A bioinformatic analysis of the RAB genes of Trypanosoma brucei

John P. Ackers a, Vivek Dhir b, Mark C. Field b,∗

a Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
b Wellcome Trust Laboratories for Molecular Parasitology, Department of Biological Sciences, Imperial College, Exhibition Road, London SW7 2AY, UK

Received 30 September 2003; received in revised form 30 November 2003; accepted 22 January 2005

Abstract

RAB proteins are small GTPases with vital roles in eukaryotic intracellular transport; orthologous RABs appear to fulfil similar functions in diverse organisms. Trypanosoma brucei spp., the causative organisms of Old World trypanosomiasis of humans and domestic animals, have extremely effective endocytic and exocytic mechanisms that are likely to be involved in maintenance of infection, making study of these systems of importance. Taking advantage of the essential completion of the T. brucei genome, we have re-examined the T. brucei RABs (TbRABs) so far described and identified a total of 16. BLAST searches and phylogenetic analysis show that nine of the TbRABs can confidently be assigned as orthologues or homologues of known RAB proteins from higher eukaryotes, and four more with reasonable probability. The core endocytic pathway is probably similar in complexity to yeast, whilst the early exocytic pathway appears to be more complex than in yeast. Two of the TbRAB family (RAB23 and 28) with clear mammalian orthologues appear to be unusual, and may be involved in nuclear processes and are described in more detail in an accompanying paper. Three TbRABs appear, however, to have no close homologues and may fulfil specialised functions in this organism. The availability of a complete set of TbRABs – which includes orthologues of the RABs responsible for control of the core of the endomembrane system (i.e. RAB1, 2, 4–7 and 11) – provides a first overview of the trafficking complexity that is present within a kinetoplastid parasite. Based on these homologies we suggest a systematic nomenclature for the TbRABs to reflect their functional homologies. This information is of importance both from the perspective of understanding the evolution and diversity of eukaryotic trafficking, but also in providing a framework by which to understand protein processing, trafficking, endocytosis and other related processes in these parasites.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Small GTPase; RAB; Trypanosoma; Intracellular transport; Endocytosis; Exocytosis

1. Introduction

The members of the Trypanosoma brucei complex are the causative agents of a number of diseases of humans and ungulates of major public health and economic importance essentially in rural areas in an irregular belt across much of equatorial Africa. Within this region T. b. brucei is not infective to humans but causes nagana, a serious wasting disease of cattle, while T. b. rhodesiense and T. b. gambiense cause East and West African human trypanosomiasis (sleeping sickness). Devastating epidemics of sleeping sickness are re-emerging in many sub-Saharan countries with possibly half a million persons infected; untreated the illness is invariably fatal [1]. The unremitting course of the disease is due to the ability of African trypanosomes to produce an extensive repertoire of antigenic variant surface glycoproteins (VSGs). In addition, when host antibody does bind to surface VSG, the parasite internalises the complex, degrades the antibody and returns the VSG to the plasma membrane [2,3]. This, plus the fact that all this traffic takes place through the relatively small flagella pocket [4], suggests that T. brucei possesses endocytic and exocytic mechanisms that are likely important...
in maintenance of infection. Furthermore, the very ability of *T. b. rhodesiense* to survive in humans requires expression of the SRA gene product (a modified VSG molecule [5]) which interacts with endocytosed apolipoprotein L-1 and/or trypanosome lytic factor (TLF) preventing release of lysozomal contents by an unknown mechanism [6]. The mechanism by which *T. b. gambiense* resists lysis by human serum is currently unknown although it does not depend on the presence of the SRA gene [7].

The secretory and endocytic mechanisms of *T. brucei* have recently been reviewed [8–10]. Central to the correct functioning of vesicle mediated transport of proteins and lipids between membrane-limited compartments are the RAB proteins—the most numerous members of the ras superfamily of small GTPases [11]. RAB proteins have been implicated in several essential steps in these processes—the tethering of vesicles to their correct membrane targets prior to SNARE-mediated membrane fusion, the interaction of vesicles with cytoskeletal elements and, possibly, vesicle budding [12]. RAB proteins are found in all eukaryotes but the number of RAB genes increases with organismal complexity from a likely minimum of seven in *Schizosaccharomyces pombe* to over 60 distinct genes in humans, with alternative splicing probably increasing the number of discrete proteins produced. In mammals at least, some are ubiquitously expressed whereas others are tissue-specific. This increased complexity is presumed to reflect the need for more distinct and specialised trafficking pathways in the higher eukaryotes.

