Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular & Biochemical Parasitology 143 (2005) 226–235
MOLECULAR & BIOCHEMICAL PARASITOLOGY

# Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*☆

Kalpana Lal[a,1], Mark C Field[b], Jane M Carlton[c], Jim Warwicker[d], Robert P Hirt[a,e,*]

[a] *Department of Zoology, The Natural History Museum, South Kensington, London SW7 5BD, UK*
[b] *The Molteno Building, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK*
[c] *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*
[d] *Faculty of Life Sciences, University of Manchester, Manchester M60 1QD, UK*
[e] *School of Biology, The Devonshire Building, Devonshire Terrace, The University of Newcastle, Newcastle upon Tyne NE1 7RU, UK*

Received 19 May 2005; accepted 23 June 2005

**Abstract**

Rab proteins are pivotal components of the membrane trafficking machinery in all eukaryotes. Distinct Rab proteins locate to specific endomembrane compartments and genomic studies suggest that Rab gene diversity correlates with endomembrane system complexity; for example unicellular organisms generally possess 5–20 Rab family members and the size of the repertoire increases to 25–60 in multicellular systems. Here we report 65 open reading frames from the unicellular protozoan *Trichomonas vaginalis* that encode distinct Rab proteins (TvRabs), indicating a family with complexity that rivals *Homo sapiens* in number. The detection of gene transcripts for the majority of these genes and conservation of functional motifs strongly suggests that TvRabs retain functionality and likely roles in membrane trafficking. The *T. vaginalis* Rab family includes orthologues of the conserved subfamilies, Rab1, Rab5, Rab6, Rab7 and Rab11, but the majority of TvRabs are not represented by orthologues in other systems and includes six novel *T. vaginalis* specific Rab subfamilies (A–F). The extreme size of the *T. vaginalis* Rab family, the presence of novel subfamilies plus the divergent nature of many TvRab sequences suggest both the presence of a highly complex endomembrane system within *Trichomonas* and potentially novel Rab functionality. A family of more than 65 Rab genes in a unicellular genome is unexpected, but may be a requirement for progression though an amoeboid life-cycle phase as both *Dictyostelium discoideum* and *Entamoeba histolytica* share with *T. vaginalis* both an amoeboid life cycle stage and very large Rab gene families.


*Keywords:* Rab; Evolution; Diversity; Phylogeny; Amoeba

## 1. Introduction

The Rab family of small GTPase proteins are central components of membrane trafficking in eukaryotes and act at multiple stages of intracellular transport processes [1]. Within a given cell type specific Rab proteins locate to distinct membrane trafficking sub-compartments and Rab diversity is often regarded as an expression of membrane trafficking complexity [2].

Unicellular eukaryotes with completed genomes encode between 5 and 20 Rab proteins, including a set of 5 core Rabs required for the basic functions of exocytosis and endocytosis; specifically Rab1, Rab5, Rab6, Rab7 and Rab11 [3]. Rab1

*Abbreviations:* RabF, Rab specific regions; RabSF, Rab subfamily specific regions; TvRab, *T. vaginalis* Rab; EST, expressed sequence tag; ORF, open reading frame; RACE, rapid amplification of cDNA ends; GAP, GTPase activating protein; GDI, guanine dissociation inhibitor

☆ *Note:* Nucleotide sequences reported in this paper are available in the GenBank™ database (accession numbers: AY896243–AY896292 and DQ019033–DQ019047).

* Corresponding author. Tel.: +44 191 246 4805; fax: +44 191 222 5228.
*E-mail address:* R.P.Hirt@ncl.ac.uk (R.P. Hirt).
[1] Present address: Department of Biological Sciences, Imperial College, South Kensington, London SW7 2AZ, UK.

0166-6851/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2005.06.008

is essential for endoplasmic reticulum to Golgi transport, Rab6 functions in *intra-Golgi* traffic, Rab5 is key to endocytosis, Rab7 functions at late endosome/lysosomes and Rab11 plays a vital role at recycling endosomes [1]. Additional Rab sequences are associated with greater complexity within the endomembrane system, e.g. multiple Rab5 isoforms [4] or with tissue complexity where tissue-specific functions are required, e.g. Rab3A function in neuronal exocytosis [5]. Thus, the multicellular organisms *Caenorhabditis elegans* and *Drosophila melanogaster* have 29 Rabs, and humans, with even greater tissue diversity, encode more than 60 Rab proteins in their genome [3].

Rabs are a monophyletic clade of the Ras-like small GTPase protein super-family which function as 'molecular switches', regulating key steps in membrane trafficking including vesicle formation, vesicle motility, membrane remodelling, vesicle docking and membrane fusion [1–3,6]. In common with all GTPases, Rab proteins possess conserved G-block sequence motifs involved in GTP/GDP binding and GTP hydrolysis [7] and can be distinguished by their typical C-terminal prenylation motifs (CCxxx, CCxx, xCCx, xxCC or xCxC) involved in targeting to membranes [8]. Detailed comparative sequence analysis has revealed diagnostic Rab specific regions (RabF) highly conserved in Rab proteins (RabFl: IGVDF, RabF2: KLQIW, RabF3: RFRSIT, RabF4: YYRGA and RabF5: LVYDIT) [3,8]. Early in eukaryotic evolution, the Rab protein family diversified into Rab subfamilies (e.g. Rab1 and Rab11) represented in diverse eukaryotes such as humans and *Trypanosoma brucei* [3,9]. Rab subfamily orthologues from different species locate at equivalent membrane trafficking compartments, performing similar functions and are often functionally complementary [1,10]. Rab subfamily orthologues are thought to bind similar effector and regulator proteins in the Rab subfamily specific sequence regions (RabSF) [3,11]. In an individual species, a Rab subfamily may be represented by closely related isoforms, e.g. *T. brucei* TbRab5a and TbRab5b which are involved in the same overriding process of endocytosis but carry distinct transport cargo [12].

*Trichomonas vaginalis* is a protozoan parasite infecting 200 million people worldwide [13], and is one of the most common sexually transmitted pathogens. *T. vaginalis* lacks a true mitochondrion and thus is an important model system to study eukaryote origins and diversification. For decades it was considered a member of an early branching primitive amitochondriate group [14] but more recent molecular phylogenetic and cell biology investigations are consistent with the hypothesis that *T. vaginalis* previously possessed mitochondria [15]. However, the phylogenetic position of *T. vaginalis* in the eukaryotic tree remains unclear. *T. vaginalis* infection is linked to a higher susceptibility to HIV infection and cervical cancer in women [13]. Upon contact with host tissue, *T. vaginalis* dramatically transforms from a free-swimming trophozoite to an adherent amoeba, which is key to establishment of infection in the vagina and subsequent pathogenesis [16,17]. Membrane trafficking processes, such as those orchestrating secretion of cysteine proteases and perforins [17], display of cell surface adhesins [13] and endocytosis of host proteins [16] or HIV particles [18], all contribute to the complex pathogenesis of *T. vaginalis,* but virtually nothing is known about their underlying molecular cell biology [16]. To initiate the molecular characterization of components regulating membrane trafficking in the protozoan *T. vaginalis* we investigated its Rab gene family repertoire in an expressed sequence tag (EST) database, 5′-RACE clones and the ongoing *T. vaginalis* genome sequencing project [19].

