doi:10.1006/jmbi.2001.5072 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **313**, 889–901

JMB

Evolution of the Rab Family of Small GTP-binding Proteins

José B. Pereira-Leal and Miguel C. Seabra*

Cell and Molecular Biology Section, Division of Biomedical Sciences, Faculty of Medicine Imperial College, London SW7 2AZ, UK

Rab proteins are small GTP-binding proteins that form the largest family within the Ras superfamily. Rab proteins regulate vesicular trafficking pathways, behaving as membrane-associated molecular switches. Here, we have identified the complete Rab families in the *Caenorhabditis elegans* (29 members), *Drosophila melanogaster* (29), *Homo sapiens* (60) and *Arabidopsis thaliana* (57), and we defined criteria for annotation of this protein family in each organism. We studied sequence conservation patterns and observed that the RabF motifs and the RabSF regions previously described in mammalian Rabs are conserved across species. This is consistent with conserved recognition mechanisms by general regulators and specific effectors. We used phylogenetic analysis and other approaches to reconstruct the multiplication of the Rab family and observed that this family shows a strict phylogeny of function as opposed to a phylogeny of species. Furthermore, we observed that Rabs co-segregating in phylogenetic trees show a pattern of similar cellular localisation and/or function. Therefore, animal and fungi Rab proteins can be grouped in ''Rab functional groups'' according to their segregating patterns in phylogenetic trees. These functional groups reflect similarity of sequence, localisation and/or function, and may also represent shared ancestry. Rab functional groups can help the understanding of the functional evolution of the Rab family in particular and vesicular transport in general, and may be used to predict general functions for novel Rab sequences.

*Corresponding author

Keywords: Rab proteins; GTPases; evolution; annotation; identification

Introduction

The recent availability of substantially completed genome sequences for several eukaryotic organisms creates new opportunities for the study of protein evolution and function. At present, the nematode (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), the budding yeast (*Saccharomyces cerevisiae*) and the fission yeast (*Schizosaccharomyces pombe*) have had their genome sequenced, and the first drafts of the complete genome of *Homo sapiens* and *Arabidopsis thaliana* were recently released. With six complete or nearly complete genomes of evolutionary distant organisms, it is now possible to start addressing the evolution of primary structure and function in the Rab protein family.

Rab proteins form the largest family of the Ras superfamily of small GTP-binding proteins and regulate intracellular trafficking pathways. More than 50 Rab proteins have been described in mammalian cells, each with a specific subcellular localisation and many with specific patterns of tissue distribution.[1–3] Rabs behave as membrane-associated molecular switches to regulate budding, transport and fusion reactions in vesicular transport.

In a previous study, we analysed sequence conservation in the mammalian Rab family[4] and observed the existence of mammalian Rab-specific motifs (RabF motifs) that clustered in and around the switch regions. This allowed us to propose criteria for Rab family classification, and to identify novel Rab sequences from the databases. We also suggested that Rab proteins use the switch regions[5,6] in addition to other regions to determine specificity of binding to protein partners, unlike Ras proteins, where specificity of binding is determined mainly by the switch regions.[7–9] These

Abbreviations used: RabF, Rab family; RabSF, Rab subfamily; pHMM, profile Hidden Markov Model; REP, Rab escort protein; Rab GDI, Rab GDP dissociation inhibitor; CTL, cytotoxic T-lymphocytes; ER, endoplasmic reticulum.

E-mail address of the corresponding author: m.seabra@ic.ac.uk

0022-2836/01/040889–13 $35.00/0

specificity-determining regions were named Rab subfamily regions (RabSF).[4,10,11]

In the present study, we identify and annotate the complete Rab family in *H. sapiens, D. melanogaster, C. elegans and A. thaliana*, and use this dataset, complemented with the complete Rab families in *S. pombe* and *S. cerevisiae*, to study their evolution. We test the hypothesis that there is a conserved mechanism of Rab interaction with regulators and effectors across evolution, and we attempt to reconstruct the multiplication of Rab proteins. This analysis suggested the existence of a higher-order hierarchy in the Rab family with implications for the function and evolution of these proteins.

Results and Discussion

Identification and annotation of complete Rab families

Previous studies have identified the Rab families in the budding and fission yeast[12–14] (shown in Tables S1 and S2 of the Supplementary Material). We first attempted to identify the complete Rab families in the human, nematode, fly, and Arabidopsis genomes. We searched the public databases with pHMM described in a previous study.[4] This criteria considers conservation of GTP-binding motifs, presence of double-cysteine prenylation motifs, and conservation of the RabF motifs. The fact that all the budding yeast Rab (Ypt/Sec4) proteins were correctly identified validated our method.

In the *C. elegans* genome, we identified 29 independent open reading frames that conform to our criteria (Tables 1, S3 and Figure 1).[4] Comparison of each sequence with pHMM describing other small GTPase families indicated that they were clearly not members of any other Ras-like small GTPase family. Using the same criteria, we identified 29 independent open reading frames in the *D. melanogaster* genome that we consider Rabs (Tables 1 and S4, Figure 1). Our strategy led to identification of more Rab sequences than two previous attempts,[14,15] suggesting that our analysis was more thorough and/or a recent improvement in the databases.

In *A. thaliana*, we identified 56 proteins that we consider *bona fide* Rabs (Tables 1 and S5, Figure 2). One additional protein, named Ara6 (accession BAB32953), exhibits some peculiar features. It possesses putative N-terminal myristoylation and pal-

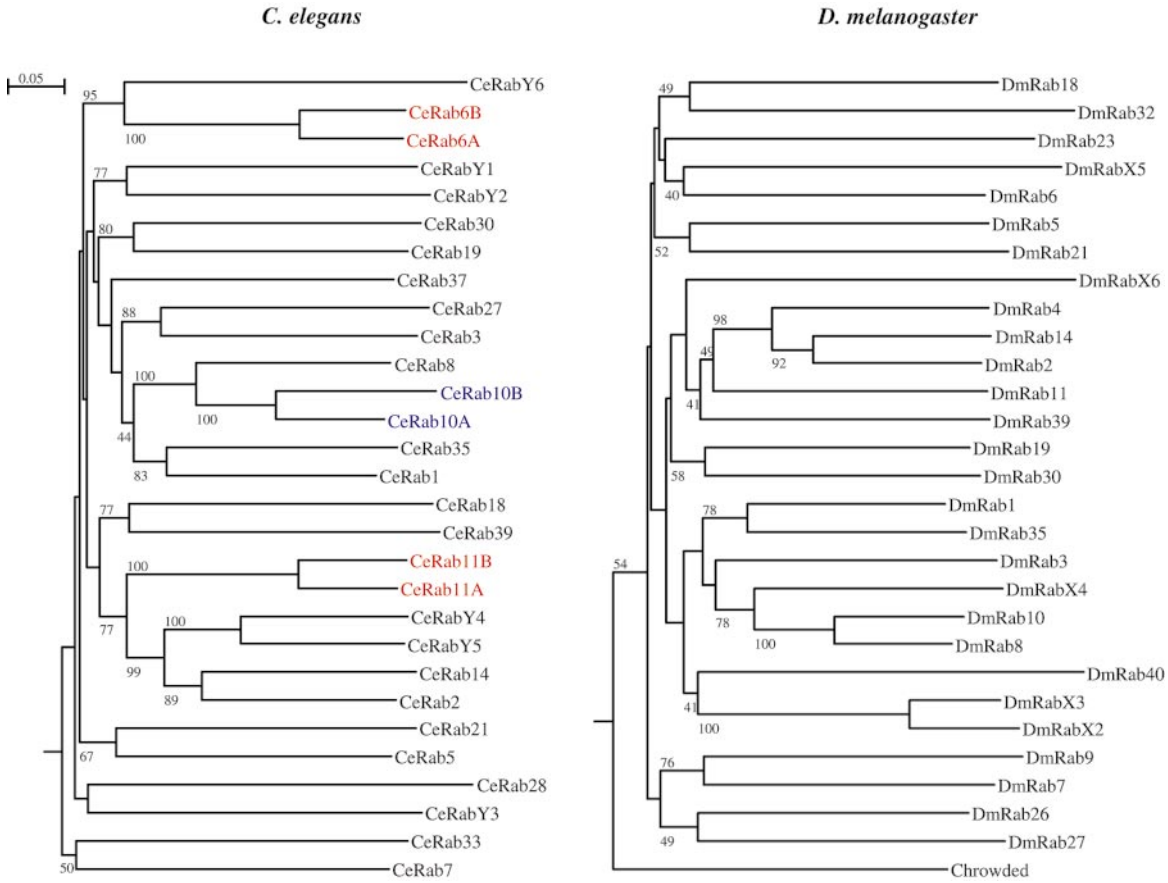

Figure 1. Neighbour-Joining tree of the *C. elegans* and *D. melanogaster* Rab families, rooted with H-Ras (not shown), scoring for amino acid difference. The numbers on the branches represent the percentage of 1000 bootstrap pseudo-samples supporting that branch; only values >40 % are shown. For clarity, subfamilies are represented in blue or red.