The ability of RAB proteins to provide specificity within the endomembrane system is related to their common structural features. The N-termini contain a number of well-conserved domains which confer GTPase activity (referred to as G (guanine) and PM (phosphate/Mg^{2+}) regions), well-conserved domains which confer GTPase activity (referred to as G (guanine) and PM (phosphate/Mg^{2+}) regions), distinguish RABs from other small GTPases (F1–F5 regions) and define their membership of a number of sub-families of RAB proteins with related functions (SF1–SF4 regions) [13]. These motifs are shown in detail in Fig. 1. The C-termini show very little conserved sequence and are believed to be responsible for the binding of a number of very diverse effector proteins which mediate the specific activities of each RAB [14]. An important insight was the realisation that different compartments in the endocytic and exocytic pathways contain distinct RABs on their surfaces [12,15]. At the same time Perira-Leal and Seabra [16] showed that in a phylogenetic tree of RABs from diverse eukaryotes the proteins grouped into families based on sequence and/or function, rather than species, suggesting that such an approach could be used to examine the trafficking potential of an organism by analysis of the RAB genes in the genome.

We have previously described several members of the trypanosome RAB family, and have reported on their functions and several other aspects [2,17–23]. As the sequencing of the *T. brucei* genome is essentially complete, we have employed a bioinformatics approach to identify and analyse all the *T. brucei* RAB sequences to determine their functional homologues and to provide detailed insight into the intracellular trafficking pathways available to this organism.

2. Materials and methods

2.1. Database extraction of *T. brucei* RAB sequences

Sixteen putative *T. brucei* RAB proteins have already been identified by us and others by BLAST search of various databases using either higher eukaryote sequences or experimentally obtained sequence data from *T. brucei* (Table 1). To ensure exhaustive coverage of the available data, each was used to BLAST search the *T. brucei* GeneDB database (http://www.genedb.org/genedb/tryp/index.jsp). Full length identical or near identical nucleotide sequences and predicted proteins were identified for 15 and a partial but identical nucleotide sequence for TmRAB1. A consensus RAB sequence (Fig. 1, [13]) was then used to search the same databases, but no new significant full-length hits were found.

2.2. RAB comparison sets

Two sets of RAB sequences were prepared for comparison with the *T. brucei* RABs—128 sequences from *Schizosaccharomyces pombe* (7), *Caenorhabditis elegans* (28), *Drosophila melanogaster* (29) and *Homo sapiens* (64) (set 1) and set 2—set 1 plus 11 additional sequences from *Saccharomyces cerevisiae* and a further 57 from *Arabidopsis thaliana*. These sequences are as listed in the supplementary information to Ref. [16] with the following alterations and additions: *C. elegans* RAB10B removed (accidental duplicate of 10A); *D. melanogaster* RAB2 replaced by AAM70817 and the human RABs replaced by an updated list kindly provided by M. Seabra (workers interested in the sequence datasets described here used here should contact the authors).