## 2. Materials and methods

### 2.1. Isolation of putative T. vaginalis Rabs genes

Putative Rab genes were identified from a *T. vaginalis* (G3 strain) expressed sequence tag (EST) gene survey with BLASTX [20]. Additional putative TvRabs were isolated using 5′-RACE (MARATHON-RACE kit, Clontech) with two degenerate primers to the highly conserved GTP binding region WDTAGQE (5′-GGCGGCGGATCCTC[T/C]TGICCIGCIGT[A/G]TCCCA-3′ and 5′-GGCGGCGGATCCTC[T/C]TG[A/T]CCAGCNGTATCCCA-3′) [21]. Amplicons of the expected size were cloned into pGEM-T easy vector (Promega) followed by sequencing and BLAST analysis. Preliminary genomic sequence data was obtained from The Institute for Genomic Research through the website at http://www.tigr.org to complement cDNA derived sequences. Genome searches were conducted of 7.2x *T. vaginalis* genome coverage of sequence data assembled 01/04/05.

### 2.2. Master Rab protein alignment construction and phylogenetic analyses

DNA sequences were conceptually translated into proteins using BioEdit sequence alignment editor 5.0.9 [22]. Rab sequences were included from *T. brucei* [4], *C. elegans* [3], *Encephalitozoon cuniculi* [23], *Plasmodium falciparum* [24], *Entamoeba histolytica* [25], *Dictyostelium discoideum* [26], *Giardia lamblia* [27] and *Saccharomyces pombe* [3]. Additional Rab sequences for *Tetrahymena thermophila* were downloaded from TIGR http://www.tigr.org, respectively. GTPase sequences from several additional taxa were obtained from NCBI (http://www.ncbi.nlm.nih.gov/). Protein sequences were aligned using CLUSTALW [28] and further manual adjustments were made in SEA VIEW [29]. Positions that could not be unambiguously aligned were excluded from the phylogenetic analyses using the SEA VIEW masking option, reducing the alignment to 161 aligned positions for analysis.

The WAG evolutionary model was deemed best fit using Tree Puzzle (http://www.tree-puzzle.de/) and was used in Bayesian analysis with MrBayes V3.0B4 [30]. The Metropolis-coupled MCMC (MCMCMC) was used with four chains with estimated discrete gamma-distributed among-site

rate variation with four categories and fraction of invariant sites. Two million generations were run, every 2000 trees kept and the first 250 samples were discarded for calculation of the consensus trees, well within the region of the MCMCMC run where all parameters values stabilized. Each analysis was conducted at least twice. Putative *T. vaginalis* Rab protein sequences were initially compared to representatives of Rab, Ran, Ras and Rho/Rop from human, yeast and *Arabidopsis,* making up a total of 98 sequences. In the dataset used to compare a broader sampling of human, yeast, plant and *T. vaginalis* Rabs, 114 sequences were aligned. A more restricted sampling or Rab subfamilies focusing on Rab1, Rab5, Rab6, Rab7 and Rab11 from a more diverse range of taxa were also analysed, making up a dataset of 51 sequences.

### 2.3. Molecular modelling

Amino acid sequences from human Rab5a and members of the TvRabAl–6, TvRabCl–9, TvRabDl–8 and TvRab5a–d clades were extracted from the master alignment and all irrelevant gaps removed. Conserved regions of each *Trichomonas* clade were mapped onto a three-dimensional model of HsRab5a [31] (PDB accession: 1TU3, the closest Rab family member as defined by Blast score to TvRab5, TvRabC and TvRabD, that has been solved structurally) with the following colour scheme: red; fully conserved in all sequences, yellow; conserved in >50% of sequences. Non-conserved positions were designated white. The locations of protein–protein interactions were determined from co-ordinates from PDB for various Rab homologues and binding partners. Comparative models were constructed with the Swiss PDB-Viewer molecular graphics program [32], which was also used for manipulation and display.

## 3. Results

### 3.1. Identification of 65 T. vaginalis Rab genes

A total of 65 putative *T. vaginalis* Rab encoding genes (TvRabs) were identified by BLAST from one of three sources; an EST project (37 distinct sequences), 5′ rapid amplification of cDNA ends (RACE) cloning (eight distinct sequences) or searches of the *T. vaginalis* genome sequencing project at The Institute for Genomic Research (TIGR) (all genes corresponding to the partial EST and RACE clones plus 20 sequences distinct from cDNA derived entries) (Table S1). For all TvRab sequences we identify full-length ORF (Table S1).

Five sequence-dependent features were used to establish the Rab status of the putative *Trichomonas* proteins and their likely functionality. Firstly, all sequences recovered as top hits Rab proteins (BLASTP) and Rab position specific scoring matrices (PSSM) (RPS-BLAST [33]), clearly differentiating them from all other small GTPases PSSM (Table S1). Secondly, all but two putative TvRab sequences possess the distinctive Rab double-cysteine prenylation motif at the C terminus [8]. It is worth noting however, that the absence of a C-terminal prenylation motif has previously been observed, albeit rarely, in Rab proteins in other species [4,8]. Thirdly, previous analysis of small GTPases from *Saccharomyces cerevisiae* [10] and *Arabidopsis thaliana* [34] have recovered Rab proteins as monophyletic and we find all of the TvRab sequences to be monophyletic with reference Rab sequences (Fig. S1). Fourthly, 70% of TvRabs demonstrate over 50% identity to previously defined diagnostic Rab specific sequence motifs (RabF) [8] (Table S1). Finally, all of the TvRabs possess conserved GTPase functional motifs including the residues in the G-domains important for GTP/Mg++ binding and GTP hydrolysis (see Figs. 3 and S2 for a subset of

Table 1
Protozoans with amoeba forms show extraordinary Rab diversity

| Species | Number of Rab proteins encoded in genome | Cellularity | Amoeboid transformation | References |
|---|---|---|---|---|
| *Trichomonas vaginalis* | >65[a] | Unicellular | + | This study |
| *Entamoeba histolytica* | 105[b] | Unicellular | + | [26] |
| *Dictyostelium discoideum* | 54[b] | Unicellular[c] | + | [27] |
| *Tetrahymena thermophila* | ∼21 | Unicellular | – | TIGR |
| *Giardia lamblia* | 8 | Unicellular | – | [35] |
| *Trypanosoma brucei* | 16 | Unicellular | – | [4] |
| *Plasmodium falciparum* | 11 | Unicellular | – | [25] |
| *Saccharomyces cerevisiae* | 11 | Unicellular | – | [3] |
| *Schisosaccharomyces pombe* | 7 | Unicellular | – | [3] |
| *Caenorhabditis elegans* | 29 | Multicellular | +[d] | [3] |
| *Drosophila melanogaster* | 29 | Multicellular | +[d] | [3] |
| *Arabidopsis thaliana* | 57 | Multicellular | – | [36,38] |
| *Homo sapiens* | ∼60 | Multicellular | +[d] | [3] |

The number of Rab proteins encoded by the genomes of a selection of taxa throughout the eukaryotic tree of life are shown, together with the capacity to form amoeboid forms.

