

Traffic Interchange

The Ypt/Rab Family and the Evolution of Trafficking in Fungi

José B. Pereira-Leal

Instituto Gulbenkian de Ciência, Apartado 14, P-2781-901
Oeiras, Portugal, jleal@igc.gulbenkian.pt

The evolution of the eukaryotic endomembrane system and the transport pathways of their vesicular intermediates are poorly understood. A common set of organelles and pathways seems to be present in all free-living eukaryotes, but different branches of the tree of life have a variety of diverse, specialized organelles. Rab/Ypt proteins are small guanosine triphosphatases with tissue-specific and organelle-specific localization that emerged as markers for organelle diversity. Here, I characterize the Rab/Ypt family in the kingdom Fungi, a sister kingdom of Animals. I identify and annotate these proteins in 26 genomes representing near one billion years of evolution, multiple lifestyles and cellular types. Surprisingly, the minimal set of Rab/Ypt present in fungi is similar to, perhaps smaller than, the predicted eukaryotic ancestral set. This suggests that the saprophytic fungal lifestyle, multicellularity as well as the highly polarized secretion associated with hyphal growth did not require any major innovation in the molecular machinery that regulates protein trafficking. The Rab/Ypt and other protein traffic-related families are kept small, not paralleling increases in genome size, in contrast to the expansion of such components observed in other branches of the tree of life, such as the animal and plant kingdoms. This analysis suggests that multicellularity and cellular diversity in fungi followed different routes from those followed by plants and metazoa.

Key words: evolution, fungi, protein traffic, Rab GTPases

Received 16 July 2007, revised and accepted for publication 25 October 2007, uncorrected manuscript published online 28 October 2007, published online 27 November 2007

A hallmark of eukaryotic cells is the presence of well-defined membrane-bound organelles, such as the Golgi apparatus and the endoplasmic reticulum (ER), within its cytosol. Vesicular trafficking pathways accomplish the movement of materials and information between cellular compartments, contribute to their biogenesis and are essential for the normal functioning of the cell. These pathways play housekeeping roles such as transport of extracellular proteins to the plasma membrane through the secretory pathway as well as a variety of specialized roles, such as pigmentation in melanocytes and antigen presentation in

immune cells (1). It is a system of fundamental biological interest underlying the molecular basis of cellular organization and specialization. It is also involved in a variety of human diseases. Malfunction of its components results in haemorrhagic disorders, immunodeficiency, mental retardation, blindness and others (2,3). Protein trafficking pathways are also frequently exploited by human pathogens to gain entry and to survive within host cells (4).

Despite the intense interest that protein trafficking has received, little attention has been devoted to its evolution. Fungi are well-suited organisms for its study. They are eukaryotic heterotrophs that digest food outside their bodies, secreting lytic enzymes to decompose complex materials into readily available nutrients (5). They form a sister kingdom to animals, estimated to have diverged approximately 0.9 billion years ago (6) (Figure 1). This kingdom is the better-sampled branch of the eukaryotic tree of life in terms of complete genome sequences, making them ideal for comparative genomic studies (6). The genomes of several closely related species and clusters of species as well as of distantly related species were sequenced, some specifically to increase phylogenetic coverage. The kingdom includes organisms that are of medical, agricultural or industrial interest as well as many model organisms. For these, there is a wealth of functional and morphological information available. Examples are the industrially important model yeast *Saccharomyces cerevisiae*, the human pathogen *Candida albicans* and the filamentous fungus *Neurospora crassa*. The kingdom Fungi encompass organisms with multiple lifestyles, such as mushrooms, molds and yeasts. In these, we find a diverse cell biology with multiple distinct cell types from the highly polarized hyphal cells displaying a fungal-specific vesicle organizing centre, the Spitzenkörper (7), to the appressorium in plant pathogens like *Magnaporthe grisea*, a cell type that generates internal pressures of up to 8 MPa (8,9) by manipulating the glycerol content of their vacuoles (10). All this diversity suggests specialization of the protein trafficking systems, a hypothesis I test here.

I will focus in this study on Rab/Ypt proteins. These form a large family of small GTP-binding proteins that regulate distinct trafficking pathways. Each Rab protein has a distinct sub-cellular localization, marking individual transport steps in different transport pathways, and many Rab proteins have specific patterns of tissue expression

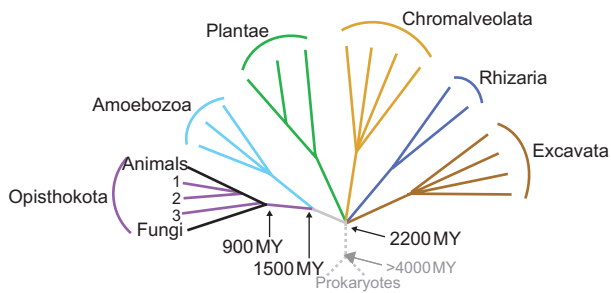


Figure 1: The eukaryotic tree of life adapted from (75). Taxa labelled 1, 2 and 3 are, respectively, Choanoflagellates, Ichthyosporidia and Nuclearioid amoebae. Estimates of divergence time, in millions of years (MY), were from (76,77) and are only for illustration purposes.

indicating that they operate in specialized, tissue-specific pathways. Rab proteins in their GTP-bound, active conformation recruit a multitude of effectors that are also specific for each organelle and pathway (11). Thus, Rabs and their effectors represent organelle- and pathway-specific markers, which can provide insights into which pathways are present in a given cell type.

Independent expansions of this protein family were observed in a variety of branches of the tree of life. For example, whereas the model unicellular organisms *S. cerevisiae* and *Schizosaccharomyces pombe* have only 11 and 7 described Rab/Ypt proteins, respectively (12), animals have near 70 (12,13), the closely related unicellular amoebas can have more than 100 proteins (14,15) and multicellular plants have near 60 (16) – see Figure 1 for evolutionary relationships of the taxonomic groups mentioned. It is clear that the expansion of the Rab family in mammals is associated with the emergence of a variety of specialized pathways and compartments. In other organisms, there is little functional information. In plants, the expansion of the family followed different routes, and the little functional information available suggests the emergence of novel subclasses and functionalities (16).

Not all branches of the eukaryotic tree display Rab expansions. For example, the Chromalveolata *Plasmodium falciparum* has only 11 Rabs (17). Furthermore, these expansions are not necessarily taxon specific – whereas *Trichomonas vaginalis* (excavate) has 65 Rabs (18), the human pathogen *Trypanosoma brucei*, also a unicellular excavate, has 16 Rabs (19). Thus, the relationship between expansions of the protein trafficking repertoire and the cell biology and lifestyle of these organisms is still unclear.