2.3. BLAST searches

Two strategies were used to identify the orthologues or homologues of the *T. brucei* RABs. Firstly, the GenBank
<table>
<thead>
<tr>
<th>Current name; accession</th>
<th>Size (AA)</th>
<th>CD a</th>
<th>Identity by reverse BLAST</th>
<th>Identity by phylogenetic reconstruction</th>
<th>Location in trypanosomes [reference]</th>
<th>Location of orthologue [reference]</th>
<th>Functional group [16]</th>
<th>Systematic name</th>
</tr>
</thead>
<tbody>
<tr>
<td>TbRAB1; AY377073</td>
<td>208</td>
<td></td>
<td>RAB1</td>
<td>Golgi [40]</td>
<td>ER, Golgi, intermediate compartment, cis-Golgi [38]</td>
<td>I TbRAB1A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB2; T. brucei GeneDB trypX-467a02.plc</td>
<td>225</td>
<td></td>
<td>RAB1 homologue</td>
<td>No data</td>
<td>ER, Golgi, intermediate compartment, cis-Golgi [38]</td>
<td>I TbRAB1B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB2B; AY370774</td>
<td>212</td>
<td></td>
<td>RAB2</td>
<td>RAB2</td>
<td>ER, Golgi [40]</td>
<td>II TbRAB2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB4; AAC46990</td>
<td>198</td>
<td></td>
<td>RAB4 orthologue</td>
<td>Functional group II member, weakly as RAB4</td>
<td>Recycling endosome [22]</td>
<td>II TbRAB4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB5A; AAC4699l (as RAB5)</td>
<td>230</td>
<td></td>
<td>RAB5 orthologue</td>
<td>RAB5 (weak)</td>
<td>Early and recycling endosome, plasma membrane [38]</td>
<td>V TbRAB5A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB5B; AAC76731</td>
<td>203</td>
<td></td>
<td>RAB5 orthologue</td>
<td>RAB5 (weak)</td>
<td>Early endosome [18,22]</td>
<td>V TbRAB5B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB6; T. brucei GeneDB TRYVjpl[p10-g10_p1c]</td>
<td>230</td>
<td></td>
<td>RAB6 orthologue</td>
<td>RAB6</td>
<td>No data</td>
<td>VI TbRAB6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB7; AAC56673</td>
<td>220</td>
<td></td>
<td>RAB7 orthologue</td>
<td>RAB7</td>
<td>Late endosome [34]c</td>
<td>VII TbRAB7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB11A; AAC39034 also as TbRAB11; AAF70820</td>
<td>216</td>
<td></td>
<td>RAB11 orthologue</td>
<td>Recycling endosome [19]</td>
<td>Recycling endosome, trans-Golgi network, plasma membrane [38]</td>
<td>II TbRAB11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB18; AAF37004</td>
<td>215</td>
<td></td>
<td>RAB18 orthologue</td>
<td>RAB18 (weak)</td>
<td>Golgi [17]</td>
<td>No group TbRAB18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB21; T. brucei GeneDB CONTG11942</td>
<td>269</td>
<td></td>
<td>RAB21 orthologue</td>
<td>RAB21</td>
<td>trans-Golgi network, apical plasma membrane [38]</td>
<td>V TbRAB21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB23; AY377075 also T91314 (as a RAB1)</td>
<td>225</td>
<td></td>
<td>RAB23 homologous</td>
<td>RAB23</td>
<td>Nuclear [41]</td>
<td>No data Group TbRAB23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB25; T. brucei GeneDB CONTG12099</td>
<td>240</td>
<td></td>
<td>RAB28 orthologue</td>
<td>RAB28</td>
<td>Nuclear [41]</td>
<td>No data Group TbRAB25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB2A; CA08211 (as RAB1)</td>
<td>218</td>
<td></td>
<td>A RAB but with no obvious homologue</td>
<td>Nearest neighbour is always TbRABX2</td>
<td>ER [21]</td>
<td>NA TbRABX1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB31; CA08210 (as RAB7)</td>
<td>219</td>
<td></td>
<td>A RAB but with no obvious homologue</td>
<td>Nearest neighbour is always TbRABX1</td>
<td>trans-Golgi [20]</td>
<td>NA TbRABX2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TbRAB11B; AAC39035</td>
<td>186</td>
<td></td>
<td>A RAB but with no obvious homologue</td>
<td>Not significantly similar to any other RAB</td>
<td>No data</td>
<td>NA TbRABX3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The current name, together with the GenBank or GeneDB accession number, are given. The size of the predicted open reading frame is shown, together with the predicted CD category. The basis for the identification of an orthologue, and the identity of the orthologue is shown for both BLAST and phylogenetic methods. Also, the experimentally determined location (or predicted in the case of very close homologues) is compared with that for available mammalian orthologues, where appropriate. The predicted functional group, based on phylogenetic analysis with sets 1 and 2, is shown, and as described in [15]. Finally, the proposed systematic name is given at right. Sequences corresponding to TbRAB5A and 4 were originally identified by El-Sayed and Donelson [39], for TbRAB1, 2A, 2B, 5B, 11A, 11B, 18, 23, and 31 were originally identified in our laboratory, and the remainder were first detected in the Sanger/TIGR databases.