[a] GenBank accession numbers for the 65 *T. vaginalis* Rabs: AY896243–AY896292 and DQ019033–DQ019047.

[b] According to the *E. histolytica* and *D. discoideum* proteome annotations available at the NCBI protein database.

[c] A portion of the life cycle is multicellular.

[d] Some cell lineages only.

these). Further, 45 of the genes are transcribed, as determined from EST or RACE cloning. Taken together, these criteria are strong evidence that these *Trichomonas* GTPases are members of the Rab family and likely to be functional. Given that the *T. vaginalis* genome has not been fully sequenced and assembled, it is possible that the TvRab family contains even more members than identified here. With at least 65 members, the size of the *T. vaginalis* Rab repertoire is extraordinary for a unicellular organism (Table 1).

### 3.2. Identification of T. vaginalis orthologues of known Rab subfamilies

A large phylogenetic analysis was performed to explore the relationships between *T. vaginalis* Rab and a reference set of experimentally characterised Rab proteins from human, *S. cerevisiae* and *A. thaliana* (Fig. 1). These analyses excluded the most divergent *T. vaginalis* Rab, TvRabX6, which position within the Rab tree was unstable, probably due to its very long branch (Fig. S1). In addition, clustering of some of the TvRab sequences with subfamily members of the reference set allowed assignment of subfamily status for a minority of TvRabs (Fig. 1). Fourteen TvRab could be assigned to previously functionally characterized Rab subfamilies [8], including TvRab1 (three isoforms), TvRab5 (four isoforms), TvRab6 (two isoforms), TvRab7 (three isoforms) and TvRab11 (two isoforms). Phylogenies including broader subfamily taxa sampling and conservation of subfamily specific sequence motifs (RabSF) [8] further confirmed these assignments (Figs. 2 and 3). The posterior probabilities (PP) supporting the TvRab sequences clustering with the Rab5 clade are low in the large phylogenetic analyses but this relationship is consistently recovered in the different analyses (Figs. 1 and S1 and data not shown). In contrast, PP are maximal for this clade in the more restricted analyses focusing exclusively on the Rab1, 5, 6, 7 and 11 clades (Fig. 2), probably due to the removal of the many divergent sequences present in the broader analyses that are likely to cause long branch attraction artefacts (Figs. 1 and S1). Moreover, these TvRab conserve Rab5 specific motifs, highlighted in Figs. 3, 4 and S2 [37], strongly implicating them as new members of the Rab5 subfamily. Interestingly, members of the Rab7 subfamily possess a distinctive four residue insert located in the RabSF3 region, loop L7 [38], which is also shared by all three *T. vaginalis* sequences that clustered into the Rab7 clade (Fig. 3). The C-terminal xCxC prenylation motif is conserved between Rab6 subfamily members from human, yeast and *T. vaginalis,* consistent with assignment as authentic Rab6 subfamily members. Overall, these data allow the assignment of 21% of the TvRab sequences to previously described Rab subfamilies.

Pair-wise comparisons of aligned subfamily sequences showed that TvRab1a–c share on average 51% sequence identity and TvRab5a–d, TvRab7a–c, TvRab6a–b and TvRab11a–b, share average identities of 46%, 52%, 38% and 81% respectively. Therefore *T. vaginalis* Rab subfamily members exhibit greater sequence variation than *H. sapiens* Rab5a–c and *A. thaliana* RabAlA–I, which share 84% and 71% sequence identity respectively. *T. vaginalis* subfamily isoform sequence variation is also clearly visible in the differences in branch lengths in the phylogenetic trees (Figs. 1, 2 and S1). Overall, the considerable sequence divergence exhibited by TvRab subfamily members suggests similar, but non-redundant functionality. Significantly, we identity *T. vaginalis* orthologues from five Rab subfamilies that are common to all eukaryotes corresponding to the functional classes I, II, V, VI and VII defined by Pereira-Leal and Seabra [3].

### 3.3. Identification and characterisation of T. vaginalis specific Rab subfamilies

A majority of the *T. vaginalis* Rab sequences exhibited no clear relationship to previously known Rab sequences, including Rabs from divergent unicellular organisms, i.e. *T. brucei, P. falciparum, Encephalitozoon cunicli, G. lamblia, Neurospora crassa, E. histolytica* or *D. discoideum* (data not shown).

Approximately half of the TvRab sequences fell into six *T. vaginalis* specific Rab clades. Individual sequences within these clades have no orthologues in other species and hence these were named TvRabA–F (Figs. 1 and S1). Individual *T. vaginalis* specific Rab clades were consistently recovered in all analyses and with high PP support, confirming their robustness (Figs. 1 and S1). These sequences are *bona fide* Rabs as defined by the criteria above, and hence *T. vaginalis* encodes a considerable number of novel Rab proteins.

Similar to the *T. vaginalis* Rabs that correspond to previously known Rab isoforms, the *T. vaginalis* specific Rab subfamilies are also present as multiple isoforms. Subfamily TvRabA includes six isoforms, subfamilies TvRabB, TvRabC, TvRabD, TvRabE and TvRabF have four, nine, eight, two and three isoforms, respectively. Pair-wise comparison of sequences within TvRab subfamilies demonstrated that the TvRabB and TvRabF subfamily members show the most intra-subfamily mean diversity with 39% sequence identity, whilst TvRabA (49%), TvRabC (53%), TvRabD (58%) and TvRabE (58%) subfamilies show less variation. Therefore, *T. vaginalis* intra-subfamily sequence variation is within the range exhibited by other subfamilies, both in *T. vaginalis* and in other species. As expected the TvRabs showed less mean sequence identity between the subfamilies (34%) than within them (50%). Sequence variability is graphically visible in the variation in branch lengths in the phylogenetic analysis (Figs. 1 and S1).

To gain additional insight into the functionality of the most abundant *Trichomonas* specific subfamilies (TvRabA, C and D), areas of sequence conservation shared by these subfamily members were mapped on to the surface of human Rab5a protein structural data bound to its effector rabaptin5 [31] (Fig. 4). Rab5 are the most similar Rab to TvRabD and TvRabC sequences and they share two motifs in the Switch