Here, I identify and annotate the complete Rab family in 26 fungi with a completely sequenced genome. Surprisingly, the diversity of fungal cellular types and near one billion years of evolution are not accompanied by diversification in the protein trafficking machinery in these organisms, even when I consider other protein trafficking-related families. My results suggest that the ancestor of fungi and animals

had already a complex endomembrane system and that the evolutionary history of protein trafficking in fungi is one of loss rather than that of gain.

Results and Discussion

The Rab family in fungi

I identified the full complement of Rab/Ypt proteins in 26 fungi with a completed genome sequence (Table 1). These organisms represent the Ascomycota, Basidiomycota and Microsporidia lineages, with a strong bias to Ascomycota (Figure 2). I identified these Rab/Ypt families using sequence similarity searches of a set of query Rab sequences from *S. cerevisiae* and *Homo sapiens* (12) against complete genome sequences. I used previously defined criteria (20), together with phylogenetic analysis, to annotate each Rab family mapping orthologues and paralogues. Figure 2 illustrates the evolutionary relationships of the fungi studied as well as the Rab proteins I identified and annotated in each organism.

The Rab family is very stable in size. Of the 26 genomes characterized, the family in 24 ranges from 8 to 12 proteins, whereas the genomes vary in size from less than 2000 genes to near 17 000 genes (*Phaeosphaeria nodorum* – 10

Table 1: Fungal species used in this study and reference of the complete paper describing the complete genome or to the web page of the genome project when a publication is not available

Species name	Genome sequence
<i>Aspergillus fumigatus</i>	(78)
<i>Cryptococcus neoformans</i>	(79)
<i>Aspergillus nidulans</i>	(80)
<i>Phanerochaete chrysosporium</i>	(81)
<i>Fusarium graminearum</i>	(82)
<i>Magnaporthe grisea</i>	(83)
<i>Phaeosphaeria nodorum</i>	(84)
<i>Neurospora crassa</i>	(6)
<i>Trichoderma reesei</i>	(85)
<i>Ustilago maydis</i>	(86)
<i>Candida albicans</i>	(87)
<i>Debaryomyces hansenii</i>	(88)
<i>Encephalitozoon cuniculi</i>	(21)
<i>Candida glabrata</i>	(88)
<i>Ashbya gossypii</i>	(89)
<i>Kluyveromyces lactis</i>	(88)
<i>Kluyveromyces waltii</i>	(90)
<i>Schizosaccharomyces pombe</i>	(91)
<i>Saccharomyces cerevisiae</i>	Saccharomyces Genome Database – (92)
<i>Saccharomyces bayanus</i>	(90,93)
<i>Saccharomyces castellii</i>	(93)
<i>Saccharomyces kluyveri</i>	(93)
<i>Saccharomyces kudriavzevii</i>	(93)
<i>Saccharomyces mikatae</i>	(90,93)
<i>Saccharomyces paradoxus</i>	(90)
<i>Yarrowia lipolytica</i>	(88)

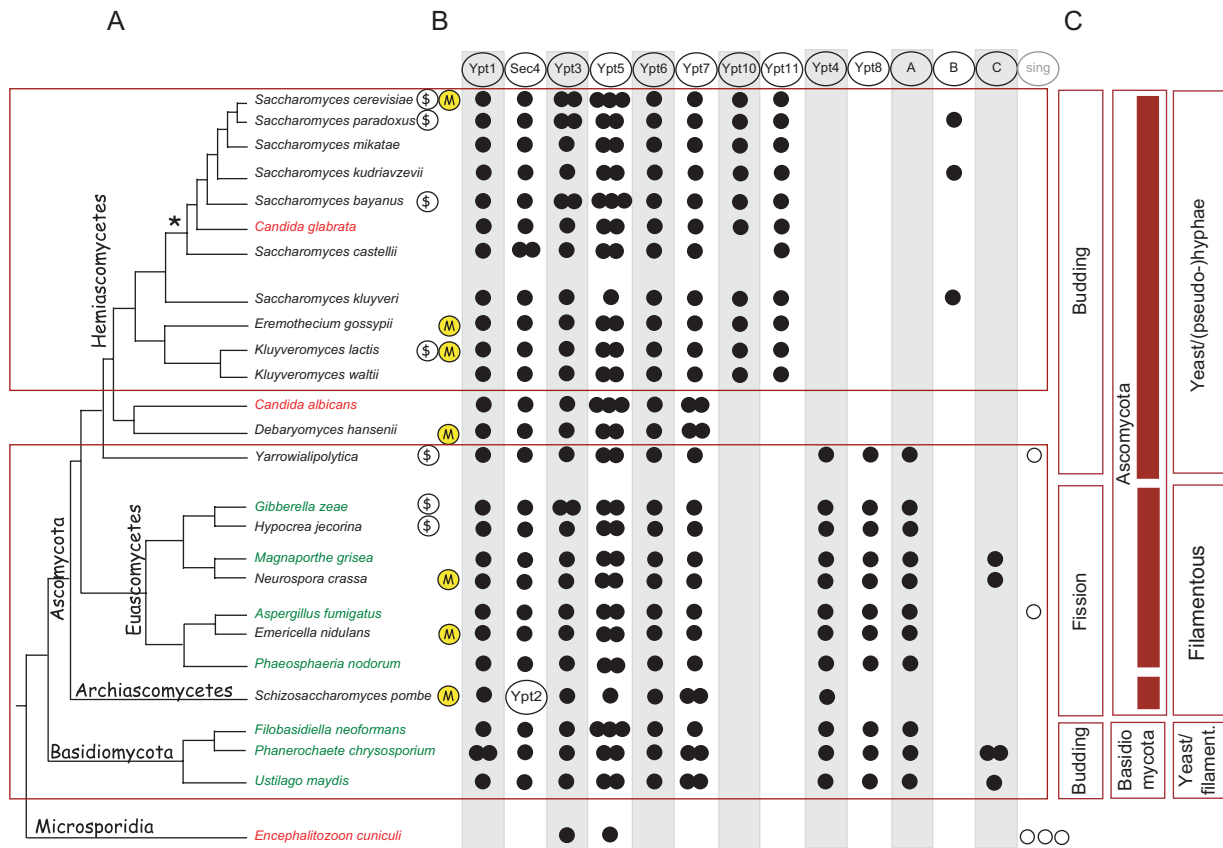


Figure 2: The Rab family in fungi. A) The tree illustrates the evolutionary relationships of the fungi studied here. The tree is loosely adapted from (70). Organisms in red and green are human and plant pathogens, respectively. The dollar and the M in front of the names highlight organisms of industrial interest and model organisms, respectively. The asterisk indicates a whole genome duplication event (37,94). B) Each column represents one Rab subfamily found in fungi. The filled circles in each column in front of the species name indicate the presence of the subfamily in that species. Multiple circles in each column indicate that the subfamily has multiple members in that organism. White circles indicate singleton Rab proteins – those that were not possible to assign to any subfamily. The red boxes overlaid on the columns with the Rab proteins highlight distinct Rab profiles. C) Boxes indicate cell division mode, taxonomy and lifestyle, illustrating that the Rab profiles are independent of these three factors.

