Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species

Andrew P. Jackson1,2, Andrew Berry1, Martin Aslett1, Harriet C. Allison2, Peter Burton3, Jana Vavrova-Anderson4, Robert Brown3, Hilary Browne1, Nicola Corton3, Heidi Hauser5, John Gamble6, Ruth Gilderthorp7, Lucio Marcello2, Jacqueline McQuillan3, Thomas D. Otto3, Michael A. Quail2, Mandy J. Sanders2, Andries van Tonder1, Michael L. Ginger4, Mark C. Field2, J. David Barry7, Christiane Hertz-Fowler8, and Matthew Berriman1

1Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; 2Department of Pathology, University of Cambridge, Cambridge CB2 1QH, United Kingdom; 3Wellcome Trust Centre for Molecular Parasitology, Institute of Infection, Immunity, and Inflammation, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow G12 8TA, United Kingdom; 4Division of Biomedical and Life Sciences, School of Health and Medicine, Lancaster University, Lancaster LA1 4YQ, United Kingdom; and 5Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, United Kingdom

Edited by Anthony A. James, University of California, Irvine, CA, and approved January 11, 2012 (received for review October 25, 2011)

Antigenic variation enables pathogens to avoid the host immune response by continual switching of surface proteins. The protozoan blood parasite Trypanosoma brucei causes human African trypanosomiasis (“sleeping sickness”) across sub-Saharan Africa and is a model system for antigenic variation, surviving by periodically replacing a monolayer of variant surface glycoproteins (VSG) that covers its cell surface. We compared the genome of Trypanosoma brucei with two closely related parasites Trypanosoma congolense and Trypanosoma vivax, to reveal how the variant antigen repertoire has evolved and how it might affect contemporary antigenic diversity. We reconstruct VSG diversification showing that Trypanosoma congolense uses variant antigens derived from multiple ancestral VSG lineages, whereas in Trypanosoma brucei VSG have recent origins, and ancestral gene lineages have been repeatedly co-opted to novel functions. These historical differences are reflected in fundamental differences between species in the scale and mechanism of recombination. Using phylogenetic incompatibility as a metric for genetic exchange, we show that the frequency of recombination is comparable between Trypanosoma congolense and Trypanosoma brucei but is much lower in Trypanosoma vivax. Furthermore, in showing that the C-terminal domain of Trypanosoma brucei VSG plays a crucial role in facilitating exchange, we reveal substantial species differences in the mechanism of VSG diversification. Our results demonstrate how past VSG evolution indirectly determines the ability of contemporary parasites to generate novel variant antigens through recombination and suggest that the current model for antigenic variation in Trypanosoma brucei is only one means by which these parasites maintain chronic infections.

Antigenic variation enables pathogens to evade immune responses by continual switching of surface proteins (1, 2). The African trypanosomes (Trypanosoma spp.) are vector-borne protozoan blood parasites that survive in their hosts by antigenic variation, periodically replacing a monolayer of variant surface glycoproteins (VSG) (3) that shield the cell surface from immune effectors (4, 5). Trypanosoma brucei is the cause of human African trypanosomiasis (or “sleeping sickness”), and the mechanisms for expression and dynamic replacement of VSG in these species are a model system for antigenic variation (4) as well as a classic example of adaptive evolution at the host–pathogen interface. Two related veterinary parasites, Trypanosoma congolense and Trypanosoma vivax, also use antigenic variation to cause devastating diseases in domesticated animals. Through their detrimental effects on livestock productivity, these species arguably present greater threats to socioeconomic well-being than T. brucei does in the agrarian societies in which they are endemic. Our understanding of how antigenic diversity is organized in T. brucei was greatly improved by the T. brucei 927 reference genome sequence (6). In this paper, we present draft genome sequences for T. congolense and T. vivax; we define their global VSG repertoires in a three-way comparative analysis with T. brucei, revealing how antigenic diversity evolved in trypanosome genomes past and present.