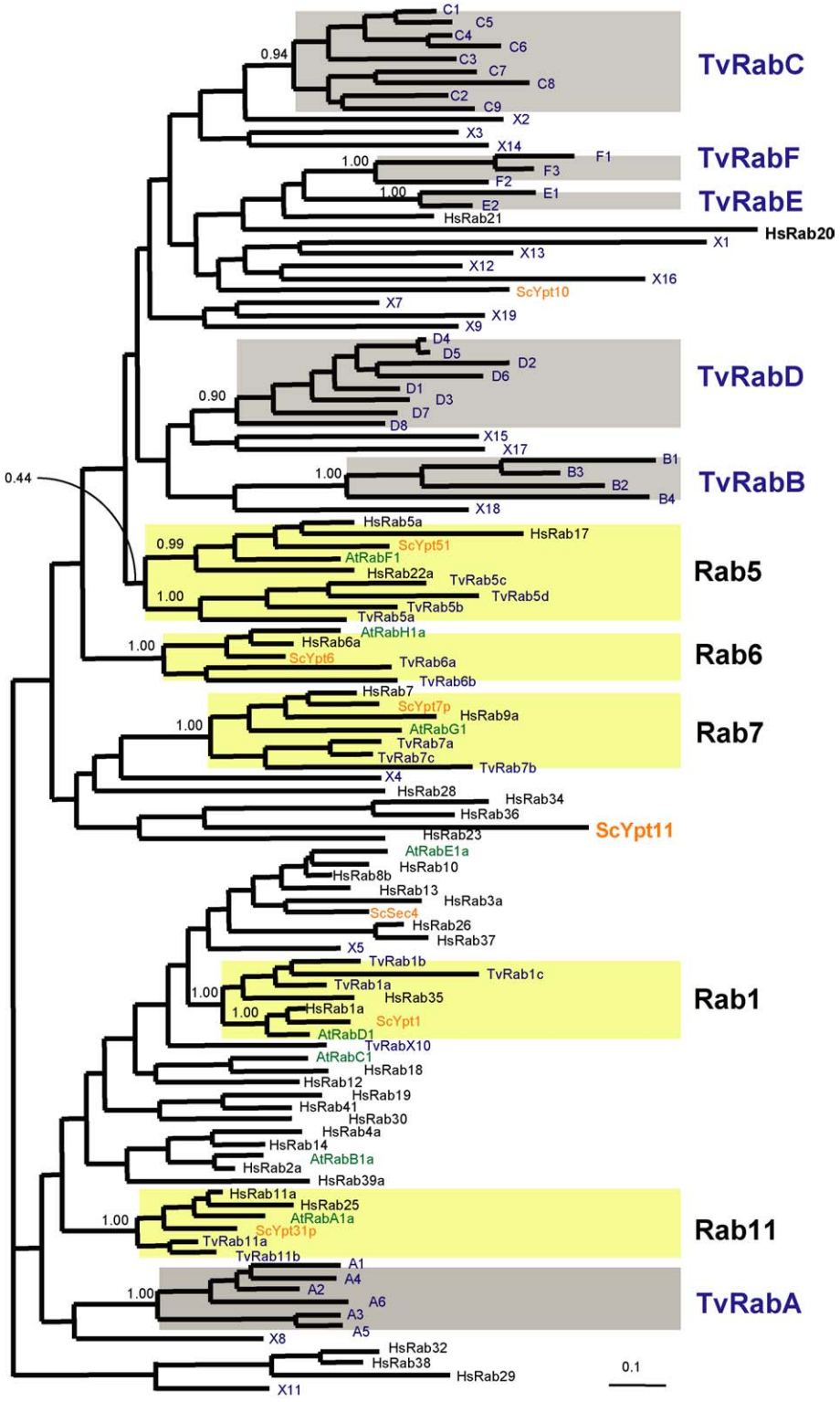


Fig. 1. Phylogenetic relationships between *Trichomonas,* human, yeast, and plant Rabs. The phylogenetic relationships between *Trichomonas vaginalis* (blue) and a selection of representative Rab from *Homo sapiens* (black), *Saccharomyces cerevisiae* (orange) and *Arabidopsis thaliana* (green) were investigated. The most divergent *T. vaginalis* sequence, TvRabX6, was not included here because of its instability within the Rab clade in all phylogenetic analyses (see Fig. S1). The shown phylogeny is a Bayesian consensus tree (see Section 2 for details of method used). Known Rab subfamily clades with broad taxonomic distributions are indicated (Rab1, 5, 6, 7 and 11, yellow shading) as are *Trichomonas* specific clades (TvRabA–F, grey shading). All, but one, of these clades are supported by very high posterior probabilities (PP) ranging between 0.94 and 1.00. The Rab5 clade is only weakly supported with a PP of 0.44. Branch lengths are a good reflection of sequence divergence. The highly divergent human Rab20 and yeast Yptl 1 are highlighted by larger font. Notably all *Trichomonas* Rab fall within the range of sequence diversity of other Rab proteins, see also text. The scale bar represents 10% sequence divergence.