Rabs). If we consider the number of distinct subfamilies instead of total number of Rabs, the picture remains unchanged. Furthermore, there is no obvious relationship between the number of Rab proteins and their distinct subfamilies with lifestyle (e.g. yeast versus filamentous), or taxonomical grouping. Thus, the size of the Rab family in fungi is kept at a relatively small and constant size rather than expanding in a neutral fashion.

Encephalitozoon cuniculi is the only fungus that shows clear correlation between Rab family size and lifestyle having undergone extensive gene loss. It belongs to a phylum that includes a diverse group of intracellular parasites, characterized by an extreme genome compaction and reduction (21,22), typical of the parasitic lifestyle. This is consistent with the loss of some trafficking functions.

A reduction in the Rab repertoire

With the catalogue of Rab proteins compiled and classified in the previous section, we can determine which Rabs are

necessary to make a fungus and which changes occurred in the Rab repertoire in the evolution of fungi from the ancestral eukaryote.

The set of Rab subfamilies common to all free-living fungi is composed of Ypt1, Sec4, Ypt3, Ypt5 and Ypt7 (Figure 3A). Their function and localization (in *S. cerevisiae*) is represented in Figure 3B and reviewed in 23–25. Briefly, Ypt1 localizes to ER and *cis* Golgi and works in the early steps of the secretory pathway, mediating ER–Golgi transport; Ypt3, like its orthologue Rab11, is involved in deep endocytic recycling (26) – earlier results suggest that it may also play a role in intra-Golgi traffic and in the budding of post-Golgi vesicles from the *trans* Golgi (27); Ypt5 mediates Golgi–endosome and plasma membrane–endosome transport; Ypt6 is involved in retrograde Golgi–ER and intra-Golgi transport; Ypt7 mediates vacuole fusion; and Sec4 mediates the delivery of *trans* Golgi network-derived vesicles into the bud, representing a form of polarized transport reminiscent of that mediated by its putative orthologue Rab8 (24,25).

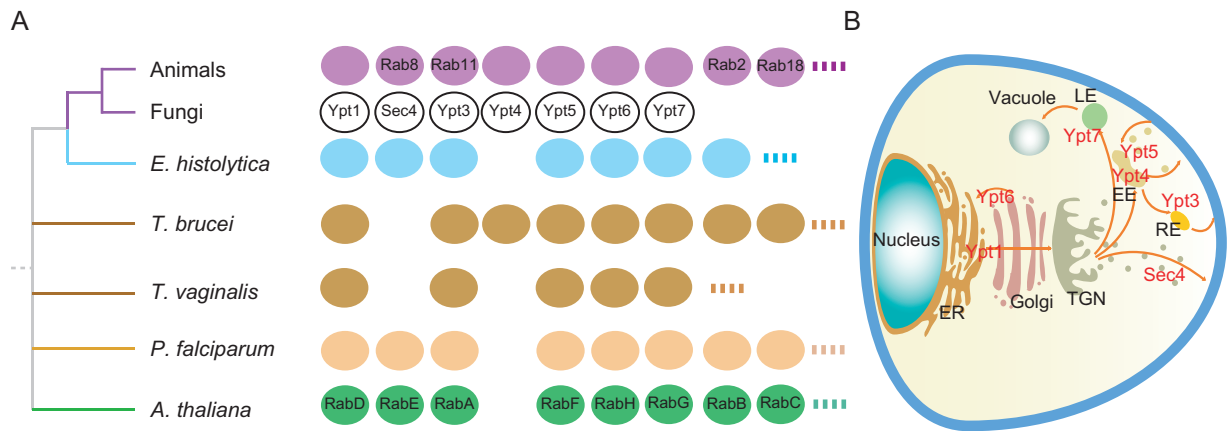


Figure 3: Rab repertoires in fungi and other eukaryotes. A) On the left is represented a simplified eukaryotic tree with the organisms for which the Rab family has been studied using the same colour scheme as that of Figure 1. Coloured circles are read vertically and represent putative orthologous subfamilies. Only those Rab subfamilies that are common to several lineages are shown, and the dots on the right indicate that the Rab family is larger than shown for that organism. When the putative orthologues are annotated differently in other species, the specific annotation is shown for that subfamily. B) Cartoon representing the localization and function of the common set of fungi Rabs. EE, early endosome; ER, endoplasmic reticulum; LE, late endosome; RE, recycling endosome; TGN, *trans* Golgi network. Full species names are *Entamoeba histolytica*, *Trypanosoma brucei*, *Trichomonas vaginalis*, *Plasmodium falciparum* and *Arabidopsis thaliana*.

This set of fungal Rab proteins is similar to the one common to most free-living organisms in which the Rab family was studied. In fact, *C. albicans* and *Debaryomyces hansenii* are restricted to this minimal set, indicating that such minimal Rab machinery is compatible with fungal life. Thus, the canonical secretory pathway appears to be sufficient for a saprophytic lifestyle. *Candida albicans* and *D. hansenii* emerge thus as particularly interesting model organisms in the study of protein trafficking as they may represent the minimal protein trafficking machinery that is compatible with free life (not parasitic). Furthermore, both organisms show dimorphic behaviour, existing both as yeast and as a filamentous form (hyphae). This strongly suggests that morphological diversity in fungi can be achieved without the need for elaborate protein trafficking pathways. Alternatively, trafficking diversity may have been achieved by Rab-independent means – below, I will show that other trafficking-related protein families display the same behaviour as the Rab family, which makes this hypothesis less likely.

It is surprising that fungal life did not require the specific gain of any Rab subfamily. It is equally surprising that it may have involved several secondary losses. The ancestral eukaryote is predicted to have included Rabs 1, 4, 5, 6, 7, 8 and 11 (28). This is based on the presence of orthologues of all these subfamilies in different eukaryotic crown groups. For example, even though Rab8 was not found in *T. brucei* and *T. vaginalis*, its presence in plants, alveolata and animals is easier to explain by their emergence in the ancestral eukaryote, followed by gene losses in the former two lineages. Otherwise, it would be necessary to invoke independent, convergent evolution in different lineages or lateral gene transfer between eukaryotes. On the same grounds, we can extend this set to include Rab18, present

in plants, alveolata, excavates and animals, and Rab2 present also in amoebas (Figure 3A). It is thus apparent that the evolution of fungi was accompanied by secondary losses of Rab2 and 18. Rab2 is involved in the early steps of the secretory pathway, mediating ER–Golgi transport as well as Golgi–ER retrograde transport (29), whereas Rab18 is involved in early endosome–plasma membrane recycling in polarized cells (29). Rab4, a regulator of rapid endocytic recycling was also lost in several branches, notably the Hemiascomycota.

