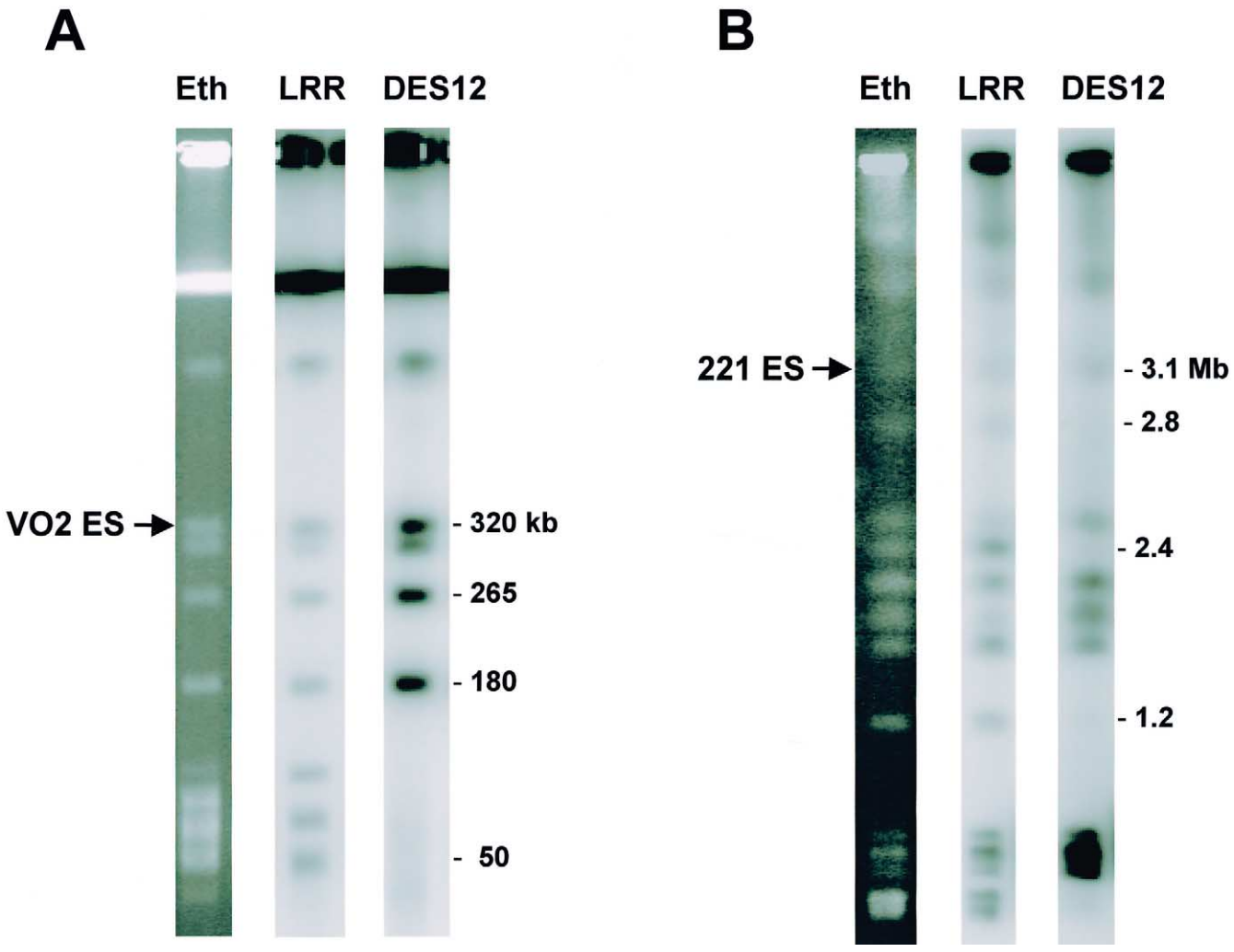

Fig. 2. The *LRRP* gene family is highly repetitive in the *T. brucei* genome. Panel A shows a CHEF pulsed field gel separation of *T. brucei* 427 chromosomes ranging from 50 to 500 kb. The panel with the ethidium bromide stained gel (Eth) has the 330 kb intermediate chromosome containing the VO2 VSG ES indicated with an arrow. A Southern blot of the gel was hybridised with a *LRRP* probe (labelled LRR) or a probe for *ESAG6* and *7* (DES12) to show the distribution of VSG ESs. CHEF separation of *T. brucei* 427 chromosomes ranging from 1 to 4 Mb is indicated in Panel B. The 3.1 Mb chromosome containing the 221 ES is indicated with an arrow. The blots were washed at high stringency (0.1 × SSC).

