Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 141 (2005) 89–97
MOLECULAR & BIOCHEMICAL PARASITOLOGY

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.

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

Abbreviations: G, RAB guanine binding region; PM, RAB phosphate/Mg^{2+} binding region; F, RAB-defining region; SF, RAB subfamily defining region; TbRAB, *Trypanosoma brucei* RAB protein

* Corresponding author. Present address: Department of Pathology, University of Cambridge, Tennis Count Road, Cambridge, CB2 1QP, UK. Tel.: +44 1233 333734.
E-mail address: mcf34@cam.ac.uk (M.C. Field).

0166-6851/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2005.01.017

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 apolioprotein L-1 and/or trypanosome lytic factor (TLF) preventing release of lysosomal 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), 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.

ydylFKllli*GDsgVGKT***SLllRFtddt***Fs**xxyks***TIG**
VDFktKTvevdgkkv***KLQIW***DTAGQE***RFrsiT***ra***YYRG**
AqGai***LVYDIT***neeSF**enlknWlkelreyaepndvvim**
LvGNKcDLeekrvvvteeAekfAeengipffET*SAK*tn
iNVeeaFetlareil**kkareldleasqn**dgslggvkle
qlppqpskqkpgccCs

Fig. 1. Profile HMM model sequence of mammalian RAB sequences. The consensus sequence taken from a Hidden Markov Model (HMM) profile [13] used for searching various databases is shown, together with critical sequence and functional regions highlighted. Bold: RAB Sub-family defining sequences SF1–SF4, italic: GTP-binding regions, bold and italic: RAB defining regions F1–F5; these distinguish RAB proteins from other members of the RAS superfamily, double underlined: prenylation signal, underlined: switch regions; these change conformation when GTP is hydrolysed.

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 TbRAB1. 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 1
Summary of reverse BLAST and phylogenetic analysis of *T. brucei* RABs, with proposed systematic nomenclature (for details see text)

Current name; accession	Size (AA)	CD[a]	Identity by reverse BLAST	Identity by phylogenetic reconstruction	Location in trypanosomes [reference]	Location of orthologue [reference]	Functional group [16]	Systematic name
TbRAB1; AY377073	208	RAB	RAB1 orthologue	RAB1	Golgi [40]	ER-Golgi, intermediate compartment, *cis*-Golgi [38]	I	TbRAB1A
TbRABZ; *T. brucei* GeneDB tryp_X-467d02.p1c	225	RAB	RAB1 homologue	No close homologue	No data	ER-Golgi, intermediate compartment, *cis*-Golgi [38]	I	TbRAB1B
TbRAB2B; AY370774	212	RAB	RAB2 orthologue	RAB2	ER/Golgi [40]	ER-Golgi, intermediate compartment, *cis*-Golgi [38]	II	TbRAB2
TbRAB4; AAC46990	198	RAB	RAB4 orthologue	Functional group II member; weakly as RAB4	Recycling endosome [22]	Early and recycling endosome, plasma membrane [38]	II	TbRAB4
TbRAB5A; AAC46991 (as RAB5)	230	RAS	RAB5 orthologue	RAB5 (weak)	Early endosome [18,22]	Plasma membrane, early endosome, clathrin-coated vesicles [38]	V	TbRAB5A
TbRAB5B; AAC78731	203	RAB	RAB5 orthologue	RAB5 (weak)	Early endosome [18,22]	Plasma membrane, early endosome, clathrin-coated vesicles [38]	V	TbRAB5B
TbRAB6; *T. brucei* GeneDB TRYPtp3p8-g10.p1c_1	230	RAB	RAB6 orthologue	RAB6	No data	ER, Golgi, *trans*-Golgi network [38]	VI	TbRAB6
TbRAB7; AAQ15670	220	RAB	RAB7 orthologue	RAB7	Late endosome [34][b]	Late endosome [38]	VII	TbRAB7
TbRAB11A; AAG39034 also as TbRAB11: AAF70820	216	RAB	RAB11 orthologue	RAB11	Recycling endosome [19]	Recycling endosome, *trans*-Golgi network, plasma membrane [38]	II	TbRAB11
TbRAB18; AAF37004	215	RAB	RAB18 orthologue	RAB18 (weak)	Golgi [17]	Early endosome, recycling endosome plasma membrane [38]	No group	TbRAB18
TbRAB21; *T. brucei* GeneDB CONTIG11942	269	RAS	RAB21 orthologue	RAB21	No data	*trans*-Golgi network, apical plasma membrane [38]	V	TbRAB21
TbRAB23; AY377075 also T09134 (as a RAB1)	225	RAS	RAB23 homologue	RAB23	Nuclear [41]	No data	No Group	TbRAB23
TbRAB28; *T. brucei* GeneDB CONTIG12099	240	RAB	RAB28 orthologue	RAB28	Nuclear? [41]	No data	No group	TbRAB28
TbRAB2A; CAA68211 (as RAB1)	218	RAB	A RAB but with no obvious homologue	Nearest neighbour is always TbRABX2	ER [21]	NA[c]	NA	TbRABX1
TbRAB31; CAA68210 (as RAB7)	219	RAB	A RAB but with no obvious homologue	Nearest neighbour is always TbRABX1	*trans*-Golgi [20]	NA	NA	TbRABX2
TbRAB11B; AAG39035	186	RAB	A RAB but with no obvious homologue	Not significantly similar to any other RAB	No data	NA	NA	TbRABX3