a CD: conserved domains identified by NCBI BLAST search.
b In Leishmania mexicana and Boshart, M., personal communication.
c NA: not applicable.
Fig. 2. Neighbour-joining phylogenetic tree of all 16 TbRABs. The tree was constructed following alignment of the complete predicted protein sequences in CLUSTAL W, with TbRAN assigned as an outgroup. Proteins are labelled with the proposed systematic names and bootstrap percentages (1000 boot replicates) given on the branches leading to the nodes.

non-redundant (nr) protein database was searched with each sequence in turn and the highest scoring non-protist hit identified. This sequence was then used to reverse BLAST search the T. brucei GeneDB databases; if the closest match was the original T. brucei RAB then this was regarded as an orthologue of the nr hit; if not as a homologue. Secondly, a local BLAST search of the set 2 sequences was carried out with each T. brucei RAB.

2.4. Phylogenetic reconstruction

Sequences were aligned using CLUSTAL W [24], manually edited with Seaview [25] or BBedit (Bare Bones Software Inc. www.barebones.com) and neighbour-joining, maximum-parsimony and maximum-likelihood trees produced and bootstrapped using CLUSTAL, PAUP* [26] and Phy lip 3.6_3 (distributed by Felsenstein, J. (2002), Department of Genome Sciences, University of Washington, Seattle). For the molecular phylogenies shown in Figs. 2 and 3, analysis was performed using PAUP V4.0b10 (Altevce) on a dual 800 MHz G4 Macintosh (Apple Computer Inc.), using the distance setting. Trees were built using neighbour-joining, and bootstrapped (1000 replicates) using either full heuristic/reten tion of groups at >50% frequency or a neighbour-joining algorithm, as dictated by processor limitations.

3. Results

3.1. The RAB family in T. brucei has 16 members

An exhaustive search of the T. brucei GeneDB database revealed full length nucleotide sequences and predicted proteins identical to 15 of the 16 TbrRABs described here (Table 1) and a partial nucleotide sequence coding for amino acids 1–187 of TbrRAB1. (The nomenclature used in column 1 is that of the original authors.) There was however no evidence to suggest the existence of any additional TbrRABs. This is consistent with an estimate of ~20 obtained by probing a T. brucei genomic library with a degenerate oligonucleotide corresponding to the highly conserved WDTAG region [23]. Table 1 (column 4) also contains the results of BLAST searches of the GenBank non-redundant protein database, T. brucei GeneDB and our set 2 databases, which suggest consistent assignment of most, but not all, T. brucei RABs (the exceptions being TbrRABs 2A, 11B and 31) as orthologues or homologues of defined RAB family members from other organisms [16].

3.2. Phylogenetic analysis allows assignment of orthologies in trypanosomes

A neighbour-joining phylogenetic tree of the 16 T. brucei RABs with T. brucei RAN (sequence assembled from Q26683 and AAA79868 [27]) as outgroup; Fig. 2) shows the close similarity between TbrRABs X1 (2A) and X2 (31) and between 5A and 5B and suggests (but with lower support) a relationship between TbrRAB 1A (1) and 1B (Z) and TbrRAB23 and 28 (current names in parentheses; see below for description of systematic nomenclature). A specific relationship between TbrRABs 6 and 21 is also indicated from this analysis.

A second phylogenetic tree of a large number of RAB sequences from higher organisms (set 1, see Section 2 for details) together with the T. brucei RABs (Fig. 3) shows again the close similarity between TbrRABs 5A and 5B and X1 (2A) and X2 (31). There is also significant support for a specific similarity between TbrRABs 1A (1), 2 (2B), 4, 6, 7, 11 (11A), 21, 23 and 28 and known RAB family members, but not (with >50% support) for any specific homologues of TbrRABs 1B (Z), 5A and 5B and 18, nor for X1, 2 or 3 (2A, 31, 11B). Note that TbrRAB4 is only just significantly segregated with a large group containing RAB4s as well as RABs 2, 11, 14 and 25, and that TbrRAB6 clusters with a group containing RAB41 as well as RAB6s – these combinations comprising functional groups II and VI, respectively (Table 2 and [16]). These data are summarised in Table 1 (column 4). These low bootstrap values are at least in part due to evolutionary distance between the trypanosome and other species included in the tree.