brucei 927 chromosome I contains a truncated and presumably non-functional ES. Two of our ES-containing BACs, particularly that containing the VO2 ES, showed considerable similarity with the left telomere of chromosome I. Three *LRRP* copies were found upstream of the 50-bp repeats in this telomere, but not elsewhere in this one megabase chromosome. A DNA segment containing *LRRP-ESAG4* pseudogene sequences was also present in this chromosome I telomere, despite the fact that this chromosome was derived from the *T. brucei* 927 rather than 427 strain. This conservation is striking, as *T. brucei* 927 and 427 strains are not obviously closely related based on their different karyotypes [20]. As there is no obvious reason why this *LRRP-ESAG4* pseudogene segment should be conserved, this could indicate that multiple ESs could have arisen from a single precursor, relatively recently.

4. Discussion

The AnTat 1.3A ES has long been considered the 'canonical' *VSG* ES [42], and appears to be highly similar to the AnTat 11.17 ES [48]. However, *T. brucei* ESs are polymorphic in size and structure, and can range from the truncated ETat1.2CR ES [5] to the extensive 221 ES described here, with its *ESAG* duplications and triplications [6]. An overview of all currently sequenced *T. brucei* ESs (Fig. 4) shows considerable diversity in the number and order of *ESAG*s. Only *ESAG*s 6 and 7 appear to have functional copies in every ES, and are presumably essential. This is difficult to test, because in addition to transcription from the active ES, there is low-level transcription of *ESAG*s 6 and 7 from many 'silent' ESs [40,41].

If only *ESAG*s 6 and 7 are essential in the bloodstream-form ES, why do most ESs contain additional