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 are 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.

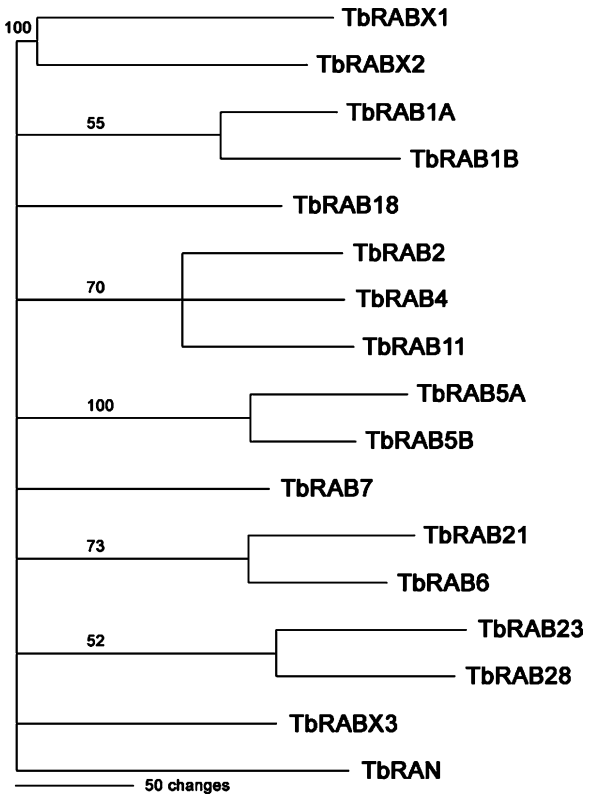

Fig. 2. Neighbour-joining phylogenetic tree of all 16 TbRABs. The tree was constructed following alignment of the complete predicted protein sequences in CLUSTAL W, with TbRAN assigned as an outgroup. Proteins are labelled with the proposed systematic names and bootstrap percentages (1000 boot replicates) 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 Phylip 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 (Altivec) 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/retention 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 TbRABs described here (Table 1) and a partial nucleotide sequence coding for amino acids 1–187 of TbRAB1. (The nomenclature used in column 1 is that of the original authors.) There was however no evidence to suggest the existence of any additional TbRABs. 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 TbRABs 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 TbRABs X1 (2A) and X2 (31) and between 5A and 5B and suggests (but with lower support) a relationship between TbRAB 1A (1) and 1B (Z) and TbRAB23 and 28 (current names in parentheses; see below for description of systematic nomenclature). A specific relationship between TbRABs 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 TbRABs 5A and 5B and X1 (2A) and X2 (31). There is also significant support for a specific similarity between TbRABs 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 TbRABs 1B (Z), 5A and 5B and 18, nor for X1, 2 or 3 (2A, 31, 11B). Note that TbRAB4 is only just significantly segregated with a large group containing RAB4s as well as RABs 2, 11, 14 and 25, and that TbRAB6 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 TbRABs were also aligned and neighbour-joining maximum-parsimony and maximum-likelihood trees constructed. The