The T. brucei genome includes many hundreds of VSG that encode a transcriptionally silent reservoir of variant antigens (7) and each cell expresses just a single gene from a specialized telomeric expression site at any time (4, 5). The parasite population collectively express multiple VSG; when the host becomes immune to the prevailing VSG, clones expressing alternative copies proliferate in a frequency-dependent manner, maintaining the infection and resulting in characteristic “waves of parasitaemia.” To survive long-term, T. brucei must generate novel VSG sequences through recombination; mechanisms may include domain shuffling (7) and gene conversion among silent, subtelomeric gene copies or possibly in situ within the expression site (8). Functional variant antigens in T. brucei consist of a- and b-type VSG (hereafter a-VSG and b-VSG), which share the cytotoxic-rich C-terminal domain (CTD) but are otherwise distantly related (9–11). Although VSG are known to occur in T. congolense and T. vivax (12–16), the repertoire of variant antigens in these species is uncharacterized. Consequently, the evolutionary diversification of the VSG gene family has not been examined, although it has been suggested that VSG are a source of novel genes. Two gene families, the expression site-associated genes (ESAG6/7) encoding transferrin receptor (TFR) and the VSG-related (VR) genes, are thought to have evolved from a-VSG (17, 18) and b-VSG (8, 11), respectively.

The antigenic variation phenotype is observed in all African trypanosomes, and it is assumed that this reflects a common physiological model, which has been defined in T. brucei. The aim of this study is to identify the evolutionary processes that have created contemporary VSG diversity and reveal any significant differences in how trypanosome species generate variant antigens. Despite their shared phenotype, our results show that species differ


The authors declare no conflict of interest.

This article is anPNAS Direct Submission. Data deposition: The draft genome sequences reported in this paper have been deposited in the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database [accession nos. HE573134–HE573524 and CAE01000001–CAE01000087 (Trypanosoma brucei) and HE573017–HE573027 and CAE01000001–CAE01000087 (Trypanosoma vivax)]. The data can be examined via GeneDB (http://www.genedb.org) and TriTrypDB (http://tritypdb.org). T. vivax transcriptome data have been submitted to the European Bioinformatics Institute Array Express Archive (accession no. E-MTAB-475). Sequence alignments and phylogenetic trees comprising the cell-surface phylome are contained in GeneDB (http://www.genedb.org/Page/trypansome_surface_phyllome).

1To whom correspondence should be addressed. E-mail: andrew.jackson@sanger.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1117313109/-DCSupplemental.
in the organization of antigenic diversity at the genome level, and they provide a basis to better understand disease progression, pathology, and host range in all African trypanosomes.

Results

VSG Gene Repertoires of T. congolense and T. vivax. We have produced high-quality draft genome sequences for T. congolense IL3000, a sister species of T. brucei, and T. vivax Y486, a third species that branches close to the root of the African trypanosome lineage (19). These genome sequences are described in Table S1 and are accessible through GeneDB (http://www.genedb.org) or TriTrypDB (http://tritrypdb.org). Comparative analysis including the existing T. brucei 927 genome sequence shows that the principal differences in genome content relate to cell-surface architecture (Tables S2–S4). To define VSG repertoires, gene sequences with predicted cell-surface roles were extracted from all three genomes and sorted with BLASTx, resulting in 81 gene families (Materials and Methods and Table S5). Phylogenies of these families were estimated that we collectively termed the “cell-surface phylome,” (http://www.genedb.org/Page/trypansomasurface_phrylome/).

The phylome contains VSG and related families already known in T. brucei but it also defines families that we believe encode the VSG repertoires of T. congolense and T. vivax.