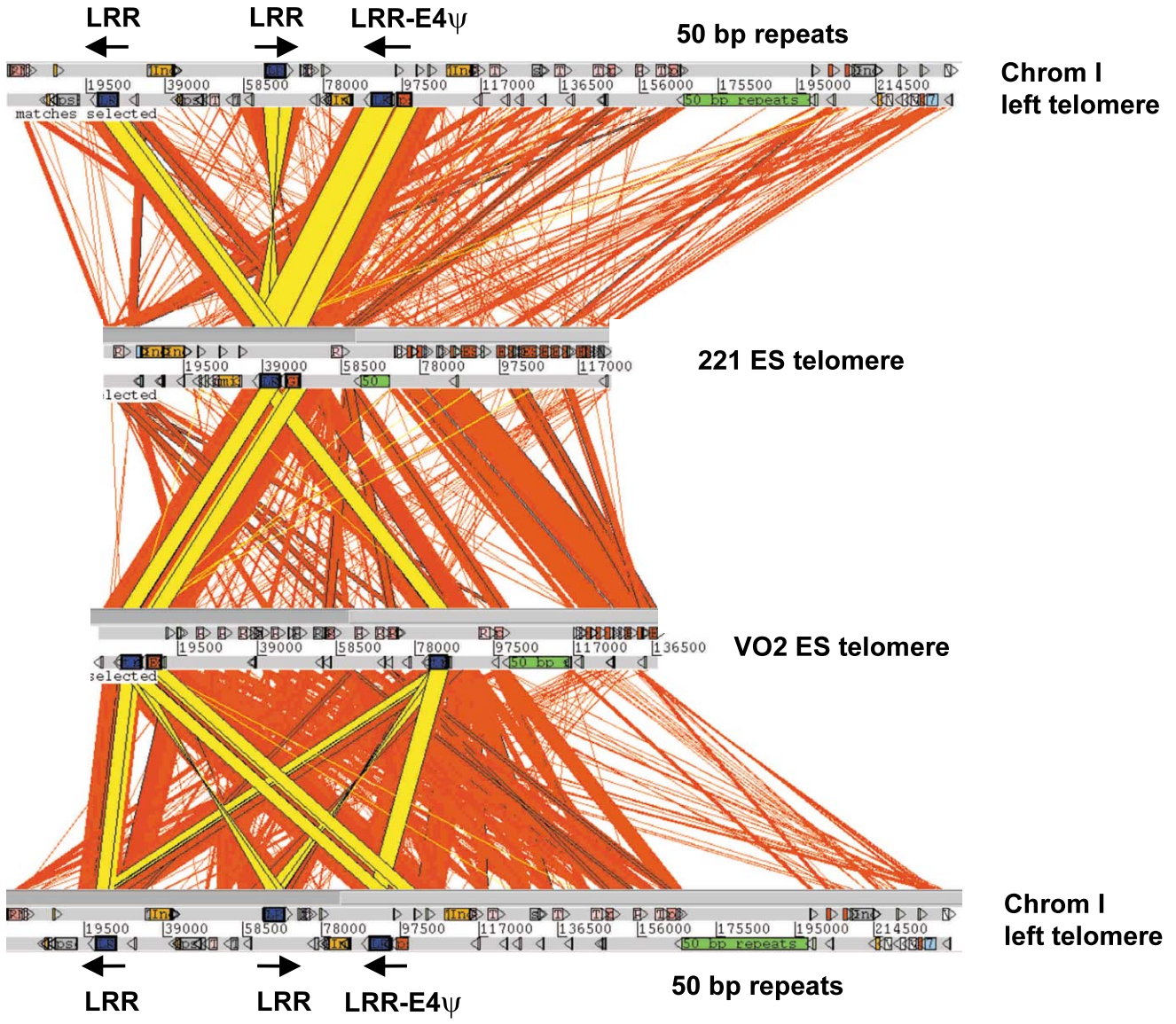

Fig. 3. Similarity in the genomic architecture of the VO2 and 221 VSG ES telomeres and the left telomere of *T. brucei* 927 chromosome I shown using Artemis Comparison Tool (ACT) in a 3-way comparison. The telomeres are arbitrarily depicted with the chromosome end on the right hand side of the figure. Comparison was performed after masking for some repetitive sequences: *RHS* coding regions, *ingi* and *RIME* elements, and the 50-bp repeats. The *LRRP* genes (LRR) and *ESAG4* pseudogene (E4ψ) are indicated above the sequence of chromosome I with arrows indicating orientation. Similarities are shown with red diagonal lines. Similarities in structure between the *LRRP* genes and *ESAG4* pseudogenes located upstream of ESs are highlighted in yellow. Sequence inversions are indicated with twisted lines. The *LRRP* genes are indicated with blue boxes, *ESAG* genes and pseudogenes with red boxes and the 50 bp repeats with green boxes.

genes? If the theory that *ESAG*s play a role in host adaptation is correct [13], other *ESAG*s could play an essential role in a host environment that has not been tested in the laboratory. Alternatively, ES-derived *ESAG*s could be non-essential but play a modulating role. Although several *ESAG*s are members of large gene families, with many copies outside ESs, genes present in ESs have different transcriptional properties to those in chromosome-internal locations. *T. brucei* chromosomes appear to be organised into large polycistronic units transcribed by RNA polymerase II, as is also the case in *Leishmania* [49,50]. ESs appear to be transcribed at a much higher rate, by RNA polymerase I [51–54]. Having some members of an *ESAG* family in an ES could allow the trypanosome to obtain higher expression of these variants.

The VO2 BAC contains approximately half of a 330-kb intermediate chromosome. We analysed all BACs hybridising with the neomycin resistance gene located in the VO2 ES, but did not find any BAC clones that extended much further upstream of the N19B2 VO2 clone sequenced here. As we were unable to identify unique sequences upstream of the 50-bp repeats of the VO2 ES (results not shown), we were unable to isolate