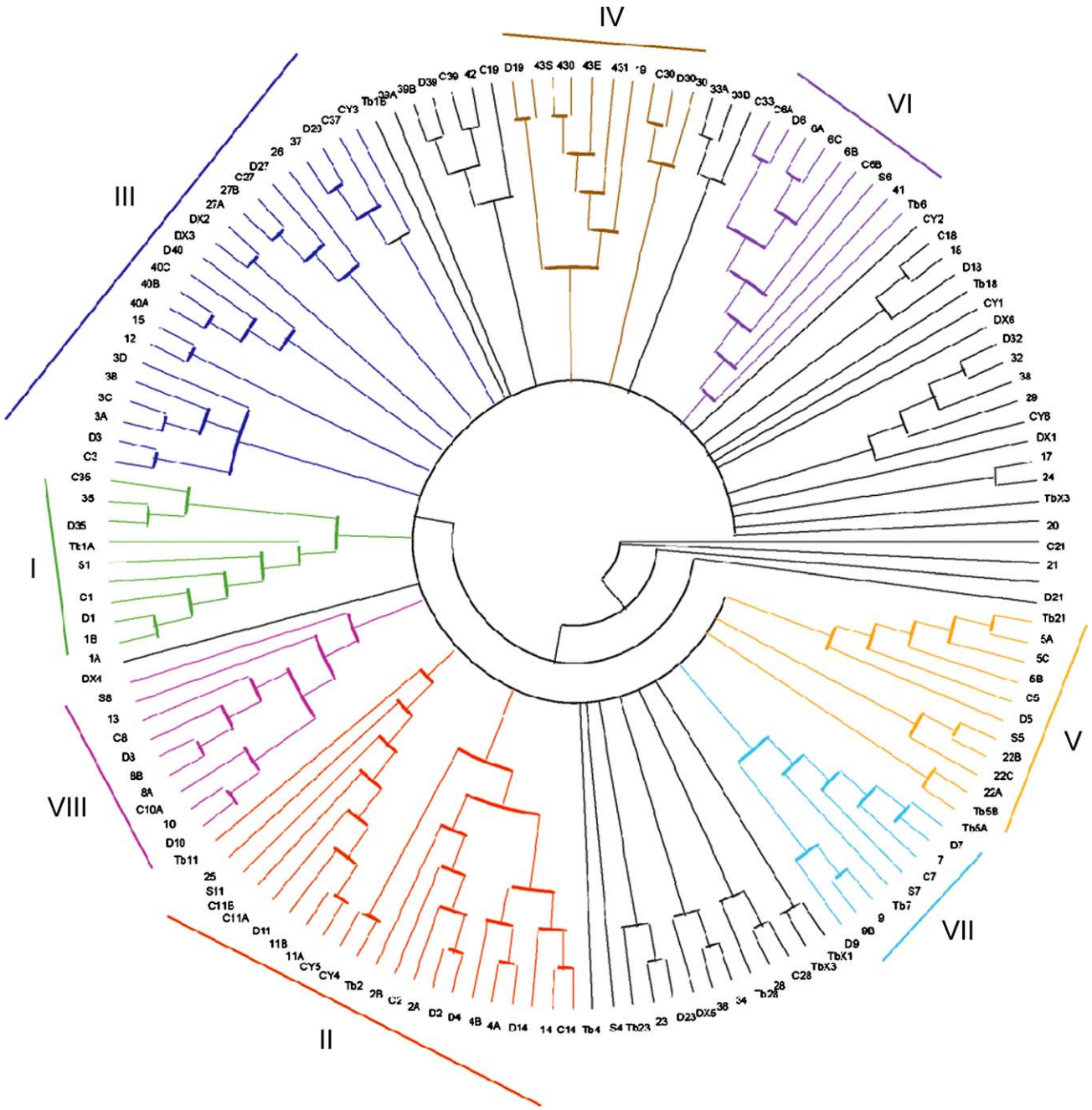

Fig. 3. Neighbour-joining circular cladogram tree of 144 RAB sequences from selected eukaryotes. The 16 TbRABs together with 128 other RAB sequences (set 1—see text) were aligned using CLUSTAL W. Some manual trimming of large, and likely erroneous, N and C-terminal extensions was performed prior to loading 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.

tree topologies in general were not significantly different from that shown in Fig. 3 except that a maximum-likelihood tree of the complete sequences also weakly (54%) supports a specific relationship between TbRAB4 and other RAB4s and also between TbRAB18 and RAB18s (60%) and a maximum-parsimony tree similarly weakly supported (55%) a relationship between TbRABs 5A and 5B and other RAB5s (data not shown). Therefore, overall the topology of the tree shown in Fig. 3 is a reliable indication of evolutionary relationships between the RAB proteins.

3.3. A systematic nomenclature for typanosome RAB proteins

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

Table 2
T. brucei RABs (current and systematic names) with their functional groups is defined by Pereira-Leal and Seabra [16]

RAB functional group[a]	Group members[b]	RAB family members present in *T. brucei*: present name (systematic name[c])	Function
I	1, 35	1, Z (1A, 1B)	Exocytosis
II	2, 4, 11, 14, 25	2B, 4, 11A (2, 4, 11)	Endocytosis and recycling
III	3, 26, 27, 37	ND	Regulated exocytosis