In summary, the minimal set of Rabs in fungi revealed no fungi-specific proteins, indicating that the saprophytic lifestyle and multicellularity can be achieved with only a minimal Rab repertoire. It is the same Rab repertoire common to all studied free-living organisms, smaller than the one predicted to have emerged in the ancestral eukaryote.

Taxon-specific Rabs: Novel functions?

There are two complementary groups of taxon-specific subfamilies, i.e. proteins that appear in only a specific branch of the fungal tree but not in others. The first cluster is formed by Ypt10 and Ypt11 and is specific to a subset of Ascomycetes (Figure 2). Ypt10 appears to be involved in ER–Golgi transport. Overexpression results in growth defects and an overabundance of vesicular and tubular structures, suggesting alterations in the function of the Golgi apparatus (30). A variety of results from large-scale studies support such a role: Ypt10 has a reported genetic interaction with the Golgi Ypt6p exchange factor Ric1p–Rgp1p (31) and physical interactions with Yip3 (32). Yip3 localizes to COPII vesicles and is proposed to participate in ER–Golgi transport (33). Thus, Ypt10 appears to participate in a very early step in the secretory pathway.

The second protein, Ypt11, localizes to the sites of bud emergence, the emerging bud tip and the bud neck during M phase (34). Like mammalian Rab11 and Rab27, it interacts with a class V myosin, specifically with myo2 (34,35). Its exact function is unclear, but overexpression of Ypt11 leads to aberrant distribution of mitochondria and cell growth arrest (34), while Ypt11 deletion leads to deficient retention of mitochondria during cell division (35). Thus, Ypt11 is a regulator of mitochondrial distribution and contributes to segregation of mitochondria in mitotic cells. Recent analysis suggests that it may function in the traffic of mitochondrial retention factors from the mother cell to the bud tip (35). There is no functional information for YptB, which only appears in three organisms (Figure 2).

Candida albicans and *D. hansenii* are two yeasts that do not display any of the above-mentioned proteins. They have instead a duplication of Ypt7, which may implicate all or some of the proteins discussed above in endocytic or vacuolar fusion.

A second taxon-specific group of Ypt proteins is observed in Basidiomycota, Euscomycetes and partially in Archiascomycetes: Ypt4, 8 and A. Ypt4 was originally identified in *S. pombe*, but no functional information has been advanced for it. The closest homologue in mammals of these two subfamilies is Rab4, and it is possible that Ypt4 is the ortholog of Rab4. They are bidirectional best hits, but phylogenetic support is weak (not shown) with difficulty in resolving the fungal sequences from Rab4, Rab2 and Rab14, all members of a functional group. A functional group is composed of Rab proteins that are more similar within the group than to other Rabs that have related functions and that we speculate represent cases of duplication and specialization of the same basic Rab function (12). Rab4 mediates rapid endocytic recycling (36); so, it is tempting to speculate that Ypt4 is involved in similar recycling events.

Ypt8, identified in this study, is a new Ypt4-related subfamily, but is not sufficiently related to Ypt4 to be classed as its isoform (20). YptA is another conserved subfamily of unknown function. The abundance of plant pathogens in the taxonomic groups where Ypt4, 8 and A are present as well the presence of some industrially relevant organisms makes these three Rab proteins attractive targets for research.

In conclusion, there are two groups of taxon-specific proteins, but little functional information. They have complementary profiles, i.e. organisms have either one or the other but never both – they seem to be mutually exclusive. The functional significance of this is unclear, and so is the evolutionary route that gave rise to these profiles, as these profiles conflict with the taxonomical groupings (Figure 2). The Hemiascomycetes-specific proteins (Ypt10, 11 and B) may be involved in secretory events, whereas the other cluster (Ypt4, 8 and A) may be associated with the endocytic pathway. Surprisingly, these taxon-specific

proteins appear to have general functions rather than specialized ones.

Multiple, independent duplications

The relative constancy of the Rab family in fungi could be a consequence of this protein family being relatively immune to duplication or the net result of a dynamic duplication – loss process with a net result of little change. To gain an insight into this question, I analyzed the evolutionary relationships in subfamilies that have multiple isoforms in fungi and in animals (Figure 4).

Ypt5 has two or three isoforms in most fungi as well as in animals (e.g. Rab5a, b and c in *H. sapiens*). Ypt53 is a recent acquisition deriving from Ypt51 as a consequence of a whole genome (37). Ypt52 and Ypt54 display a complementary phylogenetic profile, i.e. when one appears, the other one does not. This type of complementary profile is typically the outcome of scenario where an early duplication of an ancestral gene was followed by asymmetric gene loss – some organisms lost one duplicate (e.g. Ypt52), whereas the other organisms lost the other duplicate (e.g. Ypt54). The multiple isoforms of Ypt5 that appear in Basidiomycota are all derived from lineage-specific duplications. Finally, Ypt51 is present in most Ascomycota and is likely a duplication of either Ypt52 or 54 at the base of the Ascomycota branch. This expansion is independent from the mammalian Rab5 subfamily (not shown). An independent expansion of the Rab5/Ypt5 subfamily was described previously between metazoa and the distantly related *T. brucei* (38). Thus, at the root of the fungi metazoan branch, I anticipate that there was one Rab5/Ypt5 that expanded independently in metazoa and fungi.

The Ypt3 subfamily, like Ypt5, expanded independently of the orthologous Rab11a and b/Rab25 subfamily (not shown) and displays multiple duplication events in its history (Figure 4). Ypt32 resulted from a whole genome duplication (37). Independent duplications led to the multiplicity in different branches, e.g. *Gibberella zeae* and *Candida glabrata*. Just as Ypt5, I also observe a complementary phylogenetic profile in Ypt31/32 versus Ypt3, suggestive of asymmetric gene loss after an initial duplication.

Thus, these two Rab subfamilies with multiple isoforms are the result of independent duplication processes. Such duplications appear to have happened frequently in the fungal genomes, but always converging on a similar number of isoforms. This means that the constancy of size of the Rab family in the evolution of fungi is not a consequence of this family being immune to duplication – in fact, there is frequent duplication but clearly a pressure to keep the whole family and the subfamilies at relatively constant sizes.