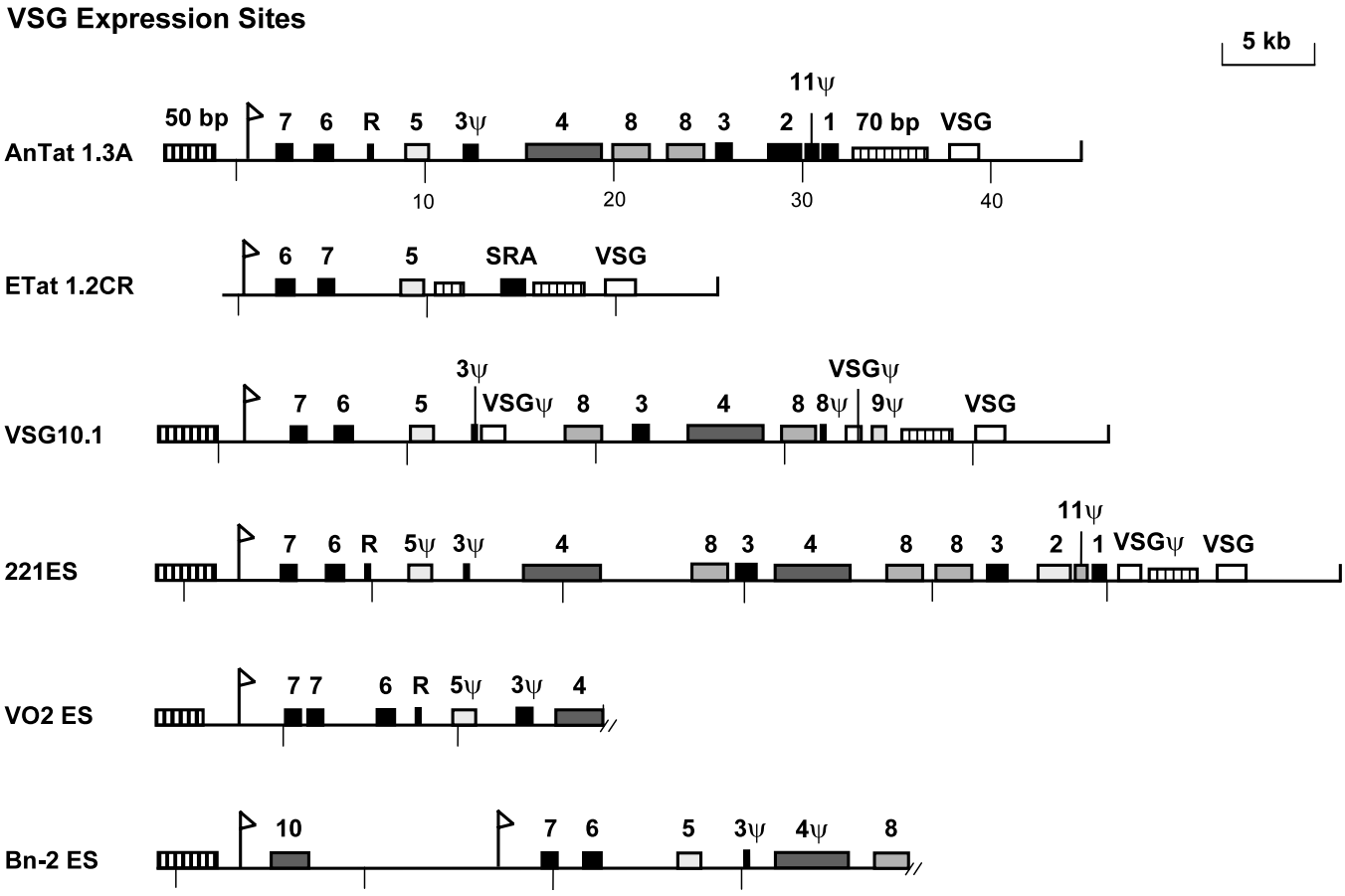

Fig. 4. Overview of sequenced *T. brucei* ESs. The promoters are indicated with flags, and *ESAG*s with numbered boxes. The *VSG* is indicated with a white box, and *SRA* [5] with a black box. Characteristic 50-bp and 70-bp repeat arrays (not drawn to scale) are indicated with striped boxes. Putative pseudogenes are indicated with ψ. The AnTat 1.3A ES was drawn from [69–71] and sequence accession numbers L20156 and AJ239060. The Etat1.2CR ES was drawn from [5] and accession number AJ010094. The VSG 10.1 ES was drawn from [44] and accession number AC087700. The 221 ES was drawn from sequence presented in this manuscript and [68]. The partial VO2 and Bn-2 ESs were drawn from the sequences presented here.