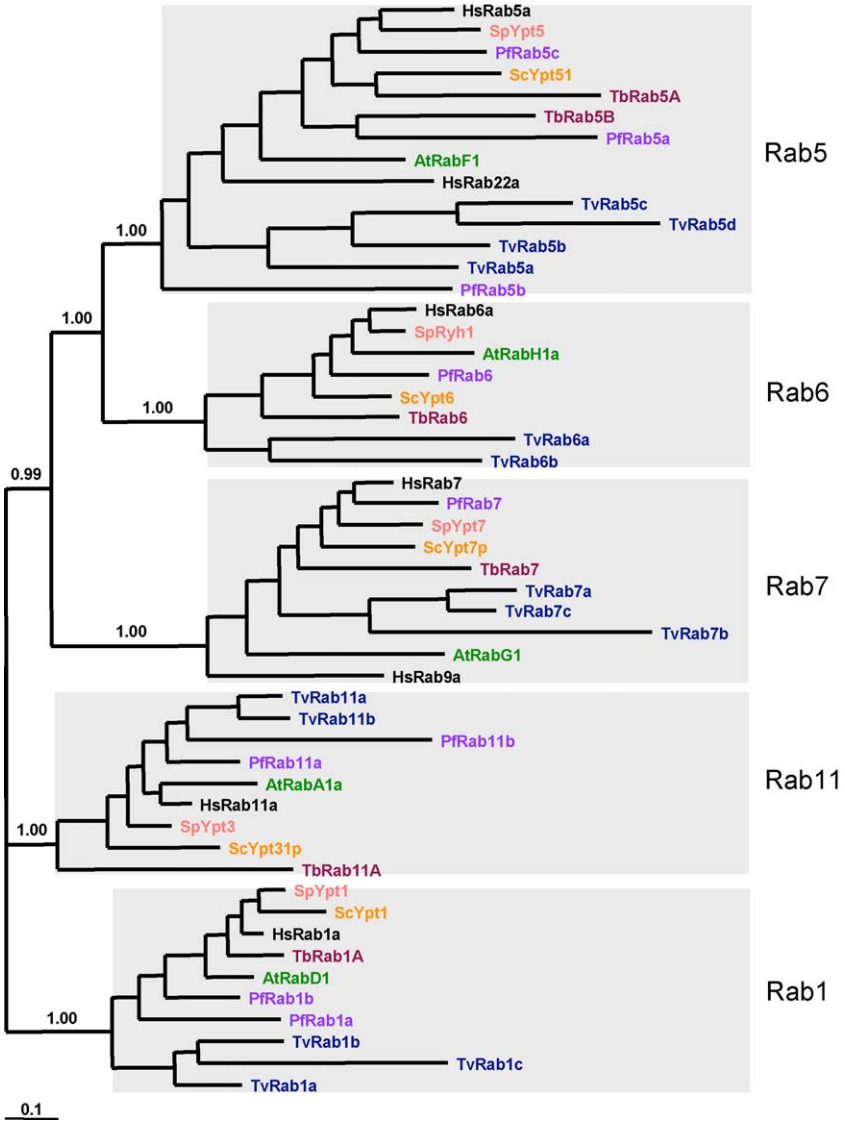


Fig. 2. Phylogeny of *T. vaginalis* Rab1, Rab5, Rab6, Rab11 and Rab7 subfamilies. A subset of Rab protein sequences were analysed from *Trichomonas vaginalis* (blue), *Homo sapiens* (black), *Saccharomyces cerevisiae* (orange), *Arabidopsis thaliana* (green), *Trypanosoma brucei* (burgundy), *Plasmodium falciparum* (purple) and *Schizosaccharomyces pombe* (pink). The Bayesian consensus tree confirmed the subfamily assignment of the TvRab proteins (see Fig. 1 and S1) and all are supported by maximum PP. The scale bar represents 10% sequence divergence.

I and II regions, TIGAAF and LAPM (Figs. 3, 4 and S2). The later is considered as a Rab5 specific sequence feature [37]. Hence, we also used TvRab5a–d sequences for comparisons. The face corresponding to the Rab5 effector rabaptin5-binding site, which greatly overlaps with the GDI binding site [31], shows the least variation within subfamily members (indicated in red and yellow in Fig. 4). In contrast, the opposite region of this face shows the least amount of conservation within subfamilies and also the largest differences in the position of the few conserved patches (data not shown). The polarity profiles on modelled individual sequences also indicate that the TvRab5 sequences are the most similar to human Rab5a (Fig. 4), as expected from their phylogenetic positions (Figs. 1 and 2). In contrast, TvRabA sequences show least similarity with in particular a negatively charged patch where TvRabC, TvRabD and TvRab5 subfamilies show an overall less polar face surrounded by three positively charged "hot spots" shared between TvbRab5 and HsRab5 (Fig. 4). Thus, this suggests that functional selection applies to similar regions of these proteins overall, but subfamily specific functional pressures (leading to different evolutionary patterns) also apply, consistent with different functionality.

### 3.4. Identification of 19 *T. vaginalis* specific orphan Rabs

A further 19 *T. vaginalis* Rab sequences did not cluster into subfamilies with either previously defined Rab clades or the TvRabA–F Rab clades (Fig. 1). Whilst retaining Rab defining

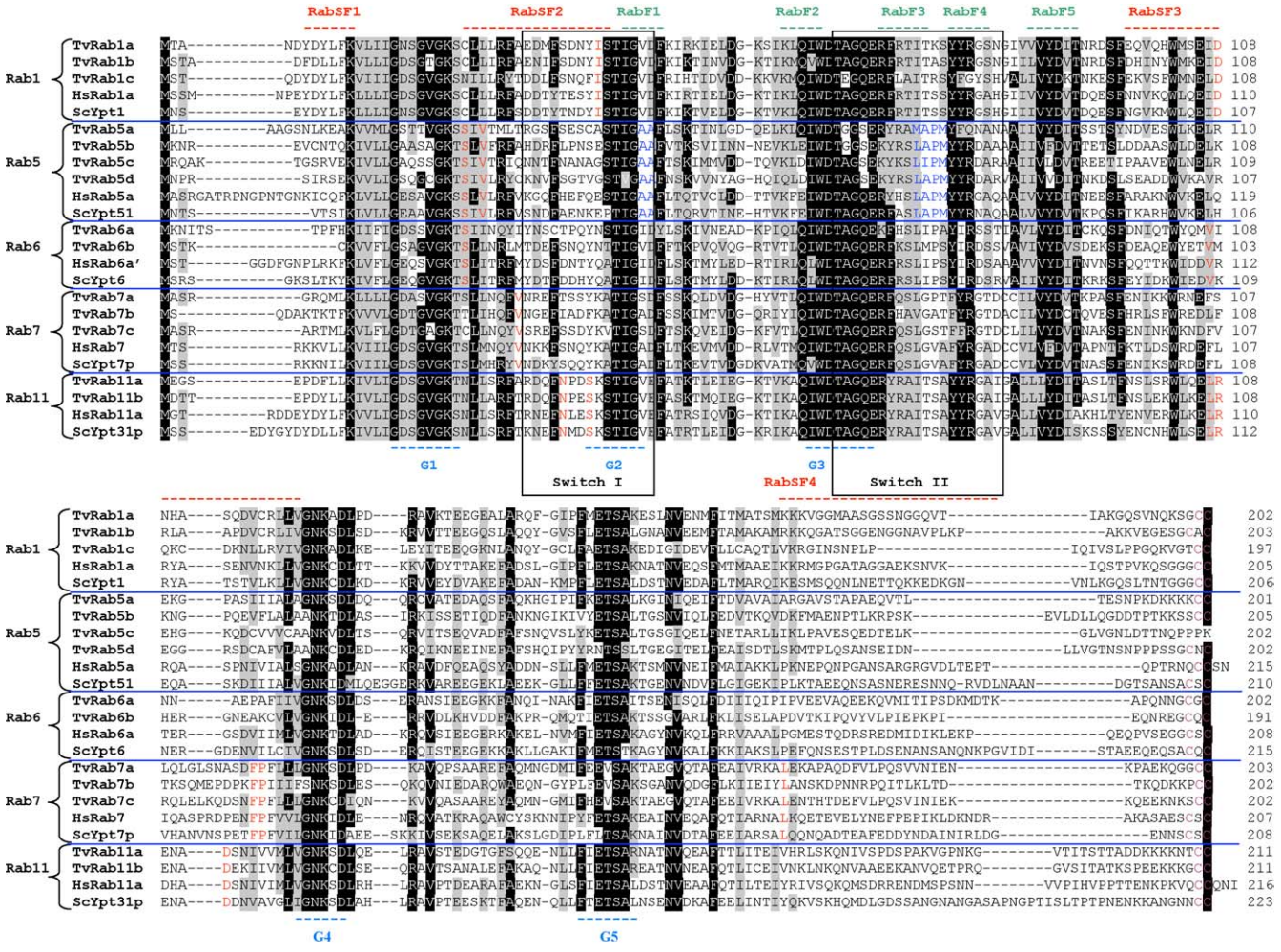


Fig. 3. Sequence comparison of *Trichomonas* Rabs with subfamily orthologues from other taxa. Comparison of sequences from *T. vaginalis* (TvRab), *Saccharomyces cerevisiae* (Ypts) and *Homo sapiens* (HsRab) of Rab subfamilies Rab1, 5, 6, 7 and 11 (bracketed and divided by a blue line). G-domains (G1-G5 shown in blue) involved in GTP/GDP/Mg++ binding are indicated below the alignment. The two conformational switch (I and II) regions [39] are boxed. All shown *Trichomonas* Rabs retain C-terminal double cysteines important to prenylation, highlighted in purple. Rab specific regions (RabF shown in green, defined in [8]) are conserved in *Trichomonas* Rab. Rab subfamily specific regions (RabSF shown in red, defined in [8]) show conservation of specific motifs highlighted in colour. Rab5 motifs highlighted in blue [37]. Residues identical in 60% or more aligned sequences are highlighted black whilst similar residues (conservative changes) are shown in grey.

features (prenylation motifs, clustering with Rabs in phylogenies and possessing conserved RabF motifs) these TvRab 'orphans' are more divergent than the core or TvRabA–F subfamily members, sharing only 28% average sequence identity in pair-wise comparisons. Hence there is a further large set of TvRab GTPases with considerable sequence variation.