What drives the independent expansion of these subfamilies? Distinct functionalities were reported for Rab5 isoforms in mammals (39–41) and Rab5 isoforms in *T. brucei* (42). Similarly, Rab11a and b in mammals were

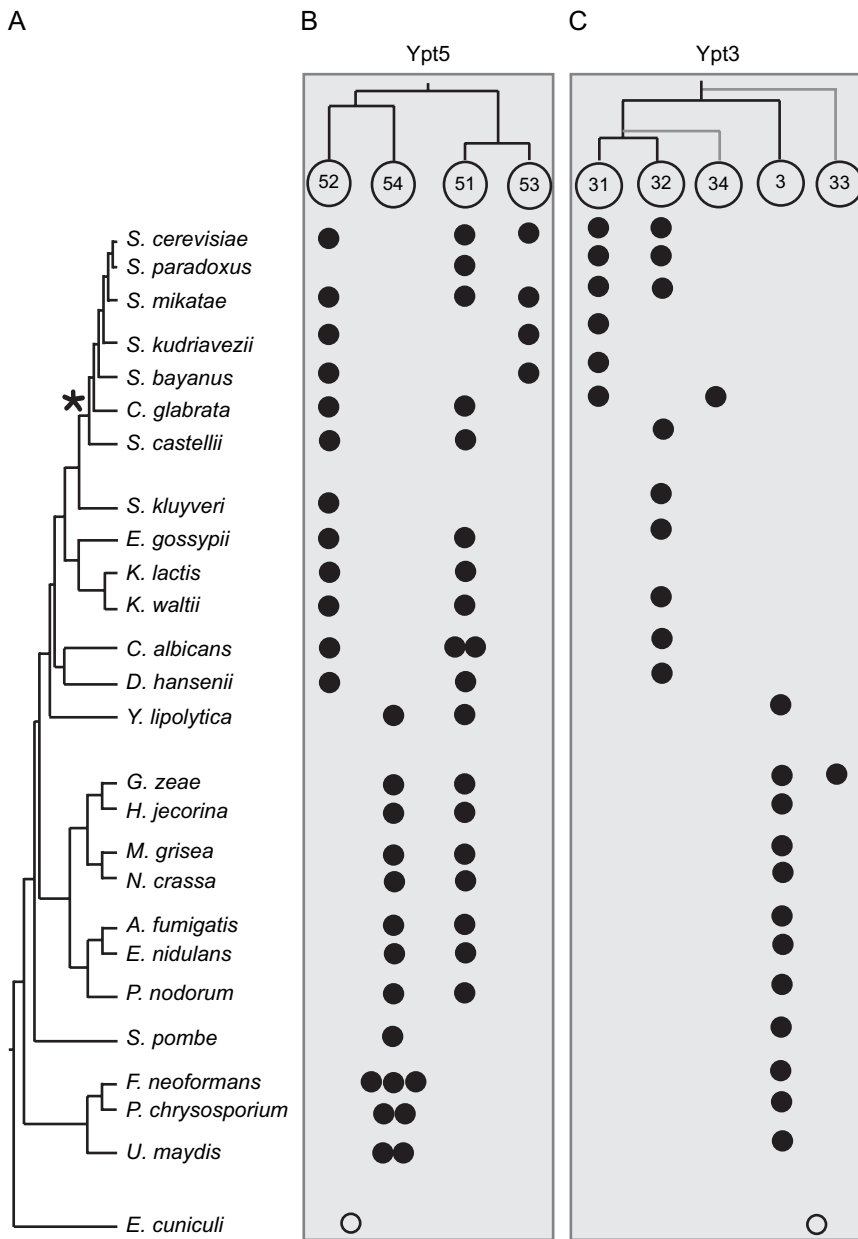


Figure 4: Duplication dynamics of Rab/Ypt subfamilies. A) The tree illustrates the evolutionary relationships of the fungi studied here. The tree is loosely adapted from (70). Panels B and C are the subfamilies Ypt5 and Ypt3, respectively. The trees on top illustrate the Neighbour-Joining distance trees of the members of the family. Each column represents one putative orthologous group, and a filled circle in that column indicates that the species has one member of that group. Multiple circles in each column indicate duplications within that putative orthology group. White circles represent proteins that cannot be assigned with confidence, and the asterisk indicates a whole genome duplication.

reported to have distinct function (43). Thus, taxon-specific functional specialization appears to be the driving force for these expansions. However, in *S. cerevisiae*, the results are less clear. Ypt31 and 32 appear to be redundant. Single deletions of Ypt31/32 are phenotypically neutral, but double knock out is lethal (27). Similarly, there seems to be some redundancy between Ypt51 and Ypt52 (44). Further supporting at least partial redundancy is the copurification of the three isoforms in the context of a large-scale proteomics screen (45). The role of Ypt53 is however less clear, and there seems to be a dosage effect limiting their genetic interactions – Ypt53 is expected to be the least abundant isoform, which would be consistent with a more specialized role (44). Thus, available evidence suggests that expansions of these subfamilies are associated with functional special-

ization, but the specific functions of the isoforms in the model organism *S. cerevisiae* remain elusive.

Other traffic-related protein families

I described above how the Rab family was kept at a relatively small, constant size in the evolution of fungi, independently of genome expansion, and that this is despite frequent duplications, which are balanced by gene losses. This is surprising as expansion of Rab families was observed in the context of cellular diversification, and fungi display diverse cell types and can exist as unicellular and/or multicellular forms. One possibility to account for this is that possible fungal innovations in protein trafficking were not mediated by expansions in the Rab family, but involved other molecular actors.

I investigated two other protein families, Rab GTPase activating proteins (RabGAPs) and SNAREs, capturing distinct aspects of Rab function and traffic in order to test the above hypothesis. These families are unequivocally identified by genome-wide structural assignments as defined in the superfamily database (46,47) (see *Data and methods* for details).

I first enumerated members of the RabGAP family. These proteins interact directly with activated Rab proteins at the membrane and increase their guanosine triphosphatase (GTPase) activity, thus switching off the signal that the activated Rab represented (48). Ten RabGAPs are predicted in *S. cerevisiae*, several with experimental support (49–54). There is scarce experimental information on the diversity of RabGAPs in other species, but sequence analysis suggests that this family expanded in parallel to Rab, for example, 52 homologues of yeast RabGAPs were detected in the human genome and 24 in the fly genome (48). Thus, RabGAPs represent a ‘positive control’ in this analysis. In fungi, I observed that this family is kept between 8 and 12 elements and its size is independent of genome size, but is correlated to the number of Rab proteins at $r = 0.49$ (Figure 5). Notable exceptions are *S. pombe* and *D. hansenii* that show an increased number of RabGAPs relative to Rabs. It is interesting to note that these organisms also have some of the most streamlined Rab families in fungi, suggesting that some diversification can occur through RabGAP family expansion.