BAC clones spanning the other half of the VO2 chromosome. Nothing is known about this presumably aneuploid size class of chromosomes, except that they frequently contain telomeric ESs [55], and none hybridised exclusively with a set of 401 unique cDNA probes [33]. The VO2 BAC is similar to the left telomere of *T. brucei* 927 chromosome I, which contains a truncated bloodstream-form ES. This similar structure could indicate that intermediate chromosomes have originated from breakage of megabase chromosomes. Chromosomal breakage resulting in deletion of hundreds of kilobases from the chromosome VIa 221 ES has frequently been seen during VSG switching [21,56].

It remains to be determined how intermediate chromosomes segregate. The VO2 BAC does not contain the 177-bp repeat arrays characteristic of mini-chromosomes [57]. These repeats could be involved in segregation of the approximately one hundred minichromosomes, which segregate differently to the megabase chromosomes [58,59]. Nothing is known about centromeres in *T. brucei*, though it seems likely that the sequences functioning as centromeres will be different for each chromosomal size class.

Bloodstream-form ESs appear to be invariably flanked upstream of the 50-bp repeat arrays by tens to hundreds of kilobases of repetitive sequences including *ingi* and *RIME* retroelements and *RHS* (pseudo)genes. This non-random distribution of repetitive elements has been seen in other organisms. For example, repetitive elements are preferentially located in islands on each of five *Arabidopsis* chromosomes reviewed in [60]. In *Saccharomyces cerevisiae* Ty5 retroposons appear to preferentially target silent chromatin [61]. It is not clear why arrays of repetitive elements are a common feature upstream of *T. brucei* bloodstream-form ESs. Although this is presumably a property of 'selfish' DNA elements, these extensive expanses of 'junk' DNA could serve the purpose of isolating chromosome-internal housekeeping genes from turbulent chromosome ends. ESs are subject to powerful silencing forces and the potentially destruc-

tive effects of the DNA rearrangements associated with *VSG* switching.

LRRP genes are present upstream of the 221 and VO2 ESs, and the truncated ES on the left telomere of *T. brucei* 927 chromosome I. In addition, this gene family appears to be present on most chromosomes containing ES sequences. This could indicate that *LRRP* genes are frequently associated with ESs. Unexpectedly, all three of these ESs contained a DNA segment including an *LRRP* and an *ESAG4* pseudogene. This conservation of a non-functional pseudogene is particularly striking, as this is found across unrelated strains: *T. brucei* 927 strain (chromosome I) and the *T. brucei* 427 strain (221 and VO2 ESs). One possibility is that multiple ESs originated relatively recently from a single precursor ES, preserving non-functional pseudogenes along with functional *ESAG*s. Other evidence for this idea is provided by the *ESAG*3 pseudogene downstream of *ESAG5*, which is found in all *T. brucei* ESs with the exception of the truncated ETat1.2CR ES. Alternatively, extensive gene conversion in *T. brucei* could have resulted in homogenisation of ES sequences. There is extensive telomere conversion in *T. brucei* [62,63], which could result in the amplification of non-functional pseudogenes. It is not known if extensive gene conversion also occurs in the repetitive regions upstream of ESs.

In conclusion, it appears that malaria parasites and the African trypanosomes have harnessed the ends of chromosomes, with their higher rates of recombination and their physical and transcriptional instability, to diversify gene families involved in phenotypic variation [64–67]. The advantages of diversity in the genes encoding the surface coat is obvious. The critical challenge will come in identifying the functional advantages of this diversity in the other genes present in the ES.

Acknowledgements

We are grateful to P. de Jong (Children's Hospital Oakland Research Institute) for constructing the BAC libraries used in this study. We thank Professor Keith Gull for stimulating discussions. This work was funded by the Wellcome Trust through its Beowulf genomics initiative (grant number 059213), a Wellcome Senior Fellowship in the Basic Biomedical Sciences to G.R., a grant to P.B. from the Netherlands Foundation for Chemical Research (CW) with financial support of the Netherlands Organisation for Scientific Research (NWO), and the National Institutes of Health (grant number AI21729 to G.A.M.C.). K.S. is a Wellcome Prize student.