The T. congolense VSG repertoire differs from that of T. brucei in three ways: First, there is no a-VSG subfamily of variant antigens; second, there are two b-VSG subfamilies, termed Fam13 (n = 302) and Fam16 (n = 512) by their phylome designations; and, third, unlike T. brucei VSG, which all share a relatively uniform CTD, T. congolense VSG have 15–20 different CTD types, each associated with a specific subset of Fam13 or Fam16 and none homologous to the T. brucei CTD. Hence, T. congolense b-VSG are more structurally heterogeneous than T. brucei b-VSG are. We know that both Fam13 and Fam16 contain functional variant antigens because each family encompasses both published T. congolense VSG (12–14) and VSG expressed sequence tags (15). Although there is no a-VSG variant antigen, there are homologs of the a-VSG–like TFR genes of T. brucei, i.e., ESA604 (Fam15; n = 43), as well as the procyclin-associated genes (PAG) (Fam14, n = 22), which also have an a-VSG–like structure (20).

VSG structural diversity is even greater in T. vivax. We have identified four VSG subfamilies (Fam23–26) that each possess definitive patterns of conserved cysteine residues (SI Results). Fam23 (n = 540) and Fam24 (n = 279) members possess sequence motifs homologous to a-VSG and b-VSG, respectively. Fam25 (n = 279) and Fam26 (n = 243) members are more closely related to b-VSG but with low (~20%) protein sequence similarity to known VSG (Fig. S1). These specific subfamilies may have evolved in T. vivax or may represent ancestral lineages not inherited by T. brucei and T. congolense. Transcriptomic data show that multiple members of all four families are transcribed in bloodstream-stage cells (Table S6). We find no orthologs to the TFR-like genes of T. brucei and T. congolense among T. vivax VSG-like genes or indeed the numerous T. vivax-specific gene families.

Amino acid sequence homology with T. brucei VSG alone does not guarantee that putative T. vivax VSG function as variant antigens. To date, only one T. vivax VSG [ILDat 2.1 (16)] has been characterized, albeit from a dissimilar strain, and it is most closely related to Fam26. Therefore, we identified an expressed VSG in the genome strain Y486 by MS analysis of a protein specific to a relaxed infection population, peptide fragments of which are 100% identical to a predicted protein in Fam23 (TY-Y486 0027060; Fig. S2). Therefore, at least one a-VSG–like (i.e., Fam23) gene in T. vivax encodes a functional variant antigen.

Phylogeny of VSG Diversification. In total, the three genome sequences yielded 1,083 a-VSG–like and 1,537 b-VSG–like full-length genes (Table S7). We estimated Bayesian and maximum likelihood phylogenies from amino acid sequence alignments (Figs. S3 and S4); however, given the large number of sequences and to enable global visualization, we also estimated a similarity network from pairwise maximum likelihood protein distances that delivered a clearer picture of relationships within the a-VSG and b-VSG lineages. The distance network includes examples of all VSG subfamilies and represents individual genes as spheres connected to others sharing identity above a threshold (Materials and Methods). The network and phylogenies are fully consistent. Fig. 1 shows the main family networks from three lineages, which show that amino acid residues crucial for transferrin binding (18) are conserved in both species (Fig. S6). Given the absence of this entire family from T. vivax, we conclude that, rather than being lost from that species, the transferrin-receptor genes evolved before the separation of T.
and T. congolense but after their split from T. vivax. Our conclusion does not preclude other T. vivax-specific proteins performing a transferrin-binding function in that species.

These results are summarized in a model of VSG evolution (Fig. 2). The ancestral African trypanosome possessed a-VSG and b-VSG type genes that probably formed multigene families or functioned as variant antigens. Both VSG types were inherited by T. vivax, which retains an a-VSG family that includes functional variant antigens. The T. brucei/T. congolense ancestor inherited both a-VSG and b-VSG lineages, and, at this point, one a-VSG gene was co-opted to a transferrin-binding role differentiated between insect and vertebrate life stages, founding a lineage that was inherited by both daughter species. Another a-VSG lineage retained its variant antigen function in T. brucei but was lost from T. congolense (SI Results). Of the ancestral b-VSG repertoire, two different lineages have been inherited by both species: The first has produced ESAG2 and Fam13 in T. brucei and T. congolense, respectively, whereas the second has produced b-VSG and VR in T. brucei and Fam16 in T. congolense. There is no step in this deduced scheme where a trypanosome lacks variant antigen. Clearly, these two species have adapted their common legacy differently. T. congolense VSG are drawn from multiple ancestral lineages, whereas T. brucei has relegated corresponding genes (VR, ESAG2, and perhaps Fam1) to novel roles and derives its variant antigens from single lineages derived after speciation. This difference in the phylogenetic diversity of VSG repertoires is important because it could affect the ability of the parasites to present novel antigens to their hosts and therefore to maintain infection.