Table 1. Rab families in six different organisms

H. sapiens	*D. melanogaster*	*C. elegans*	*S. cerevisiae*	*S. pombe*	*A. thaliana*
Rab1a,b	DmRab1	CeRab1	Ypt1	Ypt1	AtRabD1-D2c
Rab2a,b	DmRab2	CeRab2			AtRabB1a-B1c
Rab3a,b,c,d	DmRab3	CeRab3			
Rab4a,b	DmRab4				
Rab5a,b,c	DmRab5	CeRab5	Ypt5152,53	Ypt5	AtRabF1-F72b
Rab6a,b,c	DmRab6	CeRab6a,b	Ypt6	Ryh1	AtRabH1a-H1e
Rab7	DmRab7	CeRab7	Ypt7	Ypt7	AtRabG1a-G3f
Rab8a,b	DmRab8	CeRab8		Ypt2	AtRabE1a-E1e
Rab9a,b	DmRab9				
Rab10	DmRab10	CeRab10a,b			
Rab11a,b	DmRab11	CeRab11a,b	Ypt31,32	Ypt3	AtRabA1a-A6b
Rab12					
Rab13					
Rab14	DmRab14	CeRab14			
Rab15					
Rab17					
Rab18	DmRab18	CeRab18			AtRabC1-C2b
Rab19	DmRab19	CeRab19			
Rab20					
Rab21	DmRab21	CeRab21			
Rab22a,b,c					
Rab23	DmRab23				
Rab24					
Rab25					
Rab26	DmRab26				
Rab27a,b	DmRab27	CeRab27			
Rab28		CeRab28			
Rab29					
Rab30	DmRab30	CeRab30			
Rab32	DmRab32				
Rab33a,b		CeRab33			
Rab34					
Rab35	DmRab35	CeRab35			
Rab36					
Rab37		CeRab37			
Rab38					
Rab39a,b	DmRab39	CeRab39			
Rab40a,b,c	DmRab40				
Rab41					
	Chrowded/ DmRabX1	CeRabY1	Sec4	Ypt4	
	DmRabX2	CeRabY2	Ypt10		
	DmRabX3	CeRabY3	Yp11		
	DmRabX4	CeRabY4			
	DmRabX5	CeRabY5			
	DmRabX6	CeRabY6			

Presence on the same row indicates that proteins are putative orthologues or are the closest homologues. Below the dotted line are represented Rabs for which no clear bomologues are found in the human Rab family, except for *A. thaliana* for which the family members are represented according to their ''homology group''. Accession numbers for all these proteins can be found in the Supplementary Material.

mitoylation motifs, but not a C-terminal prenylation motif. Although this type of lipid modification has not been described in the Rab family before, the presence of consensus sequences in the RabF motifs suggests that this is likely a Rab protein, albeit an unusual one.

Finally, we identified 60 independent open reading frames encoding Rab proteins in the *H. sapiens* genome (Tables 1 and S6, Figure 3). We caution that this analysis represents our best guess as of May 2001 and it is likely that the present list will need updating as the quality of the sequences in the databases improves. Surprisingly, our analysis differs considerably from a previous report published in early 2001.[15] First, we identified human proteins not previously reported, namely Rab24 and Rab33b. Secondly, we chose to discard sequences that differ from established Rab sequences only by the presence of insertions (for example, IGI_M1ctg4256_3 and Rab4a) or that are virtually identical (for example, GI7705963 and Rab9), as these are suggestive of putative splice variants, misidentification of splice sites or pseudogenes. Thirdly, we renamed a few Rabs to fit our nomenclature criteria.[4] (A direct comparison between the present study and the Bock *et al.* study[15] is shown in Table S6 of the Supplementary Material.)

We detect inconsistencies not just between these two studies on the human genome but also with

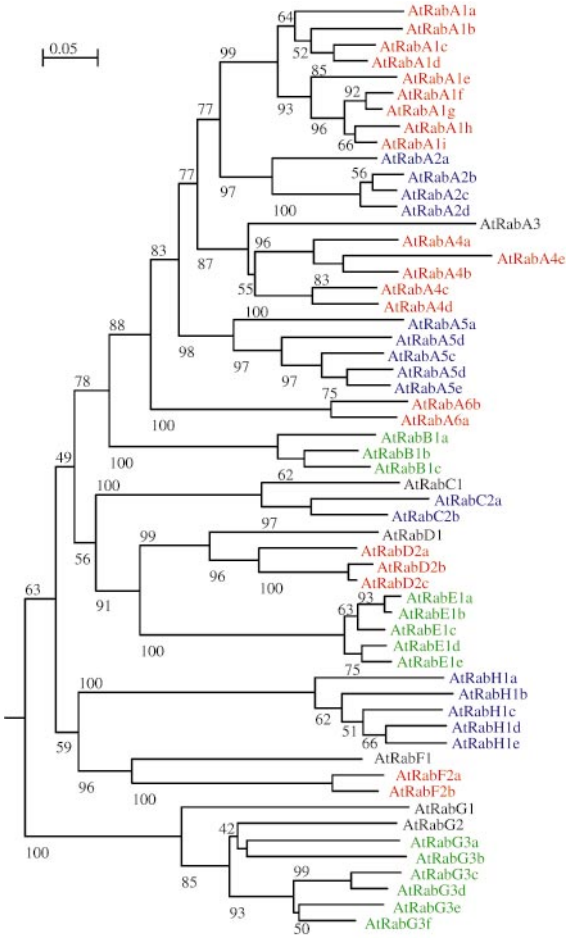

Figure 2. Neighbour-Joining tree of the *A. thaliana* Rab family, rooted with H-Ras (not shown), scoring for amino acid difference. The numbers on the branches represent the percentage of 1000 bootstrap pseudo-samples supporting that branch; only values >40 % are shown. For clarity, putative subfamilies are represented in blue, green or red.