To further extend this analysis, the regions extending from SF1 to SF4 and concatenated SF1–SF4 (Fig. 1) from all of the TbrRABs were also aligned and neighbour-joining maximum-parsimony and maximum-likelihood trees constructed. The
Fig. 3. Neighbour-joining circular cladogram tree of 144 RAB sequences from selected eukaryotes. The 16 TbRABs together with 128 other RAB sequences (set 1—see text) were aligned using CLUSTAL W. Some manual trimming of large, and likely erroneous, N and C-terminal extensions was performed prior to loading the alignment into PAUP. TbRABs are labelled with the proposed new names and indicated in bold. Other species are indicated by prefixes: Schizosaccharomyces pombe (S), Caenorhabditis elegans (C), Drosophila melanogaster (D) and Homo sapiens (no prefix). The tree was bootstrapped 1000 times. RAB families as defined in [13] are colour coded as indicated. Some minor differences between the phylogeny here and that reported in [13] are most likely the result of differences between the tree building procedures and the presence of additional sequences.

Fig. 3 is a reliable indication of evolutionary relationships between the RAB proteins.

3.3. A systematic nomenclature for trypanosome RAB proteins

From the analysis above, nine TbRABs (1A, 1B, 2B, 4, 6, 7, 11(11A), 21, 23 and 28; current names in parentheses)
Table 2

<table>
<thead>
<tr>
<th>T. brucei RABs (current and systematic names) with their functional groups is defined by Pereira-Leal and Seabra [16]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAB functional group</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td>V</td>
</tr>
<tr>
<td>VI</td>
</tr>
<tr>
<td>VII</td>
</tr>
<tr>
<td>VIII</td>
</tr>
<tr>
<td>Ungrouped</td>
</tr>
</tbody>
</table>

- For functional groups as defined in Ref. [16] and are based on phylogenetic reconstruction. These groupings frequently contain RABs from multiple families with related sequences, for example, both RAB11 and RAB4 are in group II, and play a role in recycling of endocytic cargo molecules.
- Example group members from the mammals.
- A systematic nomenclature is proposed based on the analysis in this report, to more accurately reflect evolutionary relationships and probable functional similarities between the trypanosome and higher eukaryote RABs.

These proteins can be confidently assigned as orthologues of known RAB proteins and thus justifiably be named to match. Four TbRABs (currently Z, 5A, 5B and 18) can be assigned by BLAST searching, although this is not fully confirmed by the tree shown in Fig. 3. We feel justified, however, in assigning these proteins as RAB1B, 5A, 5B and 18 for the following reasons: TbRABZ is significantly similar to TbRAB1A when TbRABs are compared amongst themselves (Fig. 2) and the assignment of 5A, 5B and 18 is supported, albeit weakly, by additional trees constructed using alternative algorithms. Most significantly, none of the large numbers of trees examined has suggested any alternative assignments, and, additionally direct functional analysis of TbRAB5A, 5B and 18 is consistent with these assignments. As already noted by Field and Field [23], the two recently duplicated TbRABs (currently designated 2A and 31) are not clearly related to other RABs; TbRAB11B is also not obviously a member of any RAB family. Following the convention of Pereira-Leal and Seabra [16] for naming RABs with no obvious homologues, we propose that the present TbRABs 2A and 31 should be called TbRABX1 and TbRABX2, respectively, and 11B TbRABX3.

3.4. Sequence conservation and deviation within trypanosome RABs

The major RAB-defining motifs (F1–F5) which distinguish RAB proteins from other members of the RAS superfamily [13], together with the C-termini are shown in Fig. 4. When these regions of the 16 T. brucei proteins are aligned with the published consensus sequences it is clear that although some motifs deviate considerably from this consensus (underlined) none has more than two out of four such aberrant motifs (F3 is excluded from this analysis as it is in general less highly conserved than the other four). In addition, deviation from the consensus is in

