Molecular Phylogenetics and Evolution 67 (2013) 123–128

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids

Paul T. Manna [a], Steven Kelly [b], Mark C. Field [a,*]

[a] Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
[b] Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

ARTICLE INFO

Article history:
Received 30 October 2012
Revised 20 December 2012
Accepted 7 January 2013
Available online 19 January 2013

Keywords:
Endocytosis
Variant surface glycoprotein
Adaptin
Trypanosoma grayi

ABSTRACT

The kinetoplastids are an important group of protozoa from the Excavata supergroup, and cause numerous diseases with wide environmental, economic and ecological impact. *Trypanosoma brucei*, the causative agent of human African trypanosomiasis, expresses a dense variant surface glycoprotein (VSG) coat, facilitating immune evasion *via* rapid switching and antigenic variation. Coupled to VSG switching is efficient clathrin-mediated endocytosis (CME), which removes anti-VSG antibody from the parasite surface. While the precise molecular basis for an extreme CME flux is unknown, genes encoding the AP2 complex, central to CME in most organisms, are absent from *T. brucei*, suggesting a mechanistic divergence in trypanosome CME. Here we identify the AP complex gene cohorts of all available kinetoplastid genomes and a new *Trypanosoma grayi* genome. We find multiple secondary losses of AP complexes, but that loss of AP2 is restricted to *T. brucei* and closest relatives. Further, loss of AP2 correlates precisely with the presence of VSG genes, supporting a model whereby these two adaptations may function synergistically in immune evasion.

1. Introduction

The Kinetoplastida are a protozoan class belonging to the Excavata supergroup, encompassing numerous medically and agriculturally important pathogens, as well as free-living representatives that have huge ecological impact. Multiple features indicate these organisms as highly divergent, and potentially early branching in the evolution of the eukaryotes (Cavalier-Smith, 2010). Within the kinetoplastida, the order Trypanosomatida contains many pathogens, including *Trypanosoma brucei*, *T. cruzi* and *Leishmania* spp., the causative agents of African trypanosomiasis, Chagas' disease and leishmaniasis respectively. These parasites have evolved diverse immune evasion strategies; while *Leishmania* spp. and *T. cruzi* exploit intracellular lifestyles by invasion of host cells, *T. brucei* persists within the host bloodstream and lymphatic system, and is therefore continually exposed to both innate and adaptive immune mechanisms. The surface of mammalian infective *T. brucei* is dominated by approximately 2×10^7 molecules of a single GPI-anchored variant surface glycoprotein (VSG), shielding invariant surface antigens from antibody recognition (Borst and Cross, 1982; Mehlert et al., 2002). VSG is however highly immunogenic and evasion of the host immune response is accomplished *via* antigenic variation, the periodic switching of mono-allelic expression between VSG genes (Pays, 2005). Additionally, clearance of host antibodies from the parasite surface *via* extremely rapid clathrin-dependent endocytosis of antibody bound surface proteins and antibody degradation aids immune evasion whilst host antibody titre is low and may also protect invariant epitopes from immunological recognition (O'Beirne et al., 1998; Allen et al., 2003; Engstler et al., 2007; Field and Carrington, 2009).

Intracellular vesicular transport requires both structural coat proteins and cargo adaptors functioning at discrete intracellular locations for faithful targeting (Bonifacino and Lippincott-Schwartz, 2003). Arguably the best understood cargo adaptors are the assembly polypeptide or AP complexes. Five AP complexes are known, each comprising two large subunits, one medium and one small subunit, together referred to as adaptins (Boehm and Bonifacino, 2001; Hirst et al., 2011). At least two AP complexes are intimately involved in clathrin-mediated transport pathways. The AP1 complex, comprising large β1 and γ subunits together with a medium μ1 and small σ1 subunit, sorts cargo into clathrin coated vesicles travelling between the *trans*-Golgi network and endosomes. The AP2 complex, (β2, α, μ2 and σ2 subunits) drives cargo recruitment and clathrin-mediated endocytosis (CME) at the plasma membrane (Jackson et al., 2010). The AP3 complex (β3, δ, μ3, σ3), mediates delivery of proteins including the vesicle tether VAMP7 to late endosomes/lysosomes and lysosome related organelles (Martinez-Arca et al., 2003; Dell Angelica, 2009). Whilst the AP3 β3 subunit encodes a clathrin interaction motif it appears

* Corresponding author.
E-mail address: mcf34@cam.ac.uk (M.C. Field).

1055-7903
http://dx.doi.org/10.1016/j.ympev.2013.01.002

that clathrin is largely dispensable for the known functions of AP3 (Peden et al., 2002). Comparatively little is known about the functions of either AP4 (β4, ε, μ4, σ4) or AP5 (β5, ζ, μ5, σ5), although both are likely involved in post-Golgi transport and neither appear to use clathrin (Burgos et al., 2010; Hirst et al., 2011). Uniquely, and apparently paradoxically given its huge reliance upon clathrin-mediated endocytosis, *T. brucei* has dispensed with the AP2 complex entirely, the only reported incidence of this secondary loss (Berriman et al., 2005; Field et al., 2007; Nevin and Dacks, 2009). We previously suggested that an extreme flux of CME in *T. brucei* relative to other eukaryotes, plus the high density of VSG, drove adaptation to a non-specific, rapid, AP2-independent mode of CME in African trypanosomes (Field and Carrington, 2009). Previously however, the limited number of genomes available from related taxa has made it difficult to assess whether the loss of the AP2 complex is indeed a unique feature of the African trypanosomes, and thus potentially related to their unusual surface architecture, or a more widely distributed adaptation among trypanosomatids and their closer relatives.