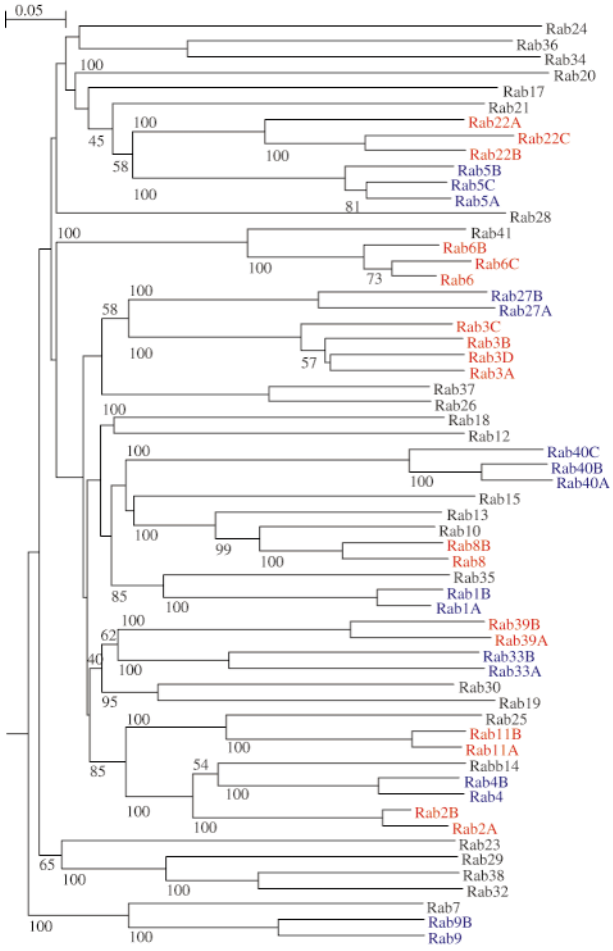

Figure 3. Neighbour-Joining tree of the *H. sapiens* Rab family, rooted with H-Ras (not shown), scoring for amino acid difference. The numbers on the branches represent the percentage of 1000 bootstrap pseudo-samples supporting that branch; only values >40 % are shown. For clarity, subfamilies are represented in blue or red.

annotation in Genbank. For example, the protein NP_003920 is annotated in GenBank as Rab7 like 1, renamed Rab42 by Bock *et al.*,[15] but we find it to be 93 % identical with rat Rab29 (BAA65444), indicating that it represents the human homologue of Rab29. Another example is the protein AF322067, which is clearly the human Rab34 protein but is annotated in GenBank as Rab39 (direct submission). This is particularly problematic in the Rab family, as the numbering of the different family members is taken frequently as indicative of function.

We observed that in phylogenetic trees of Rab proteins of different species, known orthologues always co-segregated, e.g. human Rab1 and yeast Ypt1p. This observation suggests that the strict phylogeny of function in the Ras superfamily previously observed by Valencia and co-workers[16] also applies within the Rab family (Figure 4). We thus propose that co-segregation in phylogenetic trees together with specific conservation in the RabSF regions be used as the criteria to assign putative orthologues.[4,10,11] The results of our analysis, including the proposed names for all members of the Rab families, are summarised in Table 1 (and can be found cross-referenced with the respective accession numbers in Tables S.3 to S.6 of the Supplementary Material).

Several proteins defied our annotation attempts. The nematode protein AAB04568, for example, has no clear homologue, presenting only vague homology to human Rab8. Protein AAB52431 is the clear orthologue of mammalian Rab2. However, two other proteins (CAB07356 and CAB07357) are also most similar to mammalian Rab2, but all three are not similar enough between them (i.e. <70 % identity) to be considered isoforms. In these difficult cases we decided to assign letters, rather than

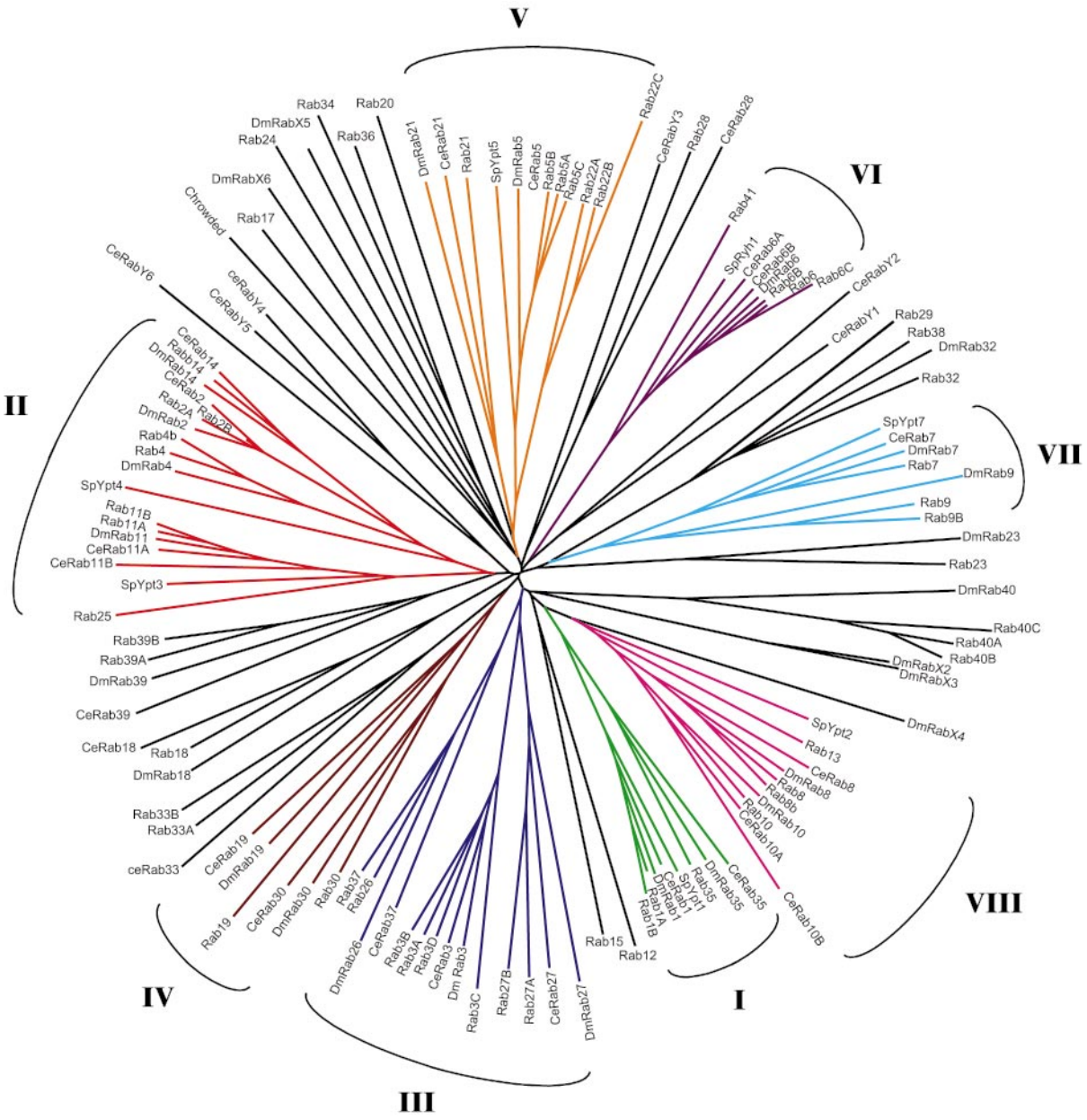

Figure 4. Neighbour-Joining dendogram of the human, nematode, fly and fission yeast Rab families. For clarity, human Rabs are not preceded by the species initials. The roman numerals (I - VIII) and the coloured branches represent Rab functional groups (see the text for further details).

the usual numbering system, to avoid confusion. For example, we propose AAB04568 to be named CeRabY1.

The Rab family of *A. thaliana* contains large putative subfamilies (up to nine isoforms in each subfamily), and in several cases the similarities to proteins of known function are not sufficiently high to allow us to ascribe putative orthologies. Unfortunately, some of the proteins have already been named, based on weak similarities, order of discovery or other undefined criteria. We now propose different criteria for annotation of plant Rab families. Inspection of the tree in Figure 2 reveals that there are eight major groups of Rab proteins in the *A. thaliana* genome that can be broadly related to mammalian Rabs by homology (Ian Moore, personal communication). Each of these groups should be labelled with capital letters, from A to H. Within these groups there are subfamilies, which can be defined based on co-segregation and conservation in the RabSF regions. Subfamilies should be defined by a number after homology group letter, which in turn is followed by a letter when there is more than one isoform (Figure 2). As

Rabs should be preceded by the species initials (e.g. AtRab), protein BAB09078 for example should be called AtRabA6a, which should be read as *Arabidopsis thaliana* Rab, group A, subtype 6, isoform a.