Tree Shape and Distribution of VSG Sequence Variation. We examined the phylogenies of VSG subfamilies within species for evidence that their distinct evolutionary legacies have affected contemporary sequence evolution. Fig. 3 demonstrates how these trees have distinct topologies because of variation in the ratio of internal to terminal branches [described by “treeness” (24), T], which is low for T. brucei (T = 0.282 and 0.275), higher for T. congolense (T = 0.376 and 0.412), and highest for T. vivax (T = 0.681 and 0.763). T. congolense and T. vivax trees are more “tree-like” because they retain information about the past in basal nodes and internal branches, whereas the T. brucei tree consists mostly of long, terminal branches. Fig. 3 also compares the distribution of VSG sequence variation, showing that T. brucei distances have much narrower distributions than either T. congolense or T. vivax VSG do because both short, terminal branches and long, basal internodes are rare. Importantly, these patterns are genome-specific rather than lineage-specific effects, i.e., a-VSG and b-VSG in T. brucei display the same dynamic despite having greater identity with subfamilies in other species. They confirm that the mechanisms for antigenic variability vary between species now and likewise in the past. Recombination is a principal evolutionary pressure affecting T. brucei VSG (5, 8), and exchange of the unique VSG CTD is well recorded (7, 11). Recombination is also the mechanism through which VSG are transposed from subtelomeric loci into the telomeric expression site (4, 5, 8). T. brucei VSG phylogenies in Fig. 3 are consistent with frequent recombination, but the cladistic structure of T. congolense and T. vivax VSG phylogenies could only persist if recombination between clades is rare. Furthermore, the incidence of pseudogenes, which result, at least partly, from gene conversion between VSG genes (5), is much lower in T. congolense (where only 21.1% of Fam13 and 29.7% of Fam16 are predicted pseudogenes) and T. vivax (15.5% and 27.2% of Fam23 and Fam24, respectively) than in T. brucei (69.2% of a-VSG and 72.2% of b-VSG) (6). Therefore, we
suspected that recombination frequency might account for species differences in sequence variation.

Contribution of Recombination to Antigenic Diversity. We examined VSG alignments for evidence of recombination in the form of phylogenetic incompatibility (PI) (25, 26), taking random samples of each alignment set and observing the proportion showing significant PI \( (P_{pi}; \text{Table S8}) \). Fig. 4 shows that \( P_{pi} \) (solid colored lines) was greatest for \( T. brucei \) a- VSG (0.392) and b-VSG (0.450) as well as for \( T. congolense \) Fam16 (0.435) and lower for \( T. congolense \) Fam13 (0.125) and \( T. vivax \) Fam23 (0.138) and Fam24 (0.126). In all cases, observed \( P_{pi} \) was significantly greater than a null distribution (black lines), confirming that PI was not solely caused by other homoplastic effects, such as rate heterogeneity (Materials and Methods). Recombination frequency is known to be proportional to sequence identity (27), and when we increased sequence identity within alignments by sampling only within crown clades, \( P_{pi} \) increased significantly (dashed colored lines) for \( T. brucei \) a-VSG (0.681) and b-VSG (0.642) as well as for \( T. congolense \) Fam13 (0.466) and Fam16 (0.823) but not for \( T. vivax \). Finally, because the CTD is known to recombine in \( T. brucei \) (7, 11), we removed the CTD from \( T. brucei \) and \( T. congolense \) alignments, which resulted in a significant decrease in \( P_{pi} \) for \( T. brucei \) a-VSG (0.152, \( P < 0.0001 \)) and b-VSG (0.234, \( P < 0.0001 \)), but, in \( T. congolense \), \( P_{pi} \) actually increased.