IV	19, 30	ND	Golgi related
V	5, 21, 22	5A, 5B, 21 (5A, 5B, 21)	Endocytosis
VI	6, 41	6 (6)	Retrograde Golgi transport
VII	7, 9	7 (7)	Late endocytosis
VIII	8, 10, 13	ND	Golgi related
Ungrouped	–	18, 23, 28, 2A, 31, 11B (18, 23, 28, X1, X2, X3)	

ND: not detected. No trypanosome RAB found that falls into this group.
[a] 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.
[b] Example group members from the mammals.
[c] 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.

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

TbRAB		Motif and consensus					
Current	Systematic	F1 IGVDF	F2 KLQIW	F3 RFrsiT	F4 YYRGA	F5 LVYDIT	C-term./prenylation …SKQKPGCCCS
1	1A	IGVDF	KLQIW	RFrtiT	YYRGA	IVYDTT	…NSGGQKSGCC
Z	1B	LSVDV	QVLMY	RFramT	FYRDA	MVYDTT	…SQRNDNCCCQ
2B	2	IGVEF	KLQIW	SFrsiT	YYRGA	LVYDVT	…SEDSESGCFC
4	4	IGVEF	KLQIW	RYksvT	YYRGA	IVYDIT	…DGPKRFSCAC
5A	5A	IGAAF	KFEIW	RYrslA	YYRGA	VVYDIT	…QQKKEGGCAC
5B	5B	GRRVF	NFDIW	RYkslA	YYRGA	VVYEIP	…STRRSSGCCG
6	6	IGIDF	RLHVW	RFrslI	YIRNS	VVYDIT	…GAGWREGGCC
7	7	IGADF	TLQIW	RFqslG	FYRGA	LVFDLT	…PSTKKSDCAC
11A	11	IGVEF	KVQIW	RFraiS	IYHGA	LVYDIT	…PAGKASGGCC
18	18	IGIDF	KLHLW	RFrtlT	YYRGA	LVYDVN	…SDPTSGAVCC
21	21	TQASM	NLNVW	RFhalG	YYRNA	LVYDVT	…ATPRRNRCCS
23	23	VGVEY	NSFLW	TVssvK	YYLDA	LVFSTD	…KSKKKKCSVM
28	28	IGLDF	KMEIW	IGgtmI	YIMGA	FVYDVT	…KKKKGKCAVM
2A	X1	IAIDF	RLQIW	RFqsvA	FYRGA	LCFDLT	…QKGSKRGGCC
31	X2	IGVDF	LLQIW	RFvslT	YYRNC	ICFDLT	…THTKKRKWRC
11B	X3	YALTL	DVDIW	RFstmH	YYHEA	LVFDVT	…MSQVLGFIKE