Rab families across evolution

Of all available Rab families, the fission yeast (*S. pombe*) presents the smallest number of Rab proteins (Table 2). In this organism, all the Rabs produce a detectable phenotype when the corresponding gene is disrupted, ranging from lethality (Ypt1p, Ypt2p, Ypt3p), temperature-sensitive growth (Rhy1p), fragmentation, size increase or reduction in number of vacuoles (Ypt7p andYpt4p, respectively) and vesicle accumulation (Ypt5p).[13,17] Furthermore, all *S. pombe* Rabs, with the exception of Ypt4p, are conserved in all other organisms.

The evolutionarily divergent budding yeast (*S. cerevisiae*), on the other hand, displays an increased number of Rab genes, many of which produce no phenotype when disrupted (Ypt31, Ypt32, Ypt51, Ypt52, Ypt53, Ypt10, Ypt11). This increased number is in part due to the appearance of subfamilies, i.e. functionally redundant isoforms (Ypt51, Ypt52,Ypt53 and Ypt31, Ypt32), but also due to the appearance of Rab proteins without clear functional or sequence homology to known Rab proteins in other organisms (Ypt10, Ypt11, and Sec4). The genome of the budding yeast contains several duplicated chromosomal regions that could underlie the appearance of subfamilies, which is consistent with the non-essentiality of the different isoforms.[18–20]

The human Rab family is the largest of all Rab families studied here and reflects the increased family size that accompanies multicellularity. Many of the Rab proteins have close homologues and form subfamilies. If we consider, in a simplistic view, that each subfamily corresponds to one function, then 39 functions are required in mammals. A significant number of these functions may involve specialised, tissue-specific trafficking pathways, as many Rabs are not expressed ubiquitously.

The nematode and the fly Rab families contain a number of Rabs intermediate between those of yeasts and mammals. Interestingly, we observe the existence of subfamilies in the nematode, but not in the fly Rab family, consistent with the recent genomic duplications observed in the nematode.[21] Analysis of these genomes suggests that there is not a linear increase in the number of Rab proteins with the number of cells, as the nematode is made of less than a thousand cells and the fruit fly contains more than three billion, and both organism have a similar number of Rab proteins. This is true even if we discard putative redundant Rabs (isoforms, i.e. Rabs forming subfamilies), assuming they represent a single function or sub-class[10] (26 Rab functions in *C. elegans* compared with 29 in *D. melanogaster*).

The Rab family of *A. thaliana* is quite different from the other Rab families considered here. It exhibits large subfamilies, which can be grouped in terms of homology to a small set of animal Rabs: Rab1, Rab2, Rab5, Rab6, Rab7, Rab8, Rab11 and Rab18. Surprisingly, the Rab11-like group in *A. thaliana* contains 26 proteins, but the significance of this fact is unclear. This Rab family organisation seems to be found in other plant Rab families (Ian Moore, personal communication). It seems obvious to suggest that Rab proteins in plants followed an evolutionary pathway different from that taken by animals or fungi. The rationalisation of this observations, however, will have to wait for more functional information on the different plant Rab proteins, as very little is known at present.

So what is the minimum number of Rab proteins required in a eukaryotic cell? One possibility is seven, as this is the number of Rabs found in the fission yeast.[13] Multicellularity and cellular specialisation may require more Rab proteins, possibly those corresponding to the conserved Rab proteins between the animal genomes considered here (Table 1). Such candidate Rabs include Rab3 and Rab27, two examples of tissue-restricted Rabs with specialised functions. The future availability of more sequenced genomes will allow a more accurate definition of the basic vesicular transport steps required in a multicellular organism.

Conserved interactions with general regulators and effectors

We recently proposed in mammalian Rabs that the RabF motifs, clustering in and around the putative switch regions, determine the interaction with

Table 2. Structure of the Rab family in six organisms

Species	Cell number	Rab number	Subfamilies
S. pombe	1	7	0
S. cerevisiae	1	11	2
C. elegans	$\sim 1 \times 10^3$	29	3
D. melanogaster	$\sim 1 \times 10^9$	29	0
H. sapiens	$\sim 1 \times 10^{13}$	$\geqslant 60$	11
A. thaliana	-	$\geqslant 57$	12

The listing of the family members and their accession numbers can be found in the Supplementary Material.

Rab-specific general regulators such as Rab escort proteins (REP) and GDP dissociation inhibitors (RabGDI). We sought to determine whether this mode of interaction with general regulators is conserved across evolution. To do so, we asked if the same regions are conserved in the different Rab familes considered in this study. We calculated pHMMs and generated model sequences for each organism Rab family. Upon alignment, we observed that the same regions are conserved in all organisms and that there is no organism-specific consensus (Figure 5). Thus, the RabF motifs seem to be a feature conserved in evolution and may indicate a conserved mode of interaction between Rabs and general regulators. Recent work showing that similar positions in yeast REP and RabGDI mediate interactions with Rabs further supports this possibility.[22]

We worried that this observation could have been biased by the Rab identification strategy followed, as the presence of RabF motifs was one of the criteria used. However, this was by no means the only criterion. We BLASTed each individual ''putative'' Rab sequence against the non-redundant and organism-specific databases to confirm the similarity with other Rabs, and to look for further members of the Rab family. Also, we checked every sequence against pHMM of other small GTPase families (Ras, Rho, Arf, Ran, Gem).

We proposed previously the existence of Rab subfamily (RabSF) regions in mammalian Rabs, possibly involved in determining binding specificity to effectors. These RabSF regions are conserved across species.[10] Consistently, Rabs from evolutionarily distant organisms exhibit functional complementation. For example, yeast Ypt1 deletions or temperature sensitive mutations can be complemented by small GTPases from *Volvox carteri*, *Chlamydomonas reinhardtii*,[23] *Brassica napus*[24] and *Mus musculus*,[25] and ypt6 null mutants can be complemented by a small GTPase from *A. thaliana*.[26]

Taken together, the absence of organism-specific consensus, the conservation of RabF regions, the cross-species functional complementation and the conservation of RabSF regions make a strong argument for a highly conserved mechanism of effector and general regulator recognition, likely to be present at the point of divergence from other small GTPases. It also suggests that this family originated by a single divergence event and that these interaction mechanisms represent a major constraint to the evolution of Rab proteins. Furthermore, the conservation of this effector recognition mechanism is indicative of effector conservation, an assumption that is supported by some recent evidence. While exchange factors for the Rab family form a very divergent class of proteins, there is a striking conservation of these proteins across evolution for known orthologues in the few known cases. For example, the mammalian and nematode exchange factors for Rab3 (Rab3GEP and Aex3) are highly conserved,[27,28] and so are the exchange factors for Ypt51 and Rab5 (Vps9 and Rabex-5).[29,30]

Prenylation and targeting motifs

Rab proteins contain one or two C-terminal cysteine residues that undergo post-translational prenyl modification.[31] These cysteine residues are arranged in a variety of prenylation motifs. Some Rabs (such as HsRab8 and HsRab23) have a single cysteine residue, fourth from the C terminus, sometimes within a CAAX box (C, cys; A, aliphatic; X, any), a motif commonly observed in the Ras and Rho families.[32] However, most Rabs have two cysteine residues arranged in different double-cysteine prenylation motifs (e.g. XXCC, XCXC, CCXX, CCXXX, XCCX), both of which are modified by geranylgeranyl moieties. Unlike Ras and Rho proteins, the prenylation motif in Rabs does not determine which prenyl transferase (and consequently which type of prenyl moiety) will modify the C-terminal cysteine residues. All Rabs appear to be substrates for a unique enzyme, Rab Geranyl-

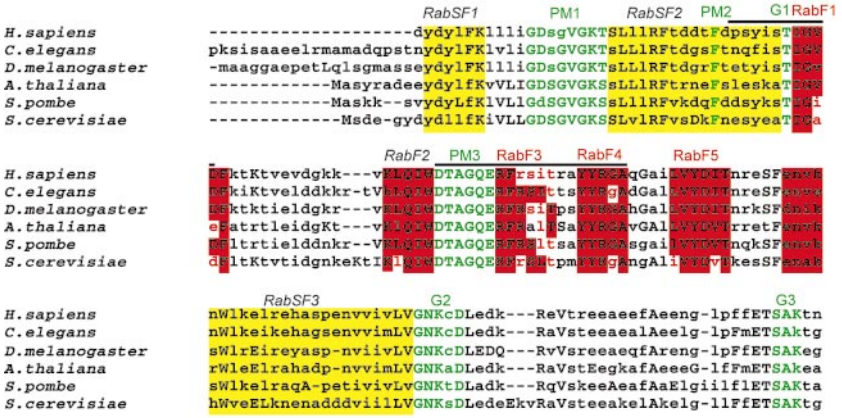

Figure 5. Partial alignment of model sequences for each family, calculated from pHMM describing the multiple sequence alignment of the complete Rab family in each organism. Positions in uppercase occur at $p > 0.5$. The PM/G motifs are highlighted in green, the RabF motifs in red boxes or red characters when the position is often occupied by a conservative substitution, and the RabSF regions in yellow boxes. Black bars above the alignment represent the switch regions.

geranyl Transferase (RGGT), following the binding to REP, a general regulator of Rabs.