References

[1] Borst P., Ulbert S.. Control of VSG gene expression sites. Mol Biochem Parasitol 2001;114:17–27.
[2] Pays E., Lips S., Nolan D., Vanhamme L., Perez-Morga D.. The VSG expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. Mol Biochem Parasitol 2001;114:1–16.
[3] Barry J.D., McCulloch R.. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. Adv Parasitol 2001;49:1–70.
[4] Vanhamme L., Pays E., McCulloch R., Barry J.D.. An update on antigenic variation in African trypanosomes. Trends Parasitol 2001;17:338–43.
[5] Xong H.V., Vanhamme L., Chamekh M., Chimfwembe C.E., Van Den Abbeele J., Pays A., Van Meirvenne N., Hamers R., De Baetselier P., Pays E.. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. Cell 1998;95:839–46.
[6] Kooter J.M., van der Spek H.J., Wagter R., d'Oliveira C.E., van der Hoeven F., Johnson P.J., Borst P.. The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei*. Cell 1987;51:261–72.
[7] Johnson P.J., Kooter J.M., Borst P.. Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell 1987;51:273–81.
[8] Vanhamme L., Lecordier L., Pays E.. Control and function of the bloodstream variant surface glycoprotein expression sites in *Trypanosoma brucei*. Int J Parasitol 2001;31:523–31.
[9] Schell D., Evers R., Preis D., Ziegelbauer K., Kiefer H., Lottspeich F., Cornelissen A.W., Overath P.. A transferrin-binding protein of *Trypanosoma brucei* is encoded by one of the genes in the variant surface glycoprotein gene expression site. Embo J 1991;10:1061–6published erratum appears in EMBO J 1993Jul;12(7):2990).
[10] Ligtenberg M.J., Bitter W., Kieft R., Steverding D., Janssen H., Calafat J., Borst P.. Reconstitution of a surface transferrin binding complex in insect form *Trypanosoma brucei*. EMBO J 1994;13:2565–73.
[11] Ross D.T., Raibaud A., Florent I.C., Sather S., Gross M.K., Storm D.R., Eisen H.. The trypanosome VSG expression site encodes adenylate cyclase and a leucine-rich putative regulatory gene. EMBO J 1991;10:2047–53.
[12] Lemley C., Yan S., Dole V.S., Madhubala R., Cunningham M.L., Beverley S.M., Myler P.J., Stuart K.D.. The Leishmania donovani LD1 locus gene ORFG encodes a biopterin transporter (BT1). Mol Biochem Parasitol 1999;104:93–105.
[13] Bitter W., Gerrits H., Kieft R., Borst P.. The role of transferrin-receptor variation in the host range of *Trypanosoma brucei*. Nature 1998;391:499–502.
[14] Chaves I., Rudenko G., Dirks-Mulder A., Cross M., Borst P.. Control of variant surface glycoprotein gene-expression sites in *Trypanosoma brucei*. EMBO J 1999;18:4846–55.
[15] Carruthers V.B., Navarro M., Cross G.A.. Targeted disruption of expression site-associated gene-1 in bloodstream-form *Trypanosoma brucei*. Mol Biochem Parasitol 1996;81:65–79.
[16] Alexandre S., Paindavoine P., Hanocq-Quertier J., Paturiaux-Hanocq F., Tebabi P., Pays E.. Families of adenylate cyclase genes in *Trypanosoma brucei*. Mol Biochem Parasitol 1996;77:173–82.
[17] Morgan R.W., El-Sayed N.M., Kepa J.K., Pedram M., Donelson J.E.. Differential expression of the expression site-associated gene I family in African trypanosomes. J Biol Chem 1996;271:9771–7.
[18] Hoek M., Cross G.A.. Expression-site-associated-gene-8 (ESAG8) is not required for regulation of the VSG expression

site in *Trypanosoma brucei*. Mol Biochem Parasitol 2001;117:211–5.

[19] Bernards A., de Lange T., Michels P.A., Liu A.Y., Huisman M.J., Borst P.. Two modes of activation of a single surface antigen gene of *Trypanosoma brucei*. Cell 1984;36:163–70.

[20] Melville S.E., Leech V., Navarro M., Cross G.A.. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* stock 427. Mol Biochem Parasitol 2000;111:261–73.

[21] Rudenko G., Chaves I., Dirks-Mulder A., Borst P.. Selection for activation of a new variant surface glycoprotein gene expression site in *Trypanosoma brucei* can result in deletion of the old one. Mol Biochem Parasitol 1998;95:97–109.

[22] Frengen E., Weichenhan D., Zhao B., Osoegawa K., van Geel M., de Jong P.J.. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics 1999;58:250–3.