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, XCXC, CCXX, CCXXX or XCCX motifs [16]; 14 of the 16 *T. brucei* RABs terminate with a sequence containing at least one cysteine. Two interesting C-termini occur in TbRABs 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 TbRAB23 [40]. The C-terminus of TbRABX2 (31) (KWRC) has not been found in other RAB proteins but still contains a cysteine.

The most unusual C-terminus is found in TbRABX3 (11B) – FIKE, which has no C-terminal cysteines. To confirm this unusual C-terminus we compared the entire sequence of TbRABX3 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 TbRABX3 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 GTPase AtRABF1 [31], *Plasmodium falciparum* RAB5b [32] and TbRABX3 are truncated. AtRABF1 has an acylated N-terminal extension that may functionally replace prenylation and PfRAB5a a putative myristylation site but there is no equivalent N-terminal sequence or myristylation site [33] in TbRABX3 or its kinetoplastid orthologues. Hence, TbRABX3 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 in 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, metazoan 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-dependant 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 RAB2, 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 from the *S. cerevisiae* genome during evolution. A RAB2 orthologue is absent from the yeast genome where ER to Golgi transport is controlled by ypt1 alone, unlike the additional complexity in mammalian cells where both RAB1 and RAB2 are required. Both TbRAB1A and TbRAB2 are essential by RNAi experiments [39], and clearly involved in trafficking though the early secretory system.

Conversely, there are significant gaps in the trypanosome repertoire when compared to higher eukaryotes, specifically functional groups III, IV and VIII [16] are not represented in the trypanosome genome. Group III contains RABs that generally fall into the category of post-Golgi exocytosis, and in particular, of regulated exocytic pathways. Hence, it is unlikely that *T. brucei* is capable of regulated secretion, which is in full agreement with experimental data where no such pathway has been detected (MCF, unpublished data). Group IV is represented by RAB19 and RAB30; very little is known about these two proteins in higher eukaryotes, but RAB30 has been reported to exhibit tissue-specific expression and to localise to the Golgi complex; this may represent a specialised function that is absent from *T. brucei* [36]. The final unrepresented group is VIII; this clade contains RAB8, RAB10 and RAB13, a highly related set of RABs that are close homologues of *S. cerevisiae* Sec4, and have been implicated in late steps in exocytosis. In mammals it is likely that several distinct exocytotic routes are present as RAB8 is ubiquitously expressed, but the absence of this group from *T. brucei* suggests a simplicity in export to the cell surface. This is also consistent with recent data suggesting that *T. brucei* exocytosis may intersect with the recycling system, so that TbRAB11 is responsible for transport to the cell surface and may replace some of the functions otherwise performed by group VIII proteins [19]. Rutherford and Moore [37] have suggested that homologues of RABs 1, 5–8 and 11 (members of functional groups I, II, V–VIII) comprise the minimal essential eukaryotic set. None of the phylogenetic trees suggests that any of the *T. brucei* unassigned RABs (X1, X2 or X3) are members of group VIII. *T. brucei* is therefore unusual amongst eukaryotes in this respect, although the very limited number of genomes which have been exhaustively examined must always be borne in mind.

The analysis described here provides a first overview of the trafficking complexity that is present within a kinetoplastid parasite. 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.