It is conceivable that the diversity of Rab prenylation motifs arises from lack of functional constrains other than those imposed by the geranylgeranylation reaction mechanism. When we compared the prenylation motifs in Rabs from all organisms, we observed that the number of cysteine residues is frequently conserved, and in many cases the topology of the prenylation motif is retained (e.g. Ypt1(CC) → Rab1(CC), Ypt6(CXC) → Rab6(CXC)). Conservation of these topologies indicates constraints to evolution possibly due to the requirement for carboxyl-methylation, even though the functional significance of Rab carboxyl-methylation affecting Rabs ending in CXC but not in CC is not understood.[33,34] Furthermore, we observed conservation of the number of cysteine residues available for prenylation. This suggests that the number of prenyl groups (one or two) may be functionally important, possibly revealing the existence of two distinct membrane-association mechanisms or perhaps different membrane targeting strategies.

We observed some unusual Rab proteins in this regard. For example, protein BAB32953 in *A. thaliana* exhibits putative N-terminal myristoylation and palmitoylation motifs, but no C-terminal prenylation motifs, and these motifs are also found in similar plant proteins, indicating a conserved feature. This type of lipid modification is novel within the Rab family. Another peculiar Rab protein is Rab24, which is thought to be cytosolic.[35] These unusual cases may represent recent evolutions of the Rab family where motifs not normally present in this family are recruited to provide for new functions.

We caution that sequencing errors or artefacts complicate annotation attempts. For example, several Rab proteins in *D. melanogaster*, *C. elegans* and *A. thaliana* do not exhibit C-terminal prenylation motifs. However, their putative orthologues in other organisms do exhibit prenylation motifs, suggesting either bad quality sequencing or deficient gene identification algorithms.

Rab functional groups

Phylogenetic trees for all the sequences considered here, reconstructed using the Neighbour-Joining method, revealed a clear phylogeny of function, as opposed to a phylogeny of species (Figures 1-4). In other words, Rab proteins of similar function in different organisms always co-segregate. As mentioned above, this represents an extension of the strict phylogeny of function previously observed in the yeast Ras superfamily.[16,36] Within the clades representing each putative Rab ''function'', we observed a phylogeny of species, with proteins segregating according to organism provenience (Figure 4).

Figures 1-3 show the trees calculated for each organism Rab family. We noted that some proteins always co-segregate, even if they do not conform to the criteria defining isoforms.[4] For example, members of the Rab1 sub-family always segregate with Rab35. Based on the tree topology, we can identify eight possible groups of co-segregating proteins (Figure 4). The proteins in each of these groups are more similar at the amino acid level than any two random Rab proteins, suggesting a higher-order organisation in the Rab family, above the subfamily level. This higher-order organisation may represent a shared ancestry between co-segregating proteins, functional relatedness or both.

To test the hypothesis that this co-segregation of ''unrelated'' proteins represents co-segregation of functional properties, we sought to identify patterns of function and/or cellular localisation in each group indicated in Figure 4. In group V, which includes the sub-families 5 and 22, we noted a pattern of subcellular localisation and possibly of function. Rab5a has been studied extensively, it localises to early endosomes and clathrin-coated vesicles, and regulates endosome budding and fusion.[2,37–41] Rab22a localises to endosomes and the plasma membrane. Over-expression of Rab22a results in the formation of abnormal endosomal structures, which is suggestive of a role in endocytosis.[42] Rab21 also segregates with the subfamilies 5 and 22, albeit showing less sequence relatedness. Rab21 seems to be specific for polarised cells, where it localises to apical vesicles and shows partial localisation to an endosomal compartment, suggesting that it may be functionally related to Rab5.[43] Interestingly, other Rab proteins that broadly segregate with Rab5 isoforms also display an endosomal localisation, namely Rab17[44,45] and Rab20,[46] but not Rab24, which is reportedly cytosolic.[35]

In group III we observed a pattern of subcellular localisation to secretory granules. Rab37 has recently been identified and localised to secretory granules in mast cells,[47] and Rab26 has been localised to secretory granules in pancreatic acinar cells.[48] Rab27a was the subject of recent work by several groups and found to localise to secretory granules (lytic granules) of cytotoxic T-lymphocytes (CTL). Defects in the *RAB27A* gene in Griscelli disease lead to haemophagocytic syndrome due to loss of CTL activity.[49,50] In melanocytes, Rab27 associates with melanosomes, lysosome-like pigment-containing organelles destined for secretion.[51–53] Rab27 appears to recruit myosinVa to regulate the transport of melanosomes to the cell periphery prior to secretion.[51–53] Rab3 isoforms co-segregate with other members of this group in the human and fly Rab tree, but not in the nematode Rab family tree, which is suggestive of a more distant relationship. Members of the Rab3 subfamily have been implicated in regulated secretory events such as neurotransmitter release and insulin secretion, and associate with secretory granules such as synaptic vesicles,[54–60] thus exhibiting a

similar type of cellular localisation as the other members of this group.

Group VII includes Rab7 and Rab9 isoforms. Both proteins show overlapping localisation to late endosomes.[61–63]

In group II, there is no apparent pattern of similar subcellular localisation. For example, Rab2 localises to the endoplasmic reticulum (ER) and Golgi,[64] and Rab4 to endosomes and plasma membrane.[65] Nevertheless, there may be a pattern of functional similarity in this group. Rab11 and 25 are proposed to be involved in recycling of proteins through the recycling endosome,[66,67] and Rab4 is proposed to be involved in endocytic recycling.[68] No functional data are available for Rab14, but Rab2 has recently been proposed to regulate a recycling step in the retrograde Golgi-ER transport.[69]

In conclusion, we suggest that there is a recognisable pattern of subcellular localisation and possibly of function, which supports the hypothesis of a phylogeny of function applying between the subfamily and the family level.

We next questioned the origin of the sequence relatedness, and functional/localisation similarity underlying the groups shown in Figure 4. The simplest explanation is that the phylogeny of function also represents the evolutionary history of the Rab family, and that members of one given branch have a shared ancestry. For example, Rab5 and Rab22 would share the same ancestor, and this ancestral protein would also be the ancestor of the budding yeast Ypt5.

This hypothesis is based on the phylogenetic reconstruction methods. However, these methods can be biased to functional similarities in highly conserved protein families, therefore, different lines of evidence need to be obtained to substantiate it. Supporting this hypothesis is the observation that there is differential conservation in discrete regions such as the PM/G and RabF motifs in some cases. For example, human group II proteins Rab25, Rab14, and the sub-families Rab11, Rab4 and Rab2, all share identical IGVEF sequence at the RabF1 motif, while human group III proteins Rab3 and Rab27 isoforms, Rab26 and Rab37 display the conserved sequence VGIDF.