To explore in more detail the evolutionary relationships between extracellular parasitism, the VSG coat and AP complex evolution, we have sequenced the genome of *Trypansoma grayi* strain ANR4, a parasite with a lifestyle similar to *T. brucei*, i.e. of African location, Tsetse fly transmitted and living extracellularly in its vertebrate (crocodile) host (Minter-Goedbloed et al., 1993; Hoare, 1929). Earlier phylogenetic studies suggest that *T. grayi* is more closely related to the intracellular South American *T. cruzi* (Stevens et al., 1999) (Stevens and Gibson, 1999) (Hamilton et al., 2007) and hence *T. grayi* is a critical taxon for understanding AP complex evolution in trypanosomes. We identified the adaptin and potential major surface protein genes from *T. grayi* and all currently available trypanosomatid and closely related bodonid genomes and examined the evolutionary history of these gene families. Our analysis shows that the loss of AP2 is restricted to the African trypanosomes and correlates with emergence of the VSG coat, lending support to the concept of co-evolution of these two features.

2. Materials and methods

2.1. Genome sequence databases

Publicly available genomes included in this analysis were Trypanosoma brucei brucei 927, *Trypanosoma brucei* gambiense, *Trypanosoma congolense*, *Trypanosoma vivax*, *Trypanosoma cruzi*, *Leishmania braziliensis*, *Leishmania infantum*, *Leishmania major*, *Leishmania mexicana*, *Leishmania tarentolae*, and were all searched at TriTrypDB, (www.tritrypdb.org), while *Bodo saltans* was from geneDB (www.genedb.org) and *Naegleria gruberi* from the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov) and the Joint Genome Initiative (www.jgi.doe.gov) respectively. *Trypanosoma carassii*, *Trypanosoma theileri*, *Trypanoplasma borrelli* were part of a transcriptome project by one of us (SK) and to be published in full later, and *Phytomonas serpens* data were kindly provided by Julius Lukes (Institute for Parasitology, České Budějovice). All sequences and alignments are available from the authors on request.

2.2. Genome sequencing

T. grayi strain ANR4 genomic DNA, a kind gift of Wendy Gibson (University of Bristol), was extracted from agarose plugs using standard phenol/chloroform methods. DNA was sequenced by 91 bp paired-end Illumina sequencing at the Beijing Genomics institute (www.genomics.cn/en/). Reads were clipped based on phred score >20 using the fastx program. Read errors and ambiguous bases were then corrected using the ALLPATHS (MacCallum et al., 2009) find read errors algorithm, with two cycles of read error correction and the default settings for *k*-mer size. Duplicate reads and reads with a post-clipped length of less than 20 nucleotides were discarded using custom Perl scripts. The clipped, corrected and filtered reads were then assembled using Velvet (Zerbino and Birney, 2008) and multiple *k*mer sizes (*k*mer = 31, 41, 51 and 61). The resulting contigs from all assemblies were then post-assembled using CAP3 (Huang and Madan, 1999) to yield a final genome assembly. *k*mer frequency analysis of filtered, clipped and corrected reads yielded an estimated genome size of 29,355,514 bases, at approximately 48*x* coverage, highly consistent with the haploid genome size estimates for other African trypanosomes.

2.3. Homology searches

Adaptin searches were performed with BLAST using *Homo sapiens* sequences as queries against genome sequence data from all available trypanosomatids and bodonids as well as the free living heterolobosid excavate *Naegleria gruberi* (Table S1). In cases where orthologs were not found, further searches were carried out using sequences from phylogenetically close organisms. All retrieved sequences were then verified by reciprocal BLAST analysis against the *H. sapiens* genome database. Major surface protein family searches were carried out with BLAST using multiple representatives of each protein family (Table S1) and genes were assigned as present based upon BLAST score ($e < 1 \times 10^{-5}$) and reciprocal BLAST retrieving members of the target gene family.

2.4. Phylogenetic reconstruction

Sequences were aligned using MUSCLE (Edgar, 2004) and edited manually to remove poorly conserved regions. For each protein family the optimal substitution model was assessed (ProtTest3 (Abascal et al., 2005)) and phylogenetic trees were generated using Bayesian (MrBayes, (Ronquist and Huelsenbeck, 2003)) and maximum likelihood (RaxML, (Stamatakis, 2006), and PhyML, (Guindon and Gascuel, 2003)) approaches. For maximum likelihood calculations, the best fitting amino acid substitution models and parameters according to ProtTest3 were; for the γ, α, δ, ε adaptin family Le and Gascuel (LG) substitution model (Le and Gascuel, 2008) with gamma (G) correction of 2.08, for β adaptins LG modified with the observed amino acid frequencies (+F)+G (1.35), for μ adaptins LG +F,+I (0.03)+G (2.54), for the σ adaptins Jones, Taylor and Thornton model (JTT) (Jones et al., 1992) +F +I (0.11)+G (1.65) and for the HSP90 gene family LG + G (0.47). PhyML was run *via* the South of France Bioinformatics Platform web server (www.atgc-montpellier.fr/phyml/). RaxML was run via the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway web server (www.phylo.org). Bayesian tree topologies and posterior probabilities were calculated using MrBayes version 3.1.2, analyses were run for 1×10^6 generations, removing all trees before a plateau, established by graphical estimation. MrBayes was run on CamGRID. All adaptin trees were rooted at AP3 (Hirst et al., 2011). New sequence data have been submitted to GenBank, and accession numbers for all sequences included in the analysis are given in Table S1.

3. Results

3.1. Representation of AP complexes in kinetoplastids

Genome searches yielded well conserved adaptin subunit homologues from all of the organisms examined. Phylogenetic analysis of the adaptin subunits produced well supported clades,