Acknowledgements

This work was supported in part by the Wellcome Trust. We are most indebted to Miguel Seabra (IC, London) for the updated set of human RAB sequences and to him, José Pereira-Leal (EBI, Cambridge) and Graham Clark (LSHTM) for helpful discussions and advice. *T. brucei* sequence data were obtained from The Sanger Institute website at http://www.sanger.ac.uk/Projects/T_brucei/ and we gratefully acknowledge the provision of this data. Sequencing of the *T. brucei* genome was accomplished as part of the Trypanosoma Genome Network with support by The Wellcome Trust. Workers interested in the sequence datasets described here should contact the authors.

References

[1] Stich A, Barrett MP, Krishna S. Waking up to sleeping sickness. Trends Parasitol 2003;19:195–7.
[2] Pal A, Hall B, Jefferies T, Field MC. RAB5 and RAB11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in *Trypanosoma brucei*. Biochem J 2003;374:443–51.
[3] Seyfang A, Mecke D, Duszenko M. Degradation, recycling, and shedding of *Trypanosoma brucei* variant surface glycoprotein. J Protozool 1990;37:546–52.
[4] Overath P, Stierhof YD, Wiese M. Endocytosis and secretion in trypanosomatid parasites – tumultuous traffic in a pocket. Trends Cell Biol 1997;7:27–33.
[5] Campillo N, Carrington M. The origin of the serum resistance associated (SRA) gene and a model of the structure of the SRA polypeptide from *Trypanosoma brucei rhodesiense*. Mol Biochem Parasitol 2003;127:79–84.
[6] Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, et al. Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature 2003;422:83–7.
[7] De Greef C, Chimfwembe E, Kihang'a Wabacha J, Bajyana Songa E, Hamers R. Only the serum-resistant bloodstream forms of *Trypanosoma brucei rhodesiense* express the serum resistance associated (SRA) protein. Ann Soc Belg Med Trop 1992;72(Suppl 1):13–21.
[8] McConville MJ, Mullin KA, Ilgoutz SC, Teasdale RD. Secretory pathway of trypanosomatid parasites. Microbiol Mol Biol Rev 2002;66:122–54.
[9] Morgan GW, Hall BS, Denny PW, Carrington M, Field MC. The kinetoplastida endocytic apparatus. Part I. A dynamic system for nutrition and evasion of host defences. Trends Parasitol 2002;18:491–6.
[10] Morgan GW, Hall BS, Denny PW, Field MC, Carrington M. The endocytic apparatus of the kinetoplastida. Part II. Machinery and components of the system. Trends Parasitol 2002;18:540–6.