In order to complement this analysis, we searched all available genomic structures of human Rab genes for an indication of shared ancestry. If proteins from one branch indeed arose from a common ancestor, we expected to find similarities in exon organisation. We retrieved available intron-exon boundaries in Rab genes from GenBank using MapView. In many cases the sequencing data are still of low quality, resulting in incomplete or no useful information at all. Using a limited number of genes, we observed many common features in genomic structures within Rab functional groups. The intron-exon boundaries are either absolutely conserved, or are close within a maximum of five codons (Figure 6). Some Rabs within each group share highly similar genomic structures. For example, Rab11 and Rab25 are almost identical, while Rab4 and Rab14 are distinguished by the appearance of an intron in Rab14 splitting in two the fifth exon in Rab4. Within subfamilies, genomic structures tend to be highly conserved if not identical (data not shown), a fact already noted by others.[70–73]

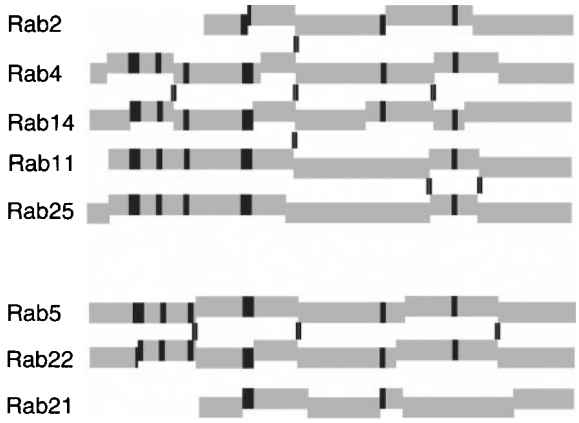

Figure 6. Representation of the coding exons (grey boxes) of available ''functional group members'', mapped to a cartoon representing a generalised Rab protein sequence, aligned by the conserved PM and G motifs (black boxes).

Based on these observations, we cannot discredit the hypothesis that co-segregating Rab proteins share a common ancestry. Consequently, the Rab family trees not only represent a separation according to function and may reflect the evolutionary history of this family of proteins. A more comprehensive analysis of gene structures of Rab genes in different organisms is required to provide clearer evidence for shared ancestry.

Conclusions

We have identified and annotated complete Rab families in all eukaryotic organisms that had their genome substantially sequenced in May 2001. We propose here objective criteria for annotation of animal and plant Rab families on the basis of recognition of putative orthologies.

Our analysis suggests that interactions between Rab proteins and their general regulators and specific effectors is conserved across evolution, as the sequence determinants of this interactions (RabF motifs and RabSF regions) are conserved in all Rab families studied here.

We addressed the evolution of the Rab family and observed a higher-order organisation within the Rab family corresponding to Rab proteins, which co-segregate in phylogenetic trees. Rabs within these groups exhibit similar function and/or cellular localisation and related genomic structures. It is tempting to speculate that early in eukaryotic evolution a minimum number of Rab

proteins provided the ''ancestral'' Rab regulatory activities. Organism specialisation and multicellularity drove the multiplication of Rab family members from the initial set of ''ancestral Rabs''. These novel Rabs appear to have maintained one or more properties that defined their ancestry, allowing us to group Rab proteins according to their ancestry, i.e. according to their putative ''ancestral Rab function''. Thus, we propose that these related functions/cellular localisations form an intermediate level of classification between family and subfamily, better described as ''Rab functional groups''. One interesting and testable possibility is that this organisation level could have predictive value to suggest a function, localisation or interactions with effectors of a given Rab protein. A possible Rab27 effector, melanophilin was identified recently and shown to be similar to Rabphilin-3a, a Rab3a effector.[74] This raises the possibility that members of one functional group will interact with a family of conserved effectors and suggest parallel evolution between Rabs and their effectors. We expect more functional groups to be defined as more functional information becomes available.

The minimal set of Rab proteins has been equated with the essential yeast Rab proteins.[13] These minimal Rab properties may represent localisation to a given cellular compartment, interaction with classes of related effectors/regulators, specific GTPase characteristics or a combination of these. A better understanding of this issue is essential to fully understand the nature of the ''ancestral Rab functions'', the way that they evolved to provide regulators for increasingly complex organisms, and to ascribe general functions to novel Rab sequences based solely on their segregation pattern in phylogenetic trees. Furthermore, the understanding of the properties shared by groups of co-segregating Rabs identified here will allow informative correlations between group-specific sequence conservation and localisation/function of Rab proteins.

Materials and Methods

We retrieved protein sequences of known Rab families from GenBank. To identify the complete Rab families in *H. sapiens, D. melanogaster, C. elegans* and *A. thaliana*, we downloaded the latest releases of the calculated open reading frames of each organism from the public databases. We then used a profile hidden Markov model (pHMM) calculated from the alignment of the mammalian Rab sequences presented previously[4] to query each database using the software HMMER 2.1.1 found at hmmer.wustl.edu.[75]

All the positive hits were then inspected visually, compared to pHMMs representing other small GTPase families and individually BLASTed against the non-redundant database in GenBank to assert if they were indeed Rab proteins, and to ensure that no sequence was missed from our analysis due to a possible bias created by the query sequences.

Protein sequences were aligned using the CLUSTAL W 1.80[76] multiple sequence alignment program with default parameters. Phylogenetic trees were reconstructed by the distance method of Neighbour-Joining,[77] scoring for observed amino acid difference and were always bootstrapped with 1000 replicates,[78] using the software Phylo_Win.[79] Genomic structures were obtained from GenBank, *via* the interface MapView.

Acknowledgements

We thank Ian Moore for helpful discussions concerning the plant Rab families and members of our laboratory for stimulating ideas. This work was supported, in part, by the Wellcome Trust. J.P.-L. is a student of ''Programa Gulbenkian de Doutoramento em Biologia e Medicina''.

References

1. Chavrier, P. & Goud, B. (1999). The role of ARF and Rab GTPases in membrane transport. *Curr. Opin. Cell Biol.* **11**, 466-475.
2. Zerial, M. & McBride, H. M. (2001). Rab proteins as membrane organisers. *Nature Rev. Mol. Cell Biol.* **2**, 107-119.
3. Stenmark, H. & Olkkonen, V. M. (2001). The Rab GTPase family. *Genome Biol.* **2**, R3007.1-R3007.7.
4. Pereira-Leal, J. B. & Seabra, M. C. (2000). 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.* **301**, 1077-1087.
5. Merithew, E., Hatherly, S., Dumas, J. J., Lawe, D., Heller-Harrison, R. & Lambright, D. (2001). Structural plasticity of an invariant hydrophobic triad in the switch regions of Rab GTPases is a determinant of effector recognition. *J. Biol. Chem.* **276**, 13982-13988.
6. Dumas, J. J., Zhu, Z., Connolly, J. L. & Lambright, D. G. (1999). Structural basis of activation and GTP hydrolysis in Rab proteins. *Structure,* **7**, 413-423.
7. White, M. A., Nicolette, C., Minden, A., Polverino, A., Van Aeist, L., Marin, M. & Wigler, M. H. (1995). Multiple ras functions can contribute to mammalian cell transformation. *Cell,* **80**, 533-541.
8. Vetter, I. R., Linnemann, T., Wohlgemuth, S., Geyer, M., Kalbitzer, H. R., Herrmann, C. & Wittinghofer, A. (1999). Structural and biochemical analysis of Ras-effector signaling via RalGDS. *FEBS Letters,* **451**, 175-180.
9. Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCormick, F. & Wittinghofer, A. (1995). The 2.2 Å crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature,* **375**, 554-560.
10. Moore, I., Schell, J. & Palme, K. (1995). Subclass-specific sequence motifs identified in Rab GTPases. *Trends Biochem. Sci.* **20**, 10-12.
11. Ostermeier, C. & Brunger, A. T. (1999). Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabbphilin-3A. *Cell,* **96**, 363-374.
12. Armstrong, J. (2000). How do Rab proteins function in membrane traffic? *Int. J. Biochem. Cell Biol.* **32**, 303-307.