Therefore, in \( T. brucei \) and \( T. congolense \), the evidence for recombination is greatest among closest-related VSG but was seldom observed in \( T. vivax \), even when sampling within clusters of highly related sequences. Although the frequency of PI is similar for \( T. brucei \) VSG and Fam16, if we compare \( P_{pi} \) in a global alignment of \( T. congolense \) b-VSG (0.163) with the corresponding value for \( T. brucei \) (0.450), it is clear that PI is prevalent throughout the \( T. brucei \) repertoire but only within \( T. congolense \) VSG clades. This is a sampling effect caused by their divergent evolutionary histories. Given that \( T. congolense \) VSG are phylogenetically diverse and have a wider distribution of sequence variation, they have proportionally more distant relationships and so more structural barriers to genetic exchange. In short, there are cohorts of \( T. congolense \) VSG that never recombine, as the topological differences in Fig. 3 suggest.

![Fig. 2. A model of VSG gene family evolution in African trypanosomes. This cartoon depicts the elaboration of VSG subfamilies in contemporary and ancestral genomes. Uncertain origins are indicated by dashed lines. An asterisk indicates that a subfamily includes a proven variant antigen, although other variant antigens may occur in unmarked subfamilies. The red arrow indicates that the CTD is uniquely shared between a-VSG and b-VSG in \( T. brucei \) and has been donated from one subfamily to the other in either direction.](image-url)
The past and present evolution of VSG can now be brought together. We have shown that the composition of contemporary VSG repertoires is determined by how each species has modified the common inheritance. T. vivax has the most structurally diverse repertoire, comprising a-VSG, b-VSG, and two additional types absent elsewhere; T. congolense combines multiple ancestral b-VSG lineages, each with a distinct CTD; and T. brucei a-VSG and b-VSG are recently derived single lineages with a common CTD. It is worth remembering that sequence mosaics generated in T. brucei infections are an additional source of diversity among expressed VSG sequences occurs in other species. Nevertheless, as a result of compositional differences, the scale of recombination varies between species, being more frequent among T. brucei and T. congolense VSG than in T. vivax, and more prevalent among T. brucei VSG than in T. congolense. Opportunities for allelic recombination among VSG may be affected by species differences in mating system; recent work indicates that sexual reproduction is frequent in T. brucei and T. congolense populations (28–29), whereas T. vivax population structure is consistent with clonal reproduction (30). However, the importance of allelic recombination to VSG diversity is debatable; because the subtelomeres of homologous chromosomes in T. brucei are frequently dissimilar, subtelomeric VSG are effectively hemizygous, and so most recombination between VSG is ectopic (11). This study shows that PI in T. brucei VSG is attributable in large part to the CTD promoting exchange throughout the repertoire, whereas the conservative CTDs of T. congolense VSG actually reduce the scale of PI and illustrate the lack of recombination between clades, emphasizing the role of gene structure in promoting ectopic recombination in mitosis rather than the frequency of allelic recombination during meiosis.