## 4. Discussion

In this report are described the identification of a very large number of Rab genes from the unicellular *T. vaginalis*; the size and diversity of the Rab repertoire in this organism is extraordinary and unexpected (Table 1). These data provide a major challenge to the paradigm that Rab gene diversity is simply a result of tissue complexity [1] and suggest the presence of additional evolutionary drivers underlying *Trichomonas* Rab family expansion.

Several lines of evidence suggest that the majority, if not all, of the identified TvRabs are likely to be functional and regulate membrane trafficking. As 45 TvRabs were identified via both mRNA and genome approaches (Table S1) they are clearly not a feature of library construction, i.e. arising through strain variation, recombination or other means of increasing sequence diversity in an artefactual manner. Moreover, the diversity between the TvRab protein sequences indicates they are encoded by distinct genes and are not minor sequence variants (Fig. 1). The identification of complete ORFs for all of the TvRabs described here, combined with their conserved sequence features (G-blocks and prenylation sites) also suggests these are unlikely to be pseudogenes. With the exception of TvRabXl, TvRabX6 and TvRabX16 (Fig. S1), all the TvRabs are more similar to the Rab dataset

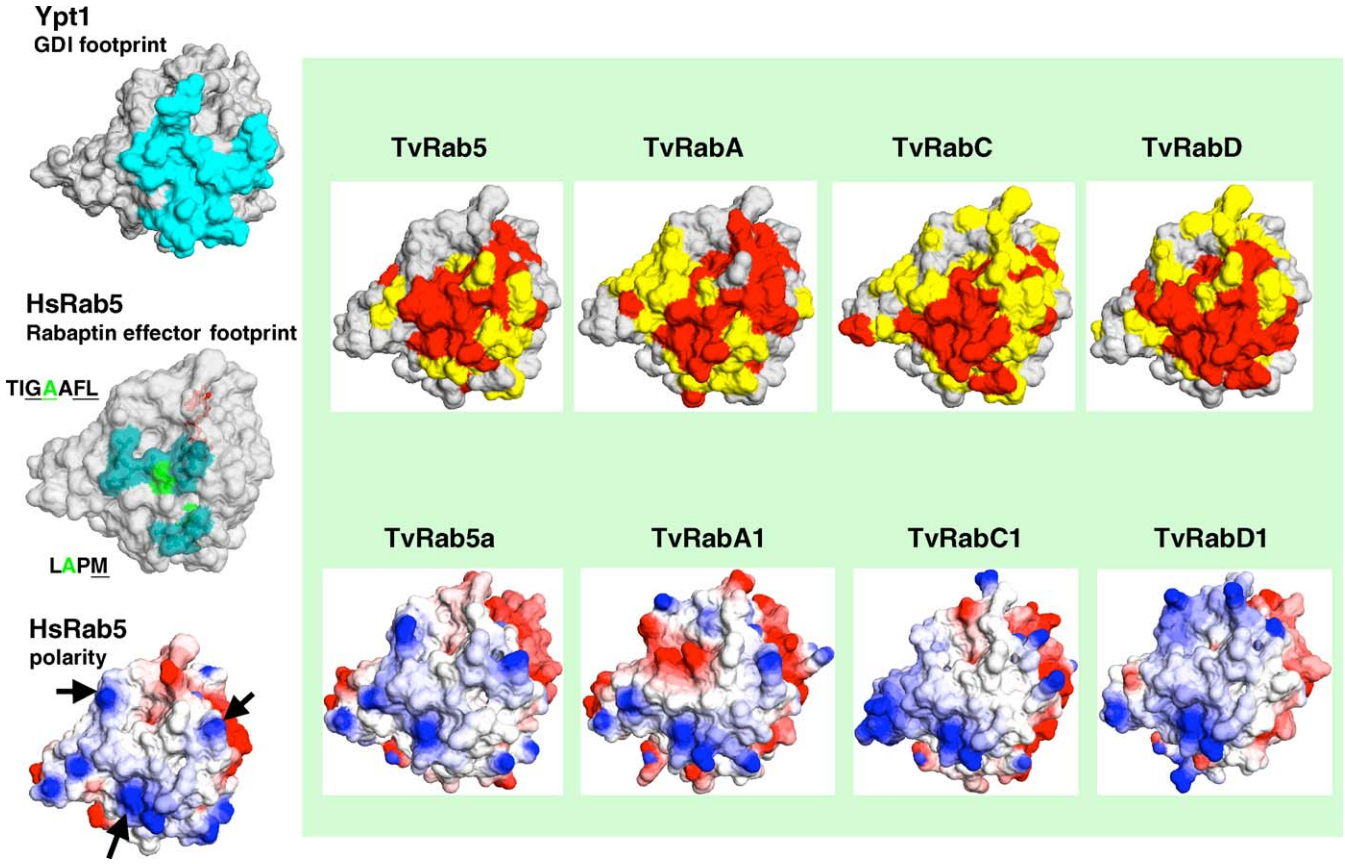


Fig. 4. Surface conservation of TvRab5, TvRabA, C and D clades. Each panel shows a Rab structure or comparative model. We choose a "front view" showing most information about Rab partner binding surfaces as defined by the human Rab5–rabaptin5 interaction [31]. A back view would be rotated relative to this by 180° about a vertical axis (see text). The left column provides context, showing (i) the *S. Cerevisiae* GDI binding surface on Yptl (top, PDB accession 1UKV) that greatly overlap with the GAP binding surface (HsRhoA PDB 10W3, not shown); (ii) position of the TIGAAFL and LAPM motifs in human Rab5a (middle, PDB accession 1TU3) mapping both switch I/II and effector binding sites (Figs. 3 and S2). Underlined residues indicate five (among 11) that binds rabaptin5 [31] (see Fig. S2, Table S2 and text). The motifs are coloured cyan, except for two alanines (the one in LAPM is not well exposed to the surface) that are discussed further in the text (green). Red denotes the GTP binding site within this transparent surface; (iii) also in the left hand column, the surface polarity for human Rab5a is displayed (bottom, 1TU3). All surface polarity plots are made with contours/colours ranging from −3 kT/e (red) to +3 kT/e (blue), with white indicating relatively non-polar regions; kT is thermal energy and e is unit charge. In HsRab5a, TvRab5, TvRabC and TvRabD, the effector binding region is overall less polar (surrounded by three positive hot spots, arrows in HsRab5a, well conserved in TvRab5a) then the other parts of this face. The corresponding region in TvRabA is characterised by a large negative patch absent in the other sequences. Each of the four right hand columns (all within the green background) pertains to the noted TvRab clades. Each column contains one plots of sequence conservation within a clade, and one plot of surface polarity for a comparative model of a representative from within that clade, all of these correspond to the orientation shown for HsRab5a on the left column. For the conservation plots, 100% identity within a TvRab clade is red, 50% identity is yellow and grey otherwise.

than two of the most divergent members of the Rab family, ScYptl 1 and HsRab20, indicating that the TvRab sequence diversity is contained within known functional Rab diversity.