Fig. 4. Consensus sequences of trypanosome RAB proteins. The consensus RAB-defining motifs F1–F5 which distinguish RAB proteins from other members of the RAS superfamily [13] (amino acids in upper case are conserved in more than 50% of the RAB sequences examined) and the aligned F1–F5 regions and C-terminal amino acids of the 16 TbRABs are shown. Residues in bold differ from the consensus; motifs underlined have more than two such residues. F3 is excluded from this classification because it is less highly conserved than the other four.
most cases represented by a conservative substitution, and hence no major sequence feature appears to differentiate the trypanosome RAB family from those of higher eukaryotes. Most RAB proteins are modified by the addition of a geranylgeranyl group to one or two C-terminal cysteines contained in CAAX, XXCC, XCXX, CCXX, CCXXX or CCXX motifs [16]; 14 of the 16 T. brucei RABs terminate with a sequence containing at least one cysteine. Two interesting C-termini occur in TbrRABs 23 and 28; these are rich in lysine and terminate with -CSVM and -CAVM, respectively; these are normally substrates for protein farnesyltransferase [28] (although the prenylating enzymes of T. brucei may differ in substrate specificity from mammals [29]). It is tempting to link this atypical C-terminus to the (so far unique) nuclear location of TbrRAB23 [40]. The C-terminus of TbrRABX2 (31) (KWRC) has not been found in other RAB proteins but still contains a cysteine.

The most unusual C-terminus is found in TbrRABX3 (11B) – FIKE, which has no C-terminal cysteines. To confirm this unusual C-terminus we compared the entire sequence of TbrRABX3 as determined in our laboratory (GenBank AAG39035) to that obtained by the sequencing project – the sequences were identical suggesting no sequencing errors. Secondly we searched GeneDB for orthologues of TbrRABX3 in other kinetoplastids and detected them in T. cruzi, T. vivax, T. congolense, L. major and L. infantum. Although the T. cruzi sequence is missing its N-terminus and the T. congolense homologue its C-terminus they are otherwise extremely similar. The L. major, L. infantum and T. vivax RABX3s contain an -ILKE and T. cruzi a -LIKE motif near the C-terminus (data not shown). Although very rare, the absence of a prenylation motif in RAB family proteins is not unprecedented. Some of these abnormal RABs have an apparently typical prenylation motif, but this is followed by a C-terminal extension (for example H. sapiens RAB6C WTH3 [30]), whereas others such as the very unusual plant RAB5-like GT-Pase AaRABF1 [31], Plasmodium falciparum RAB5b [32] and TbrRABX3 are truncated. AaRABF1 has an acylated N-terminal extension that may functionally replace prenylation and PIRAB5a a putative myristoylation site but there is no equivalent N-terminal sequence or myristoylation site [33] in TbrRABX3 or its kinetoplastid orthologues. Hence, TbrRABX3 is highly unusual in displaying no clear mechanism for membrane attachment. Clearly, direct experimentation is required to investigate this issue further.

4. Discussion

T. brucei has 16 RABs, a complement substantially greater than S. pombe which has only seven, S. cerevisiae with a complement of 11 and Plasmodium with 12. The number is rather less than the multicellular systems, for example Caenorhabditis elegans has 29 RABs and humans have over 60, but does suggest that trypanosomes have a comparatively sophisticated endomembrane system for a single cell organism. Some newly identified T. brucei RABs have been named by analogy with known proteins while others have simply been numbered in an arbitrary manner, giving rise to some ambiguities in the literature and on the databases. The work of Pereira-Leal and Seabra [13,16], extending the earlier observations of others has led to a consistent nomenclature for the majority of yeast, metazoa and plant RABs from six fully-sequenced genomes (these comprise the molecules included in set 2). It is thus possible to identify orthologues or homologues of newly described RABs and map them onto the established numbering system for the human RABs. Such a consistent scheme of nomenclature, if carefully assigned, has the major advantage of providing suggestions as to function, and here we propose a systematic set of names for T. brucei in the final column of Table 1. Are all of the 16 actually RAB homologues? Firstly the top BLAST hit in the NCBI nonredundant database is always a RAB protein although in three cases the conserved domain identified was the highly related RAS domain (Table 1). Secondly all of these proteins have at least two of Pereira-Leal and Seabra’s [13] five RAB-defining motifs (F1–F5) which are highly similar to the consensus sequences. Hence, all of these sequences can be assigned as RAB family members.