Differences in the role of the CTD indicate that, in addition to scale, the mechanism of recombination varies between species. The CTD is exchanged between T. brucei VSG but is not exposed to antibodies and therefore may not directly contribute to antigenic diversity (31). However, the CTD-encoding sequence may have a role in the duplicative transposition of VSG, which promotes the diversification of the silent archive (11) and is of paramount importance to antigenic variation because it is required to move VSG from silent, subtelomeric loci into the telomeric expression site (4, 8). It is thought that VSG transposition is facilitated by an upstream 70-bp repeat region and the conserved CTD (8), with the transposed region terminating around the 3′ end of the coding sequence (32). Hence, in 37 recent such duplicative events among silent VSG in the T. brucei 927 genome (11), 16 terminated within the CTD region. In showing that the majority of PI in T. brucei VSG alignments concerns the CTD, our data demonstrate that a major recombination breakpoint occurs between the N-terminal domains and CTDs, supporting a principal role for the CTD in promoting VSG switching. Immediately, we can see that this mechanism cannot operate in T. congolense, where the CTDs are heterogeneous and have no role in promoting exchange. Hence, we propose that the preeminence of the CTD in PI among T. brucei VSG reflects the frequent transposition of N-terminal domains,
and, through its solitary CTD type, which originated uniquely through horizontal transfer between VSG lineages, T. brucei may have evolved a distinct mechanism for the VSG transpositions that are key for diversification and antigen switching. Antigenic variation is central to the host–Trypanosoma relationship, intimately linked to the course and severity of disease, to parasite transmission and host range, and therefore to disease epidemiology. All African trypanosomes display antigenic variation, and, although the current T. brucei-based model might adequately describe the general phenomenon, this study shows that the genomic basis for antigenic variation has diverged among trypanosomes in a manner consistent with distinct mechanisms for generating antigenic variability. Consequently, we now have reason to expect substantial species differences beneath the general phenotype, a framework to dissect this variation, and so a basis for understanding how the enigmatic VSG connects with the wider disease.

Materials and Methods

Genome Sequencing and Annotation. T. congolense IL3000 and T. vivax Y486 genomes were capillary-sequenced with a whole-genome shotgun strategy as described previously (6).

Annotation of VSG Genes. T. congolense and T. vivax VSG were identified by BLASTp-based homology searches and hidden Markov models. The boundaries of all VSG ORFs were manually checked against global sequence alignments.

Comparison of Gene Content. OrthoMCL was used to examine putative gene gains and losses. All putative losses were confirmed by examining expected genomic position and by searching unassembled sequence reads for reciprocal sequence matches with tBLASTn/BLASTx.

T. vivax Transcriptome. T. vivax Y486 was grown from stabitate in BALB/c mice immunosuppressed with cyclophosphamide and was amplified at patient parasitaemia in three immunosuppressed mice, from which whole blood was collected. The blood was treated with the erythrocyte lysis buffer, and RNA was isolated from the pellet.

Analysis of Fam1 Gene Expression. To determine mRNA expression levels of Fam1 family members, quantitative real-time PCR (qRT-PCR) was carried out on total RNA.

Transfection and Fam1 Protein Localization. Ectopic expression of HA epitope-tagged Tb927.6.1310 at the N terminus (after the predicted signal peptide sequence) was carried out with constitutive and inducible expression vectors.

VSG Purification and Sequencing. T. vivax cell extracts were run in 1D SDS/PAGE, and three bands in the estimated size range were extracted from each, trypsinized, and subjected to liquid chromatography/tandem MS analysis.

Cell-Surface Phylome. Homologs to each T. brucei “surface” gene were identified among all T. brucei, T. congolense, T. vivax, and Trypanosoma cruzi predicted genes with wuBLAST. Gene family phylogenies were estimated with maximum likelihood and Bayesian methods.

Recombination Analysis. PI within VSG sequence alignments was used as a measure of recombination. The pairwise homology index (26) returns a single probability value for PI, which was applied to amino acid sequence alignments for seven VSG subfamilies (Table S8), subsampled 1,000 times. The proportion of subalignments with significant PI, termed P_{<0.05}, was compared between species. Additional information on methods is provided in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Phelxu Majiwa for sample preparation and Sara Melville for advice and advocacy during project planning. Barbara Marchetti (University of Glasgow) prepared RNA for transcriptomic analysis, Richard Burchmore (University of Glasgow) performed 2D electrophoresis, and Douglas Lamont (University of Dundee) performed liquid chromatography/tandem MS analysis. Arnab Pain and Chris Newbold (Wellcome Trust Sanger Institute) provided help and advice during manuscript preparation. We thank our colleagues in the sequence and informatics groups at the Wellcome Trust Sanger Institute. This work was funded by the Wellcome Trust Grants WT085775/Z/08/Z, 055558/Z/98/A, and 055558/Z/98/C. The Wellcome Trust Centre for Molecular Parasitology is supported by core funding from the Wellcome Trust (Grant 085349/Z/08/Z).