[23] Zeng C., Kouprina N., Zhu B., Cairo A., Hoek M., Cross G., Osoegawa K., Larionov V., de Jong P.. Large-insert bac/yac libraries for selective re-isolation of genomic regions by homologous recombination in yeast. Genomics 2001;77:27–34.

[24] Harris D.E., Murphy L.. Sequencing bacterial artificial chromosomes. In: Starkey M.P., Elaswarapu R., editors. Genomics protocols. Totowa, NJ: Humana Press, 2001:217–34.

[25] Wilson R., Ainscough R., Anderson K., et al. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 1994;368:32–8.

[26] Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., Rajandream M.A., Barrell B.. Artemis: sequence visualisation and annotation. Bioinformatics 2000;16:944–5.

[27] Bringaud F., Biteau N., Melville S.E., Hez S., El-Sayed N., Berriman M., Hall N., Donelson J.E., Baltz T.. A new, expressed multigene family containing a hot spot of insertion for retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. Eukaryotic Cell 2002;1:137–51.

[28] Kimmel B.E., ole-MoiYoi O.K., Young J.R.. Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol Cell Biol 1987;7:1465–75.

[29] Murphy N.B., Pays A., Tebabi P., Coquelet H., Guyaux M., Steinert M., Pays E.. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. J Mol Biol 1987;195:855–71.

[30] Hasan G., Turner M.J., Cordingley J.S.. Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. Cell 1984;37:333–41.

[31] Zomerdijk J.C., Ouellette M., ten Asbroek A.L., Kieft R., Bommer A.M., Clayton C.E., Borst P.. The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J 1990;9:2791–801.

[32] Sambrook J., Russell D.W.. Molecular Cloning: a Laboratory Manual, 3rd ed.. New York, USA: Cold Spring Harbour Press, 2001.

[33] Melville S.E., Leech V., Gerrard C.S., Tait A., Blackwell J.M.. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. Mol Biochem Parasitol 1998;94:155–73.

[34] Rudenko G., Blundell P.A., Taylor M.C., Kieft R., Borst P.. VSG gene expression site control in insect form *Trypanosoma brucei*. EMBO J 1994;13:5470–82.

[35] Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y., Simon M.. Cloning and stable maintenance of 300-kb-pair fragments of human DNA in *Escherichia coli* using an *F*-factor-based vector. Proc Natl Acad Sci USA 1992;89:8794–7.

[36] Osoegawa K., Woon P.Y., Zhao B., Frengen E., Tateno M., Catanese J.J., de Jong P.J.. An improved approach for construction of bacterial artificial chromosome libraries. Genomics 1998;52:1–8.

[37] Gottesdiener K., Chung H.M., Brown S.D., Lee M.G.S., van der Ploeg L.H.T.. Characterization of VSG gene expression site promoters and promoter-associated DNA rearrangement events. Mol Cell Biol 1991;11:2467–80.

[38] Gottesdiener K., Goriparthi L., Masucci J.P., van der Ploeg L.H.T.. A proposed mechanism for promoter-associated DNA rearrangement events at a variant surface glycoprotein gene expression site. Mol Cell Biol 1992;12:4784–95.

[39] Levy D.D., Cebula T.A.. Fidelity of replication of repetitive DNA in mutS and repair proficient *Escherichia coli*. Mutat Res 2001;474:1–14.

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

[41] Vanhamme L., Poelvoorde P., Pays A., Tebabi P., Van Xong H., Pays E.. Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. Mol Microbiol 2000;36:328–40.

[42] Pays E., Tebabi P., Coquelet H., Revelard P., Salmon D., Steinert M.. The genes and transcripts of an antigen gene expression site from *T. brucei*. Cell 1989;57:835–45.

[43] Melville S.E., Gerrard C.S., Blackwell J.M.. Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. Chromosome Res 1999;7:191–203.

[44] LaCount D.J., El-Sayed N.M., Kaul S., Wanless D., Turner C.M., Donelson J.E.. Analysis of a donor gene region for a variant surface glycoprotein and its expression site in African trypanosomes. Nucleic Acids Res 2001;29:2012–9.

[45] Revelard P., Lips S., Pays E.. 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 1990;18:7299–303.

[46] Smiley B.L., Stadnyk A.W., Myler P.J., Stuart K.. 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 1990;10:6436–44.