[11] Stenmark H, Olkkonen VM. The RAB GTPase family. Genome Biol 2001;2, reviews3007.1–7.
[12] Zerial M, McBride H. RAB proteins as membrane organizers. Nat Rev Mol Cell Biol 2001;2:107–17.
[13] Pereira-Leal JB, Seabra MC. The mammalian RAB family of small GTPases: definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the Ras superfamily. J Mol Biol 2000;301:1077–87.
[14] Gonzalez Jr L, Scheller RH. Regulation of membrane trafficking: structural insights from a RAB/effector complex. Cell 1999;96:755–8.
[15] Pfeffer SR. RAB GTPases: specifying and deciphering organelle identity and function. Trends Cell Biol 2001;11:487–91.
[16] Pereira-Leal JB, Seabra MC. Evolution of the RAB family of small GTP-binding proteins. J Mol Biol 2001;313:889–901.
[17] Jeffries TR, Morgan GW, Field MC. TbRAB18, a developmentally regulated Golgi GTPase from *Trypanosoma brucei*. Mol Biochem Parasitol 2002;121:63–74.
[18] Pal A, Hall BS, Nesbeth DN, Field HI, Field MC. Differential endocytic functions of *Trypanosoma brucei* RAB5 isoforms reveal a glycosylphosphatidylinositol-specific endosomal pathway. J Biol Chem 2002;277:9529–39.
[19] Jeffries TR, Morgan GW, Field MC. A developmentally regulated rab11 homologue in *Trypanosoma brucei* is involved in recycling processes. J Cell Sci 2001;114:2617–26.
[20] Field H, Sherwin T, Smith AC, Gull K, Field MC. Cell-cycle and developmental regulation of TbRAB31 localisation, a GTP-locked RAB protein from *Trypanosoma brucei*. Mol Biochem Parasitol 2000;106:21–35.
[21] Field H, Ali BR, Sherwin T, Gull K, Croft SL, Field MC. TbRAB2p, a marker for the endoplasmic reticulum of *Trypanosoma brucei*, localises to the ERGIC in mammalian cells. J Cell Sci 1999;112:147–56.
[22] Field H, Farjah M, Pal A, Gull K, Field MC. Complexity of trypanosomatid endocytosis pathways revealed by RAB4 and RAB5 isoforms in *Trypanosoma brucei*. J Biol Chem 1998;273:32102–10.
[23] Field H, Field MC. Tandem duplication of rab genes followed by sequence divergence and acquisition of distinct functions in *Trypanosoma brucei*. J Biol Chem 1997;272:10498–505.
[24] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 1994;22:4673–80.
[25] Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 1996;12:543–8.
[26] Swofford DL. PAUP*: phylogenetic analysis using parsimony (and other methods). Sunderland, MA: Sinauer Associates; 2002.
[27] Field MC, Field H, Boothroyd JC. A homologue of the nuclear GTPase ran/TC4 from *Trypanosoma brucei*. Mol Biochem Parasitol 1995;69:131–4.
[28] Roskoski Jr R. Protein prenylation: a pivotal posttranslational process. Biochem Biophys Res Commun 2003;303:1–7.
[29] Yokoyama K, Lin Y, Stuart KD, Gelb MH. Prenylation of proteins in *Trypanosoma brucei*. Mol Biochem Parasitol 1997;87:61–9.
[30] Shan J, Yuan L, Budman DR, Xu HP. WTH3, a new member of the RAB6 gene family, and multidrug resistance. Biochim Biophys Acta 2002;1589:112–23.
[31] Ueda T, Yamaguchi M, Uchimiya H, Nakano A. Ara6, a plant-unique novel type RAB GTPase, functions in the endocytic pathway of *Arabidopsis thaliana*. EMBO J 2001;20:4730–41.
[32] Quevillon E, Spielmann T, Brahimi K, Chattopadhyay D, Yeramian E, Langsley G. The *Plasmodium falciparum* family of RAB GTPases. Gene 2003;306:13–25.
[33] Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal *N*-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. J Mol Biol 2002;317:523–40.
[34] Denny PW, Lewis S, Tempero JE, et al. *Leishmania* RAB7: characterisation of terminal endocytic stages in an intracellular parasite. Mol Biochem Parasitol 2002;123:105–13.
[35] Jedd G, Mulholland J, Segev N. Two new ypt GTPases are required for exit from the yeast trans-Golgi compartment. J Cell Biol 1997;137:563–80.
[36] de Leeuw HP, Koster PM, Calafat J, et al. Small GTP-binding proteins in human endothelial cells. Br J Haematol 1998;103:15–9.
[37] Rutherford S, Moore I. The *Arabidopsis* RAB GTPase family: another enigma variation. Curr Opin Plant Biol 2002;5:518–28.
[38] Seabra MC, Mules EH, Hume AN. RAB GTPases, intracellular traffic and disease. Trends Mol Med 2002;8:23–30.
[39] El-Sayed NM, Alarcon CM, Beck JC, Sheffield VC, Donelson JE. cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. Mol Biochem Parasitol 1995;73:75–90.
[40] Dhir V, Goulding D, Field MC. TbRAB1 and TbRAB2 mediate trafficking through the early secretory pathway of *Trypanosoma brucei*. Mol Biochem Parasitol 2004;137:253–65.
[41] Dhir V, Field MC. TbRAB23; a nuclear-associated RAB protein from *Trypanosoma brucei*. Mol Biochem Parasitol 2004;136:297–301.