Rab subfamily orthologues from diverse taxa share similar intracellular locations and functions [1,2]. Eukaryotes share a core set of Rab subfamily proteins responsible for coordination of exocytosis and endocytosis and *T. vaginalis* also performs these functions. The identification of *Trichomonas* orthologues of Rab1, Rab5, Rab6, Rab7 and Rab11 indicates conservation of one important aspect of the molecular machinery regulating exocytotic and endocytic pathways. Retention of subfamily sequence motifs over the large evolutionary distance between *T. vaginalis,* yeast and humans further extends published comparative studies [3] identifying residues of likely functional importance as effector and/or regulator binding motifs (Figs. 3, 4 and S2). Particularly well conserved are the Rab5 specific alanine in Switch I (TIGAAFL motif) and Switch II regions LAPM (Figs. 3 and S2). These sites corresponds to two conformational determinants [39] modulating the structural difference between Rab subfamilies, of which the former is conserved among Rab5 (A only) and other endocytic Rab (A/S), whereas exocytic Rab possess instead valine or isoleucine [39] (see Rab5 and 7 versus Rab1 and 11 in Fig. 3). This also suggests that TvRabC and TvRabD clade members are involved in endocytosis, whereas TvRabA sequences (V or I) are involved in exocytosis (Fig. S2).

The *T. vaginalis* specific Rab subfamilies (A–F) cluster into separate clades that are distinct from those defined by Rab sequences from other taxa (Fig. 1). Distinct sequence

patches of conservation within each clade corresponding to contiguous regions on the surface of the predicted proteins in subfamilies TvRab5, TvRabA, TvRabC and TvRabD can be observed by mapping conserved motifs onto a human Rab5a molecular structure (Figs. 4 and S2). The overlapping GAP and GDI binding regions share areas of high conservation between subfamilies, suggesting similar regulation mechanisms for these TvRabs (Figs. 4, S2). In contrast, the sites corresponding to effector binding regions have important differences between the four TvRab clades, with in particular a batch of negatively charged residues in the TvRabA clade that is absent in the three other clades (Figs. 4 and S2). Differences between the TvRab sequences corresponding to human Rab5a-rabaptin5 interactions sites (11 residues mainly in switch I/II region) [31] (Table S2, Fig. S2), or Rab3a-rabphilin-3A interactions sites (27 residues with 60% of them outside the switch domains) [40] (Fig. S2), suggest that these four *Trichomonas* clades are likely to bind different sets of clade specific effector proteins.

What are the evolutionary drivers for Rab gene family expansion, and in *T. vaginalis* in particular? It is possible that the Rab gene family expansion correlates with intracellular membrane complexity. Indeed, *T. thermophila,* a unicellular ciliate, demonstrates complex endomembrane organisation and has a comparably large number of Rabs (∼21) for a unicellular organism (Table 1), although not on the scale of *T. vaginalis.* Membrane trafficking complexity may partly explain TvRab gene proliferation and it will be interesting to see if further genome studies reveal corresponding gene expansion of complicit proteins involved in orchestrating membrane trafficking. Trafficking complexity may be apparent in distinct life cycle stages but little is known of the *T. vaginalis* life cycle [16]. However, *T. vaginalis* has only one host (human) and is unlikely to equal the number of morphologies achieved by *Plasmodium falciparum,* which has only 11 Rab genes [24].

Interestingly, the unicellular organisms *E. histolytica* and *D. discoideum* also have large Rab gene families, with at least 105 and 54 Rab genes, respectively [25,26] (Table 1). As for the lineage-specific Rab gene family expansions observed in animals versus plants [3,36], *Trichomonas* has expanded its Rab repertoire independently from the two amoebozoa, to which it is unlikely to be closely related [14]. Amoebozoa and *T. vaginalis* are phagocytic but phagocytosis *per se* does not require Rab gene amplification as the phagocytic protozoan *G. lamblia* [41] only has 8 Rabs and the number of metazoan Rabs dedicated to phagocytic pathways is limited. Although these three protozoa occupy drastically different ecological niches, they all have the capacity to form amoebae [42,43]. This intriguing parallel may point to a role for a large Rab gene family in cellular transformation to amoebae. Indeed, one *Entamoeba* Rab was recently shown to be involved in amoeboid movement by localizing to the leading edge of motile amoeboid cells [42]. It will also be interesting to compare the expression patterns TvRab genes between the trophozoites and the amoeba form bound to host tissue, possibly involving the TvRabA and/or TvRabB in regulated secretion of pathogenic factors upon host tissue [e.g. 17]. Since at least one Rab protein was shown to directly interact with a endocytosed protein cargo, Rab3a with the polymeric Ig receptor [44], and that we observe a large set of Rab5-like sequences (TvRab5a–d, TvRabCl-9 and TvRabDl-8), it would also be interesting to investigate if *Trichomonas* genome encodes a large repertoire of surface proteins that could be binding different sets of the broad TvRab5-like set. To conclude, it is likely that combinations of several evolutionary drivers are responsible for the large *T. vaginalis* Rab gene family. Clearly, comparative genomics and cell biology of *Trichomonas* and other protozoa should give new and fascinating insights into Rab evolution and function and the molecular basis of membrane trafficking. Furthermore, the *Trichomonas* specific sets of Rab clades identified here, clearly distinct from its host Rab repertoire, could also lead to the identification of interesting drug targets [45] to be used to disrupt specific membrane trafficking steps to control the parasite.

## Note added in proof

A related paper describing detailed analyses of 91 Rab proteins from *Entamoeba histolytica* was published. Saito-Nakano Y, Loftus BJ, Hall N, Nozaki T. The diversity of Rab GTPases in Entamoeba histolytica. Exp Parasitol 2005;10:244–52.

## Acknowledgments

This research was supported by a Wellcome Trust University award to RPH. MCF's laboratory is supported by Wellcome Trust project and program grants. We are grateful to Phillip Carter (Imperial College, London) for initial 3D analysis of TvRab sequences and Martin Embley for comments. Sequencing of the *Trichomonas vaginalis* genome project at TIGR is being accomplished with support from National Institutes of Allergy and Infectious Diseases, National Institutes of Health. Preliminary *Tetrahymena thermophilia* sequence data was obtained from The Institute for Genomic Research website at http://www.tigr.org.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2005.06.008.

## References

[1] Stenmark H, Olkkonen VM. The Rab GTPase family. Genome Biol 2001:2.