Next, I enumerated SNARE proteins in fungal genomes. SNAREs are components of protein complexes that are critical for membrane fusion in the secretory pathway (55). They have also been extensively used as markers of protein trafficking diversity and evolution (13,55–59). In plants, an expansion in this family appears to correlate with the emergence of multicellularity (57). Such increase is not paralleled in metazoa, where the fly and the worm (13) display similar numbers of SNAREs to the unicellular *Leishmania major* (56) and *Plasmodium falciparum* (58). An expansion is, however, observed in the human genome, with nearly twice the numbers of fly and worm and appearing to correlate with an increase in the number of distinct tissues (13). Thus, SNAREs are not only central components of the trafficking machinery but also their numbers carry some evolutionary signal.

Did fungal diversity demand an increase in the numbers of SNAREs? Gupta and Heath (59) performed previously a detailed analysis of SNARE proteins in two complete and four draft fungal genomes. Their preliminary results suggested little variation in total number of SNAREs as well as no apparent correlation with genome size. I now address these questions in the 26 fungal genomes considered here. To do so, I considered two SNARE-related structural superfamilies: the ‘SNARE-like’ and the ‘t-SNARE’ superfamilies. The ‘SNARE-like’ superfamily includes R-SNAREs (60) as well as other proteins involved in

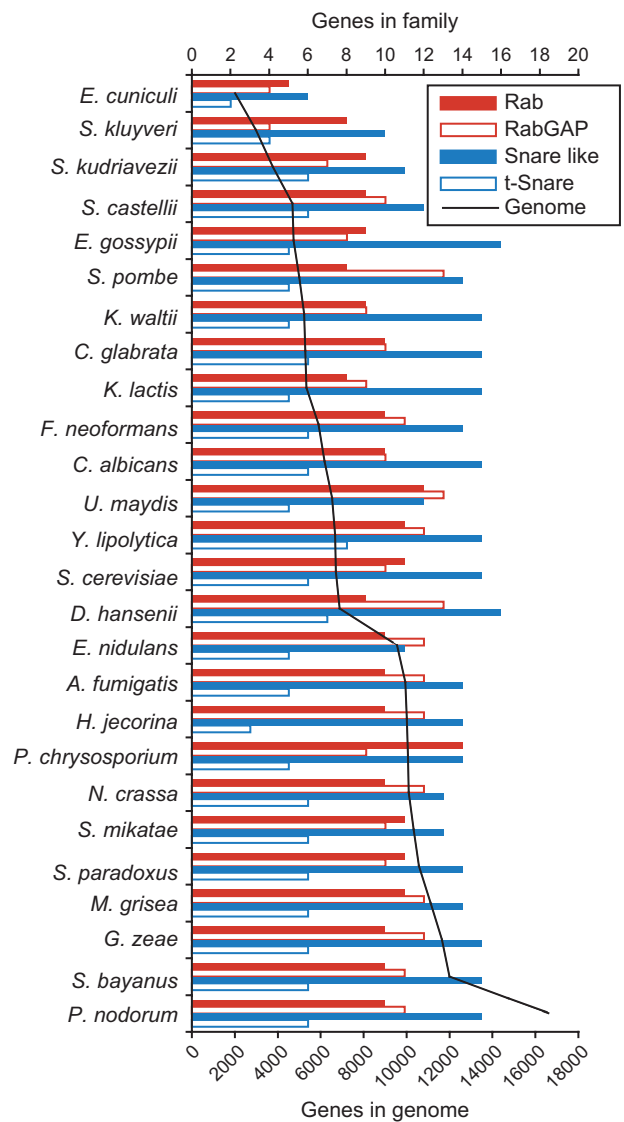


Figure 5: Protein trafficking-related families in fungi. The black line represents genome size measured in number of genes and is read on the bottom axis. The plot is sorted by ascending genome size. For each of the protein families studied, the number of genes that can be classified in that family using structural domain assignments in the Superfamily database (46) is plotted and read on the top axis. These protein families did not expand in the evolution of fungi, despite the growth in genome size as well as the multiple lifestyles these organisms developed.

distinct trafficking pathways and complexes. It includes components of the TRAPP complex, a complex present on the *cis* Golgi that acts prior to SNARE complex assembly (61,62). It also includes subunits of the adaptor protein complexes, which are components of the protein coats that associate with the cytoplasmic face of organelles and are involved in the formation of intracellular transport vesicles and cargo selection (63), as well as subunits of the coatomer complex (COPI). The ‘t-SNARE’ superfamily contains Q-SNAREs (60). SNARE proteins were identified

in a variety of organisms. *Saccharomyces cerevisiae* has 21 distinct proteins, *Caenorhabditis elegans* 23, *Drosophila melanogaster* 20 and *H. sapiens* 35 (13). In the fungi studied here, there is little variation in the number of either member of these superfamilies. *Yarrowia lipolytica* and *D. hansenii* have the highest number (23 distinct proteins) and *E. cuniculi* the lowest (eight proteins – Figure 5). Variations in either superfamily size are independent of genome size variation. They are in fact correlated with Rab numbers ($r = 0.51$), suggesting that Rab numbers may be an accurate indicator of protein trafficking complexity. However, simple enumerations of members of protein families need to be considered with care as they may miss functional information. For example, although *L. major* has comparable numbers of SNARE proteins to metazoa, it misses some of the subfamilies present in humans (e.g. SNAP-25 like) but has others that appear to be species or taxon specific (56).

I further investigated two other protein families involved in trafficking that were not expected to have expanded, representing thus a type of 'negative control' in this analysis. The first clathrin heavy chain – most organisms have a single gene, with the exception of the human genome that has a specialized muscle-specific isoform (64). As expected, I found that all fungi have a single gene for clathrin heavy chain, except for *E. cuniculi* that has none (not shown). The other family was class I myosins, motor proteins involved in a variety of trafficking events (65,66). In *S. cerevisiae*, two type I myosins (Myo3 and Myo5) localize to endocytic sites and are essential for endocytosis. They are thought to be partially redundant. In all fungi analyzed, I observed that only *C. glabrata* had two – all other fungi have a single myosin type 1 gene (not shown).

In summary, the traffic-related protein superfamilies considered here display the same trend in the evolution of fungi. That is, there is little variation of family size in the organisms studied, and this variation is uncorrelated to genome size, taxonomy or lifestyle or between any combinations of these families (not shown), but RabGAPS and SNARE numbers are correlated with Rab numbers. This observation supports the picture that emerged from the detailed analysis of Rab proteins: the diversification of the cellular biology of fungi does not appear to have required an expansion of the protein trafficking machinery. It is possible, however, that there were expansions in other protein families not investigated here. Only an exhaustive enumeration of all protein families involved in protein trafficking could clarify this point.