As S. cerevisiae and mammals are the best understood model systems in terms of intracellular transport processes, a comparison between these organisms and T. brucei is illuminating. Firstly, the major core functionality is well conserved, so that the trypanosome RAB proteins that are likely responsible for control of the basic exocytic and endocytic pathways have clear orthologues in yeast and mammals. This core set includes RAB1A, 1B, 2, 4, 5A, 5B, 6, 7 and 11 and most significantly, where experimental data are available, these fully support the in silico assignments made here [17–22,34]. This indicates that trypanosomes likely utilise a conserved mechanism for ER to Golgi transport (RAB1 and RAB2), trafficking through the early endosome (RAB5), recycling (RAB4 and RAB11), delivery to the late endosome/lysosome (RAB7) and retrograde transport through the Golgi complex (RAB6). It is also interesting to note that the recycling system, represented in trypanosomes by RAB4 and RAB11, is rather more similar to the situation found in mammals than in the yeast where a similar role is performed by ypt31 and ypt32; however in mammals and trypanosomes RAB4 and RAB11 conduct distinct functions, whereas ypt31 and ypt32 are functionally redundant [35]. It is also interesting that trypanosomes have two RAB5 proteins, whilst S. cerevisiae and mammals have three; phylogenetic reconstruction indicates that the multiple members of the RAB5 family have arisen in a lineage-dependent manner, presumably reflecting adaptation to a specific lifestyle [22]. Therefore, the basic plan of the endomembrane system appears to be very ancient indeed, whilst gene duplication has facilitated the building of additional, lineage-specific complexity.

Secondly, a proportion of the additional RAB complement in trypanosomes appears to have arisen through the presence of three divergent RABs, i.e. RABX1, RABX2 and RABX3.
Two of these, RABX1 and RABX2 (formerly RAB2 and RAB31), arose as a tandem duplication that predates the speciation of Old World and New World trypanosomatids, as near identical genes are found in syntenic arrangement in the L. major, T. vivax and T. cruzi genomes (data not shown). Whilst close homologues of RABX3 are present in Leishmania and other trypanosomes, no clear orthologue could be found amongst the crown eukaryotes. Further, T. brucei also possesses orthologues of mammalian RAB proteins that are not present in the yeast genome. These include RAB2B, RAB18, RAB21, RAB23 and RAB28. RAB18 is localised to the Golgi apparatus [17], whilst RAB21 remains uncharacterised at this time. Preliminary data suggests that RAB23 at least may play a novel role in the nucleus and interestingly, is highly similar to RAB28. Further, phylogenetic reconstruction suggests that RAB23 and RAB28 are ancient RAB proteins that arose before the speciation event separating trypanosomes from the mammalian lineage, and therefore have most likely been lost before the speciation event separating trypanosomes from the mammalian lineage, and therefore have most likely been lost before the speciation event separating trypanosomes from the mammalian lineage.

References

Pal A, Hall BS, Nesbitt DN, Field HI, Field MC. Differential en-

Jeffries TR, Morgan GW, Field MC. TbRAB18, a developmentally

Pereira-Leal JB, Seabra MC. The mammalian RAB family of small

Pereira-Leal JB, Seabra MC. Evolution of the RAB family of small

Field H, Ali BR, Sherwin T, Gull K, Croft SL, Field MC. TbRAB2p, a marker for the endoplasmic reticulum of


Seabra MC, Mules EH, Hume AN. RAB GTPases, intracellular traf-


Yoshida T, Yamaguchi M, Uchimya H, Nakano A, Arakawa, a plant-

Ueda T, Yamauchi N, Nakano A, Arakawa, a plant-

Ueda T, Yamaguchi M, Uchimya H, Nakano A, Arakawa, a plant-

Seagram MC, Mules EH, Hume AN. RAB GTPases, intracellular traf-


Jordi G, Mullolland J, Seely N. Two new ypt GTPases are re-

de Lecque HP, Koster PM, Calcott J, et al. Small GTP-binding pro-

Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-

Field H, Ali BR, Sherwin T, Smith AC, Gull K, Field MC. Tandem duplication of rab genes fol-


Roskoski Jr R. Protein prettlination: a pivotal posttranslational pro-

Field MC, Field H, Boothroyd JC. A homologue of the nuclear


GTPase family: an-


Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-


Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-


de Lecque HP, Koster PM, Calcott J, et al. Small GTP-binding pro-

Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-

Jordi G, Mullolland J, Seely N. Two new ypt GTPases are re-

Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-

Pereira-Leal JB, Seabra MC. The mammalian RAB family of small

Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-