13. Lazar, T., Gotte, M. & Gallwitz, D. (1997). Vesicular transport: how many Ypt/Rab-GTPases make a eukaryotic cell? *Trends Biochem. Sci.* **22**, 468-472.
14. Armstrong, J. (2000). Membrane traffic between genomes. *Genome Biol.* **1**, 104.1-104.4.
15. Bock, J., Matern, H., Peden, A. & Scheller, R. H. (2001). A genomic prespective on membrane compartment organization. *Nature,* **409**, 839-841.
16. Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. (1991). The ras protein family: evolutionary tree and role of conserved amino acids. *Biochemistry,* **30**, 4637-4648.
17. Olkkonen, V. M. & Stenmark, H. (1997). Role of Rab GTPases in membrane traffic. *Int. Rev. Cytol.* **176**, 1-85.
18. Seoighe, C. & Wolfe, K. (1998). Extent of genomic rearrangement after genome duplication in yeast. *Proc. Natl Acad. Sci. USA,* **95**, 4447-4452.
19. Wolfe, K. & Shields, D. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature,* **387**, 708-713.
20. Achaz, G., Coissac, E., Viari, A. & Netter, P. (2000). Analysis of intrachromosomal duplications in yeast *Saccharomyces cerevisiae*: a possible model for their origin. *Mol. Biol. Evol.* **17**, 1268-1275.
21. Friedman, R. & Hughes, A. (2001). Gene duplication and the structure of the eukaryotic genomes. *Genome Res.* **11**, 373-381.
22. Alory, C. & Balch, W. E. (2000). Molecular basis for Rab prenylation. *J. Cell Biol.* **150**, 89-103.
23. Fabry, S., Steigerwald, R., Berknlau, C., Dietmaier, W. & Schmidt, R. A. (1995). Structure-function analysis of small G proteins from volvox and *Chlamidomonas* by complementation of *Saccharomyces cerevisiae* YPT/SEC mutations. *Mol. Gen. Genet.* **247**, 265-274.
24. Palme, K., Diefenthal, T. & Moore, I. (1993). The ypt gene family from maize and Arabidopsis: structural and functional analysis. *J. Exp. Bot.* **44**, 183-195.
25. Haubruck, H., Prange, R., Vorgias, C. & Gallwitz, D. (1989). The ras-related mouse ypt1 protein can functionally replace the YPT1 gene product in yeast. *EMBO J.* **8**, 1427-1432.
26. Bednarek, S., Reynolds, T., Schroeder, M., Grabowski, R., Hengst, L., Gallwitz, D. & Raikhel, N. (1994). A small GTP-binding protein from *Arabidopsis thaliana* functionally complements the yeast ypt6 null mutant. *Plant Physiol.* **104**, 591-596.
27. Iwasaki, K., Staunton, J., Saifee, O., Nonet, M. & Thomas, J. H. (1997). aex-3 encodes a novel regulator of presynaptic activity in *C. elegans*. *Neuron,* **18**, 613-622.
28. Wada, M., Nakanishi, H., Satoh, A., Hirano, H., Obaishi, H., Matsuura, Y. & Takai, Y. (1997). Isolation and characterization of a GDP/GTP exchange protein specific for the Rab3 subfamily small G proteins. *J. Biol. Chem.* **272**, 3875-3878.
29. Hama, H., Tall, G. G. & Horazdovsky, B. F. (1999). Vps9p is a guanine nucleotide exchange factor involved in vesicle-mediated vacuolar protein transport. *J. Biol. Chem.* **274**, 15284-15291.
30. Horiuchi, H., Lippe, R., McBride, H. M., Rubino, M., Woodman, P., Stenmark, H. *et al*. (1997). A novel Rab5 GDP/GTP exchange factor complexed to Rabaptin-5 links nucleotide exchange to effector recruitment and function. *Cell,* **90**, 1149-1159.
31. Seabra, M. C. (2000). Biochemistry of Rab geranylgeranyl transferase. In *The Enzymes* (Tamanoi, F. & Sigman, D. S., eds), vol. 21, pp. 131-154, Academic Press, San Diego.
32. Casey, P. J. & Seabra, M. C. (1996). Protein prenyltransferases. *J. Biol. Chem.* **271**, 5289-5292.
33. Smeland, T. E., Seabra, M. C., Goldstein, J. L. & Brown, M. S. (1994). Geranylgeranylated Rab proteins terminating in Cys-Ala-Cys, but not Cys-Cys, are carboxyl-methylated by bovine brain membranes *in vitro*. *Proc. Natl Acad. Sci. USA,* **91**, 10712-10716.
34. Bergo, M. O., Gordon, K., Ambroziak, P., Otto, J. C., Casey, P. J.,Gomes, A. Q. *et al*. (2001). Isoprenylcysteine carboxyl methyltransferase deficiency in mice. *J. Biol. Chem.* **276**, 5841-5845.
35. Erdman, R. A., Shellenberger, K. E., Overmeyer, J. H. & Maltese, W. A. (2000). Rab24 is an atypical member of the Rab GTPase family. Deficient GTPase activity, GDP dissociation inhibitor interaction, and prenylation of Rab24 expressed in cultured cells. *J. Biol. Chem.* **275**, 3848-3856.
36. Garcia-Ranea, J. A. & Valencia, A. (1998). Distribution and functional diversification of the ras superfamily in *Saccharomyces cerevisiae*. *FEBS Letters,* **434**, 219-225.
37. Bucci, C., Parton, R. G., Mather, I. H., Stunnenberg, H., Simons, K., Hoflack, B. & Zerial, M. (1992). The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway. *Cell,* **70**, 715-728.
38. Barbieri, M. A., Hoffenberg, S., Roberts, R., Mukhopadhyay, A., Pomrehn, A., Dickey, B. F. & Stahl, P. D. (1998). Evidence for a symmetrical requirement for Rab5-GTP in *in vitro* endosome-endosome fusion. *J. Biol. Chem.* **273**, 25850-25855.
39. Roberts, R. L., Barbieri, M. A., Pryse, K. M., Chua, M., Morisaki, J. H. & Stahl, P. D. (1999). Endosome fusion in living cells overexpressing GFP-rab5. *J. Cell Sci.* **112**, 3667-3675.
40. Rybin, V., Ullrich, O., Rubino, M., Alexandrov, K., Simon, I., Seabra, M. C. *et al*. (1996). GTPase activity of Rab5 acts as a timer for endocytic membrane fusion. *Nature,* **383**, 266-269.
41. Stenmark, H., Parton, R., Steele-Mortimer, O., Lutcke, A., Gruenberg, J. & Zerial, M. (1994). Inhibition of Rab5 GTPase activity stimulates membrane fusion in endocytosis. *EMBO J.* **13**, 1287-1296.
42. Olkkonen, V. M., Dupree, P., Killisch, I., Lutcke, A., Zerial, M. & Simons, K. (1993). Molecular cloning and subcellular localization of three GTP-binding proteins of the rab subfamily. *J. Cell Sci.* **106**, 1249-1261.
43. Opdam, F. J., Kamps, G., Croes, H. J., van Bokhoven, H., Ginsel, L. A. & Fransen, J. A. (2000). Expression of Rab small GTPases in epithelial Caco-2 cells: Rab21 is apically located GTP-binding protein in polarised intestinal epithelial cells. *Eur. J. Cell Biol.* **79**, 308-318.
44. Hunziker, W. & Peters, P. J. (1998). Rab17 localizes to recycling endosomes and regulates receptor-mediated transcytosis in epithelial cells. *J. Biol. Chem.* **273**, 15734-15741.
45. Zacchi, P., Stenmark, H., Parton, R. G., Orioli, D., Lim, F., Giner, A. *et al*. (1998). Rab17 regulates membrane trafficking through apical recycling endosomes in polarized epithelial cells. *J. Cell Biol.* **140**, 1039-1053.
46. Lutcke, A., Parton, R., Murphy, C., Olkkonen, V. M., Dupree, P., Valencia, A. *et al*. (1994). Cloning and subcellular localization of novel rab proteins reveals polarized and cell type specific expression. *J. Cell Sci.* **107**, 3437-3438.