[2] Zerial M, McBride H. Rab proteins as membrane organizers. Nat Rev Mol Cell Biol 2001;2:107–17.
[3] Pereira-Leal JB, Seabra MC. Evolution of the Rab family of small GTP-binding proteins. J Mol Biol 2001;313:889–901.
[4] Ackers JP, Dhir V, Field MC. A bioinformatic analysis of the RAB genes of *Trypanosoma brucei*. Mol Biochem Parasitol 2005;141:89–97.
[5] Oishi H, Sasaki T, Nagano F, et al. Localization of the Rab3 small G protein regulators in nerve terminals and their involvement in $Ca^{2+}$-dependent exocytosis. J Biol Chem 1998;273: 34580–5.
[6] Segev N. Ypt and Rab GTPases: insight into functions through novel interactions. Curr Opin Cell Biol 2001;13:500–11.
[7] Sprang SR. G protein mechanisms: insights from structural analysis. Annu Rev Biochem 1997;66:639–78.
[8] 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.
[9] 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.
[10] Garcia-Ranea JA, Valencia A. Distribution and functional diversification of the ras superfamily in *Saccharomyces cerevisiae*. FEBS Lett 1998;434:219–25.
[11] Pereira-Leal JB, Strom M, Godfrey RF, Seabra MC. Structural determinants of Rab and Rab Escort Protein interaction: Rab family motifs define a conserved binding surface. Biochem Biophys Res Commun 2003;301:92–7.
[12] 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.
[13] Lehker MW, Alderete JF. Biology of trichomonosis. Curr Opin Infect Dis 2000;13:37–45.
[14] Simpson AGB, Roger AJ. Excavata and the origin of amitochondriate eukaryotes. In: Hirt RP, Horner DS, editors. Organelles, genomes and eukaryote phylogeny: an evolutionary synthesis in the age of genomics. Boca Raton: CRC Press; 2004. p. 27–53.
[15] Hrdy I, Hirt RP, Dolezal P, et al. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature 2004;432:618–22.
[16] Petrin D, Delgaty K, Bhatt R, Garber G, Clinical. microbiological aspects of *Trichomonas vaginalis*. Clin Microbiol Rev 1998;11:300–17.
[17] Fiori PL, Rappelli P, Addis MF. The flagellated parasite *Trichomonas vaginalis:* new insights into cytopathogenicity mechanisms. Microbes Infect 1999;1:149–56.
[18] Rendon-Maldonado J, Espinosa-Cantellano M, Soler C, Torres JV, Martinez-Pal omo A. *Trichomonas vaginalis:* in vitro attachment and internalization of HIV-1 and HIV-1-infected lymphocytes. J Eukaryot Microbiol 2003;50:43–8.
[19] Lyons EJ, Carlton JM. Mind the gap: bridging the divide between clinical and molecular studies of the trichomonads. Trends Parasitol 2004;20:204–7.
[20] Hirt RP, Lal K, Pinxteren J, et al. Biochemical and genetic evidence for a family of heterotrimeric G-proteins in *Trichomonas vaginalis*. Mol Biochem Parasitol 2003;129:179–89.
[21] Chavrier P, Simons K, Zerial M. The complexity of the Rab and Rho GTP-binding protein subfamilies revealed by a PCR cloning approach. Gene 1992;112:261–4.
[22] Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999;41:95–8.
[23] Katinka MD, Duprat S, Cornillot E, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 2001;414:450–3.
[24] Quevillon E, Spielmann T, Brahimi K, Chattopadhyay D, Yeramian E, Langsley G. The *Plasmodium falciparum* family of Rab GTPases. Gene 2003;306:13–25.
[25] Loftus B, Anderson I, Davies R, et al. The genome of the protist parasite *Entamoeba histolytica*. Nature 2005;433:865–8.
[26] Eichinger L, Pachebat JA, Glockner G, et al. The genome of the social amoeba *Dictyostelium discoideum*. Nature 2005;435:43–57.
[27] Langford TD, Silberman JD, Weiland ME, et al. Giardia lamblia: identification and characterization of Rab and GDI proteins in a genome survey of the ER to Golgi endomembrane system. Exp Parasitol 2002;101:13–24.
[28] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–80.
[29] Galtier N, Gouy M, Gautier C, SeaView. Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 1996;12:543–8.
[30] Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 2001;17:754–5.
[31] Zhu G, Zhai P, Liu J, Terzyan S, Li G, Zhang XC. Structural basis of Rab5-Rabaptin5 interaction in endocytosis. Nat Struct Mol Biol 2004;11:975–83.
[32] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 1997;18:2714–23.
[33] Marchler-Bauer A, Anderson JB, DeWeese-Scott C, et al. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 2003;31:383–7.
[34] Vernoud V, Horton AC, Yang Z, Nielsen E. Analysis of the small GTPase gene superfamily of *Arabidopsis*. Plant Physiol 2003;131:1191–208.
[35] Marti M, Regos A, Li Y, et al. An ancestral secretory apparatus in the protozoan parasite Giardia intestinalis. J Biol Chem 2003;278:24837–48.
[36] Rutherford S, Moore I. The *Arabidopsis* Rab GTPase family: another enigma variation. Curr Opin Plant Biol 2002;5:518–28.
[37] Collins RN, Brennwald P. Rab. In: Hall A, editor. Frontiers in molecular biology: GTPases, vol. 24. Oxford: Oxford University Press; 2000. p. 137–75.
[38] Constantinescu AT, Rak A, Alexandrov K, Esters H, Goody RS, Scheidig AJ. Rab-subfamily-specific regions of Ypt7p are structurally different from other RabGTPases. Structure (Camb) 2002;10:569–79.
[39] Merithew E, Hatherly S, Dumas JJ, Lawe DC, Heller-Harrison R, Lambright DG. Structural plasticity of an invariant hydrophobic triad in the switch regions of Rab GTPases is a determinant of effector recognition. J Biol Chem 2001;276:13982–8.
[40] Ostermeier C, Brunger AT. Structural basis of Rab effector specificity: Crystal structure of the small G protein Rab3A complexed with the effector domain of Rabphilin-3A. Cell 1999;96:363–74.
[41] Sogayar MI, Gregorio EA. Uptake of bacteria by trophozoites of *Giardia duodenalis* (Say). Ann Trop Med Parasitol 1989;83: 63–6.
[42] Welter BH, Temesvari LA. A unique Rab GTPase, EhRabA, of *Entamoeba histolytica,* localizes to the leading edge of motile cells. Mol Biochem Parasitol 2004;135:185–95.
[43] Furtado MB, Benchimol M. Observation of membrane fusion on the interaction of *Trichomonas vaginalis* with human vaginal epithelial cells. Parasitol Res 1998;84:213–20.
[44] van ISC, Tuvim MJ, Weimbs T, Dickey BF, Mostov KE. Direct interaction between Rab3b and the polymeric immunoglobulin receptor controls ligand-stimulated transcytosis in epithelial cells. Dev Cell 2002;2:219–28.
[45] Stein MP, Dong J, Wandinger-Ness A. Rab proteins and endocytic trafficking: potential targets for therapeutic intervention. Adv Drug Deliv Rev 2003;55:1421–37.