Reconstructing the evolutionary history of the Rab family

In order to understand how the different Rab profiles highlighted in Figure 2 could have emerged, I now attempt to reconstruct the evolutionary history of the Rab family in fungi. This serves the additional purpose of identifying those proteins likely to have been present in the ancestral fungus.

Evolutionary reconstructions always rely on a number of assumptions. The present one relies on the assumption that those proteins that exist in most organisms were present at the base of the tree. Supporting this assumption is the observation by Kunin and Ouzounis that in bacterial and archaeal genomes, gene losses are more frequent than gene gains (*gene genesis* in the original) and horizontal gene transfer (67). In eukaryotes, horizontal gene transfer is poorly understood. It is clearly possible as shown recently by Friesen et al. in the case of *Stagonospora nodorum*'s ToxA toxin that was laterally transferred from the fungus *Pyrenophora tritici-repentis* (68). However, there is no evidence that such events are frequent. Thus, in this reconstruction, I will consider horizontal gene transfer between eukaryotes unlikely. I consider scenarios that involve small number of events more likely than those that involve multiple ones.

Figures 2 and 4 illustrate that species-specific duplications are very frequent. The whole genome duplication that happened in the *Saccharomyces* lineage created the isoforms Ypt53 and Ypt32, but had otherwise little impact in the evolution of Rabs in fungi. Thus, segmental duplication appears to be the most important type of duplication in the evolution of Rab proteins and the generation of novel functions.

The most parsimonious scenario is the one presented in Figure 6. Ypt4, Ypt8 and YptA were present in the ancestor of Basidiomycota and Ascomycota and were retained as the two lineages separated. As the Ascomycota separated, seven gene losses could account for the present constellation of Rab proteins, accompanied by the appearance of three novel Rabs in the Saccharomycetaceae family (Ypt10, 11 and B) (Figure 6). As Ypt10, 11 and B are not monophyletic, they are likely independent duplicates of distinct genes.

This complicated scenario is required to account for the Rab family of *Y. lipolytica*, which although belonging to a Hemiascomycete appears to be more similar to the Rab family of Euascomycetes and Basidiomycota than that of other Hemiascomycetes. In fact, this early diverging Hemiascomycete has other features that are more akin to Euascomycetes than to Hemiascomycetes. One example is ribonuclease III processing signals present in all studied Hemiascomycetes except for *Y. lipolytica* (69). Any alternative scenario, in which Ypt4, 8 and A are not present at the base of the tree, requires several lateral gene transfer steps as well as several gene losses. In the absence of any other data and considering that the phylogenetic position of *Y. lipolytica* is not in doubt as it was recently confirmed by several groups (70,71), I must at present opt for the most parsimonious scenario.

It seems, thus, clear that the evolution of the Rab family in fungi is a story of frequent duplications within subfamilies, contrasting with several losses of whole subfamilies. The

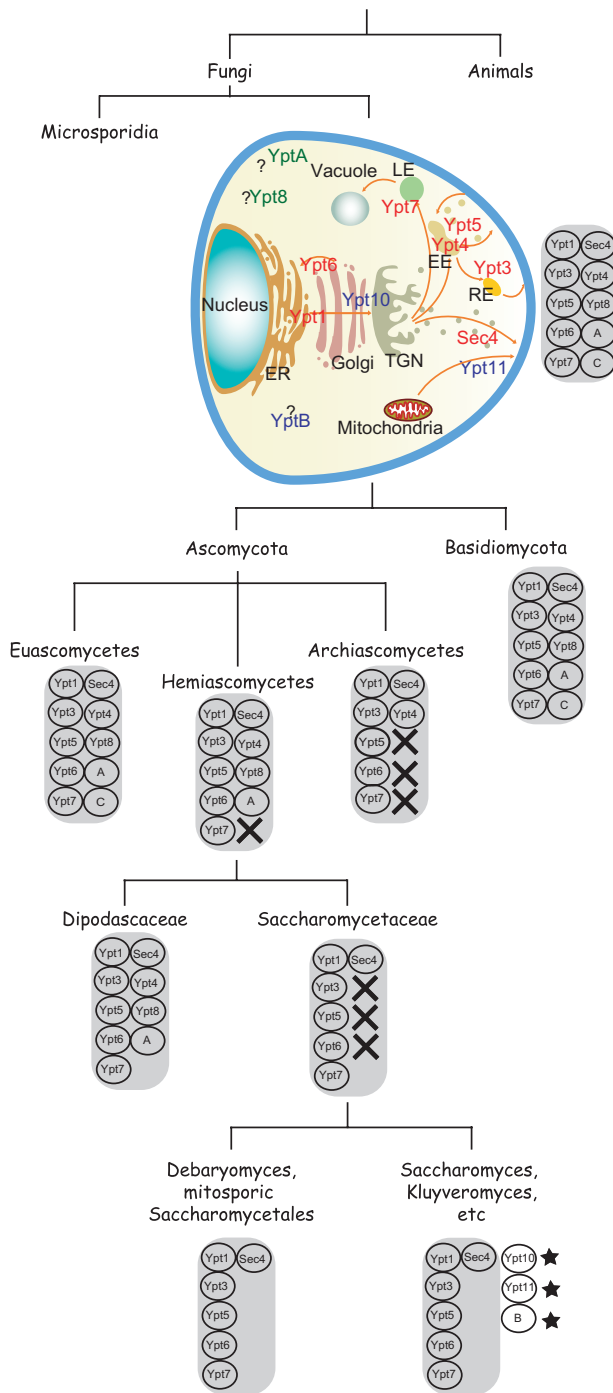


Figure 6: Evolution of Rabs in fungi. The whole figure represents the major evolutionary steps predicted in the emergence of extant Rab families in Fungi. It starts with a cartoon representing the putative cellular compartments of the ancestor of Basidiomycota and Ascomycota, with the Rab complement it is expected to have possessed. The core Rabs, present in all fungi, are shown in red, whereas the two groups of taxon-specific Rabs are shown in green and blue. Below this cartoon is a simplified evolutionary tree in which the grey rectangles show the constellation of Rab subfamilies present or predicted to be present in the taxon. Gene losses are represented by crosses (x), and gene gains, i.e. Rabs that appear for the first time, are represented outside the grey boxes and highlighted with a star (★). EE, early endosome; ER, endoplasmic reticulum; LE, late endosome; RE, recycling endosome; TGN, *trans* Golgi network. Question marks next to protein names indicate that no information on their localization and/or function is available.

includes several human and plant pathogens as well as organisms of scientific and industrial interest. I observed that there is little variation in the Rab repertoire of these species, despite the large evolutionary distance covered in the analysis and large range of genome sizes. Surprisingly, the minimal fungal Rab set is similar to, perhaps even smaller than, the set of Rabs speculated to have been present in the last eukaryotic common ancestor. This is surprising because it means that the specialized saprophytic lifestyle of fungi as well as the highly polarized secretion associated to hyphal growth did not require expansions of Rab repertoires, as could have been predicted by previous analysis (59). It also means that multicellularity in fungi was achieved without resource to Rab family expansions, unlike what was observed for two other types of multicellularity – the metazoa (12,13) and plants (12,16). An enumeration of SNARE proteins showed that SNARE numbers do not vary much in fungal genomes, which is also in contrast with the notion that multicellularity implies expansion of this family, at least in plants (57).