47. Masuda, E. S., Luo, Y., Young, C., Shen, M., Rossi, A. B., Huang, B. C. *et al*. (2000). Rab37 is a novel mast cell specific GTPase localized to secretory granules. *FEBS Letters,* **470**, 61-64.
48. Yoshie, S., Imai, A., Nashida, T. & Shimomura, H. (2000). Expression, characterization, and localization of Rab26, a low molecular weight GTP-binding protein, in the rat parotid gland. *Histochem. Cell Biol.* **113**, 259-263.
49. Haddad, E., Wu, X., Hammer, J., III & Henkart, P. (2001). Defective granule exocytosis in Rab27a-deficient lymphocytes from ashen mice. *J. Cell Biol.* **152**, 835-842.
50. Stinchcombe, J. C., Barral, D. C., Mules, E. H., Booth, S., Hume, A. N., Machesky, L. M. *et al*. (2001). Rab27a is required for regulated secretion in cytotoxic t lymphocytes. *J. Cell Biol.* **152**, 825-834.
51. Hume, A. N., Collinson, L. M., Rapak, A., Gomes, A. Q., Hopkins, C. R. & Seabra, M. C. (2001). Rab27a regulates the peripheral distribution of melanosomes in melanocytes. *J. Cell Biol.* **152**, 795-808.
52. Wu, X., Rao, K., Bowers, M., Copeland, N., Jenkins, N. & Hammer, J. (2001). Rab27a enables myosinVa-dependent melanosomes capture. *J. Cell Sci.* **114**, 1091-1100.
53. Bahadoran, P., Aberdam, R., Mantoux, F., Busca, R., Bille, K., Yalman, N. *et al*. (2001). Rab27a. A key to melanosome transport in human melanocytes. *J. Cell Biol.* **152**, 843-850.
54. Iezzi, M., Escher, G., Meda, P., Charollais, A., Baldini, G., Darchen, F. *et al*. (1999). Subcellular distribution and function of Rab3A, B, C, and D isoforms in insulin-secreting cells. *Mol. Endocrinol.* **13**, 202-212.
55. Tuvim, M. J., Adachi, R., Chocano, J. F., Moore, R. H., Lampert, R. M., Zera, E., Romero, E. *et al*. (1999). Rab3D, a small GTPase, is localized on mast cell secretory granules and translocates to the plasma membrane upon exocytosis. *Am. J. Respir. Cell Mol. Biol.* **20**, 79-89.
56. Al-Matubsi, H. Y., Breed, W., Jenkin, G. & Fairclough, R. J. (1999). Co-localization of Rab3B and oxytocin to electron dense granules of the sheep corpus luteum during the estrous cycle. *Anat. Rec.* **254**, 214-221.
57. Raffaniello, R. D., Lin, J., Schwimmer, R. & Ojakian, G. K. (1999). Expression and localization of Rab3D in rat parotid gland. *Biochim. Biophys. Acta,* **1450**, 352-363.
58. Fischer, V. M., Mignery, G. A., Baumert, M., Perin, M. S., Hanson, T. J., Burger, P. M. (1990). rab3 is a small GTP-binding protein exclusively localized to synaptic vesicles. *Proc. Natl Acad. Sci. USA,* **87**, 1988-1992.
59. Darchen, F., Zahraoui, A., Hammel, F., Monteils, M. P., Tavitian, A. & Scherman, D. (1990). Association of the GTP-binding protein Rab3A with bovine adrenal chromaffin granules. *Proc. Natl Acad. Sci. USA,* **87**, 5692-5696.
60. Oishi, H., Sasaki, T., Nagano, F., Ikeda, W., Ohya, T., Wada, M. *et al*. (1998). Localization of the Rab3 small G protein regulators in nerve terminals and their involvement in Ca^{2+}-dependent exocytosis. *J. Biol. Chem.* **273**, 34580-34585.
61. Feng, Y., Press, B. & Wandinger-Ness, A. (1995). Rab7: an important regulator of late endocytic membrane traffic. *J. Cell Biol.* **131**, 1435-1452.
62. Lombardi, D., Soldati, T., Riederer, M. A., Goda, Y., Zerial, M. & Pfeffer, S. R. (1993). Rab9 functions in transport between late endosomes and the trans Golgi network. *EMBO J.* **12**, 677-682.
63. Meresse, S., Gorvel, J. P. & Chavrier, P. (1995). The Rab7 GTPase resides on a vesicular compartment connected to lysosomes. *J. Cell Sci.* **108**, 3349-3358.
64. Tisdale, E. J., Bourne, J. R., Khosravi-Far, R., Der, C. J. & Balch, W. E. (1992). GTP-binding mutants of rab1 and rab2 are potent inhibitors of vesicular transport from the endoplasmic reticulum to the Golgi complex. *J. Cell Biol.* **119**, 749-761.
65. Sonnichsen, B., De Renzis, S., Nielsen, E., Rietdorf, J. & Zerial, M. (2000). Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11. *J. Cell Biol.* **149**, 901-914.
66. Casanova, J. E., Wang, X., Kumar, R., Bhartur, S. G., Navarre, J., Woodrum, J. E. *et al*. (1999). Association of Rab25 and Rab11a with the apical recycling system of polarized Madin-Darby canine kidney cells. *Mol. Biol. Cell,* **10**, 47-61.
67. Ullrich, O., Reinsch, S., Urbe, S., Zerial, M. & Parton, R. G. (1996). Rab11 regulates recycling through the pericentriolar recycling endosome. *J. Cell Biol.* **135**, 913-924.
68. van der Sluijs, P., Hull, M., Webster, P., Male, P., Goud, B. & Mellman, I. (1992). The small GTP-binding protein rab4 controls an early sorting event on the endocytic pathway. *Cell,* **70**, 729-740.
69. Tisdale, E. J. (1999). A Rab2 mutant with impaired GTPase activity stimulates vesicle formation from pre-Golgi intermediates. *Mol. Biol. Cell,* **10**, 1837-1849.
70. Adachi, R., Nigam, R., Tuvim, M. J., DeMayo, F. & Dickey, B. F. (2000). Genomic organization, chromosomal localization, and expression of the murine RAB3D gene. *Biochem. Biophys. Res. Commun.* **273**, 877-883.
71. Bhartur, S. G., Calhoun, B. C., Woodrum, J., Kurkjian, J., Iyer, S., Lai, F. & Goldenring, J. R. (2000). Genomic structure of murine Rab11 family members. *Biochem. Biophys. Res. Commun.* **269**, 611-617.
72. Ramalho, J. S., Tolmachova, T., Hume, A., McGuigan, A., Gregory-Evans, C., Huxley, C. & Seabra, M. C. (2001). Chromosomal mapping, gene structure and characterization of the human and murine RAB27B gene. *BMC Genet.* **2**, 2.
73. Tolmachova, T., Ramalho, J. S., Anant, J. S., Schultz, R. A., Huxley, C. M. & Seabra, M. C. (1999). Cloning, mapping and characterization of the human *RAB27A* gene. *Gene,* **239**, 109-116.
74. Matesic, L. E., Yip, R., Reuss, A. E., Swing, D. A., O'Sullivan, T. N., Fletcher, C. F. *et al*. (2001). Mutations in *Mlph*, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice. *Proc. Natl Acad. Sci. USA,* **98**, 10238-10243.
75. Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics,* **14**, 755-763.
76. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.
77. Saitou, N. & Nei, M. (1987). The neighbour-joining method: a new method to reconstruct phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425.

78. Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* **39**, 783-791.
79. Galtier, N., Gouy, M. & Gautier, C. (1996). SeaView and Phylo_Win, two graphic tools for sequence alignment and molecular phylogeny. *Comput. Applic. Biosci.* **12**, 543-548.

Edited by J. Thornton

(Received 11 June 2001; received in revised form 4 September 2001; accepted 5 September 2001)

http://www.academicpress.com/jmb

Supplementary Material comprising six Tables is available on IDEAL