[47] Kobe B., Deisenhofer J.. The leucine-rich repeat: a versatile binding motif. Trends Biochem Sci 1994;19:415–21.

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

[49] Marchetti M.A., Tschudi C., Silva E., Ullu E.. Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersionof RNA polymerase II- and III-transcribed genes. Nucleic Acids Res 1998;26:3591–8.

[50] Myler P.J., Stuart K.D.. Recent developments from the *Leishmania* genome project. Curr Opin Microbiol 2000;3:412–6.

[51] Kooter J.M., Borst P.. Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucleic Acids Res 1984;12:9457–72.

[52] Laufer G., Schaaf G., Bollgonn S., Gunzl A.. In vitro analysis of alpha-amanitin-resistant transcription from the rRNA, procyclic acidic repetitive protein, and variant surface glycoprotein gene promoters in *Trypanosoma brucei*. Mol Cell Biol 1999;19:5466–73.

[53] Laufer G., Gunzl A.. In-vitro competition analysis of procyclin gene and variant surface glycoprotein gene expression site transcription in *Trypanosoma brucei*. Mol Biochem Parasitol 2001;113:55–65.

[54] Navarro M., Gull K.. A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. Nature 2001;414:759–63.

[55] Van der Ploeg L.H., Cornelissen A.W., Michels P.A., Borst P.. Chromosome rearrangements in *Trypanosoma brucei*. Cell 1984;39:213–21.
[56] Cross M., Taylor M.C., Borst P.. Frequent loss of the active site during variant surface glycoprotein expression site switching in vitro in *Trypanosoma brucei*. Mol Cell Biol 1998;18:198–205.
[57] Sloof P., Menke H.H., Caspers M.P., Borst P.. Size fractionation of *Trypanosoma brucei* DNA: localisation of the 177-bp repeat satellite DNA and a variant surface glycoprotein gene in a minichromosomal DNA fraction. Nucleic Acids Res 1983;11:3889–901.
[58] Ersfeld K., Gull K.. Partitioning of large and minichromosomes in *Trypanosoma brucei*. Science 1997;276:611–4.
[59] Gull K., Alsford S., Ersfeld K.. Segregation of minichromosomes in trypanosomes: implications for mitotic mechanisms. Trends Microbiol 1998;6:319–23.
[60] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature, 2000;408:796–815.
[61] Zou S., Voytas D.F.. Silent chromatin determines target preference of the *Saccharomyces* retrotransposon Ty5. Proc Natl Acad Sci USA 1997;94:7412–6.
[62] McCulloch R., Rudenko G., Borst P.. Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-bp repeat sequences. Mol Cell Biol 1997;17:833–43.
[63] Robinson N.P., Burman N., Melville S.E., Barry J.D.. Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. Mol Cell Biol 1999;19:5839–46.
[64] Rudenko G.. Genes involved in phenotypic and antigenic variation in African trypanosomes and malaria. Curr Opin Microbiol 1999;2:651–6.
[65] Scherf A., Figueiredo L.M., Freitas-Junior L.H.. *Plasmodium* telomeres: a pathogen's perspective. Curr Opin Microbiol 2001;4:409–14.
[66] Freitas-Junior L.H., Bottius E., Pirrit L.A., Deitsch K.W., Scheidig C., Guinet F., Nehrbass U., Wellems T.E., Scherf A.. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature 2000;407:1018–22.
[67] Gottschling D.E., Aparicio O.M., Billington B.L., Zakian V.A.. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. Cell 1990;63:751–62.
[68] Bernards A., Kooter J.M., Borst P.. Structure and transcription of a telomeric surface antigen gene of *Trypanosoma brucei*. Mol Cell Biol 1985;5:545–53.
[69] Lips S., Revelard P., Pays E.. 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 1993;62:135–7.
[70] Alexandre S., Guyaux M., Murphy N.B., Coquelet H., Pays A., Steinert M., Pays E.. Putative genes of a variant-specific antigen gene transcription unit in *Trypanosoma brucei*. Mol Cell Biol 1988;8:2367–78.
[71] Redpath M.B., Windle H., Nolan D., Pays E., Voorheis H.P., Carrington M.. ESAG11, a new VSG expression site-associated gene from *Trypanosoma brucei*. Mol Biochem Parasitol 2000;111:223–8.