I found two clusters of completely taxon-specific Rab proteins displaying complementary phylogenetic profiles. Surprisingly, these proteins have mostly unknown functions and have had little experimental attention, even though their specificity makes them attractive targets in several fungi that are human or plant pathogens.

Overall, this analysis of Rab proteins, and of other protein trafficking-related protein families, suggests that cellular diversity in fungi did not evolve the same routes as taken in other branches of the tree of life, notably in animals and plants.

Data and Methods

The organisms used in this study are listed in Table 1 together with the reference to the genome source.

Sequence similarity searches were performed using Smith–Waterman algorithm, as implemented in a Time Logic Decypher™ machine, at a significance threshold of

ancestral of fungi had an elaborate endomembrane system that was likely similar to that of extant Basidiomycota and Euascomycota. It included all housekeeping functions that have been found in other distantly related eukaryotes.

Conclusions

I identified and annotated the Rab family in 26 fungi genomes. These represent three major fungi lineages,

$p < 0.02$, requiring an overlap of at least 40 residues, an alignment score of at least 100. All sequences were masked for low-complexity regions using Seg (72). Rab proteins were identified as all those that were more similar to a reference set of Rab proteins than to other small GTPases. The Rab reference set included *S. cerevisiae* and *H. sapiens* families and were compiled in (12). All sequences are available from the author's website: <http://eao.igc.gulbenkian.pt/CGL/>.

Classification of Rab proteins was based on protein sequence analysis and performed in three stages. The first was to define putative orthologous sequences as bidirectional best hits using the human and yeast families as reference (12). All other sequences were classified as related to their best reference sequence hit. We then aligned all the reference sequences using CLUSTAL W 1.83 (73) and calculated a Neighbour-Joining, bootstrapped tree using the same program and used these trees to assign a classification to the sequences not automatically assigned in the previous step. All situations that were not clear by the combination of these two approaches were further investigated using phylogenetic analysis by maximum likelihood, implemented in TREE-PUZZLE (74).

Annotated Rab families from other organisms were obtained from the following references: the animals *Homo sapiens*, *Drosophila melanogaster* and *Caenorhabditis elegans* and the plant *Arabidopsis thaliana* from (12). The four unicellular eukaryotes, representing three crown eukaryotic groups and their source are *Trypanosoma brucei* (19), *Trichomonas vaginalis* (18), *Entamoeba histolytica* (15) and *Plasmodium falciparum* (17).

The identification of other protein families was made using Superfamily automated genome-wide domain assignments. The Superfamily database compiles structural domain assignments that are based on profile Hidden Markov Models derived from known protein structures (46,47). The Superfamily accession codes of the families we considered are RabGAP (47923), SNARE like (64356), t-SNARES (47661), clathrin heavy chain (50989) and myosin type II (52540 and 50044).

Acknowledgment

I thank Alistair Hume for helpful comments as well as Mónica Bettencourt-Dias, Rui Martinho and Zita Santos for critical reading of the manuscript.

References

- Blott EJ, Griffiths GM. Secretory lysosomes. *Nat Rev Mol Cell Biol* 2002;3:122–131.
- Aridor M, Hannan LA. Traffic jam: a compendium of human diseases that affect intracellular transport processes. *Traffic* 2000;1:836–851.
- Aridor M, Hannan LA. Traffic jams II: an update of diseases of intracellular transport. *Traffic* 2002;3:781–790.
- Gruenberg J, van der Goot FG. Mechanisms of pathogen entry through the endosomal compartments. *Nat Rev Mol Cell Biol* 2006;7:495–504.
- Naglik J, Albrecht A, Bader O, Hube B. *Candida albicans* proteinases and host/pathogen interactions. *Cell Microbiol* 2004;6:915–926.
- Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 2003;422:859–868.
- Virag A, Harris SD. The Spitzenkörper: a molecular perspective. *Mycol Res* 2006;110:4–13.
- Howard RJ, Ferrari MA, Roach DH, Money NP. Penetration of hard substrates by a fungus employing enormous turgor pressures. *Proc Natl Acad Sci U S A* 1991;88:11281–11284.
- Bechinger C, Giebel KF, Schnell M, Leiderer P, Deising HB, Bastmeyer M. Optical measurements of invasive forces exerted by appressoria of a plant pathogenic fungus. *Science* 1999;285:1896–1899.
- Weber RW, Wakley GE, Thines E, Talbot NJ. The vacuole as central element of the lytic system and sink for lipid droplets in maturing appressoria of *Magnaporthe grisea*. *Protoplasma* 2001;216:101–112.
- Grosshans BL, Ortiz D, Novick P. Rabs and their effectors: achieving specificity in membrane traffic. *Proc Natl Acad Sci U S A* 2006;103:11821–11827.
- Pereira-Leal JB, Seabra MC. Evolution of the Rab family of small GTP-binding proteins. *J Mol Biol* 2001;313:889–901.
- Bock JB, Matern HT, Peden AA, Scheller RH. A genomic perspective on membrane compartment organization. *Nature* 2001;409:839–841.
- Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, Nozaki T, Suh B, Pop M, Duchene M, Ackers J et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 2005;433:865–868.
- Saito-Nakano Y, Loftus BJ, Hall N, Nozaki T. The diversity of Rab GTPases in *Entamoeba histolytica*. *Exp Parasitol* 2005;110:244–252.
- Rutherford S, Moore I. The *Arabidopsis* Rab GTPase family: another enigma variation. *Curr Opin Plant Biol* 2002;5:518–528.
- Quevillon E, Spielmann T, Brahimi K, Chattopadhyay D, Yeramian E, Längsley G. The *Plasmodium falciparum* family of Rab GTPases. *Gene* 2003;306:13–25.
- Lal K, Field MC, Carlton JM, Warwicker J, Hirt RP. Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*. *Mol Biochem Parasitol* 2005;143:226–235.
- Ackers JP, Dhir V, Field MC. A bioinformatic analysis of the RAB genes of *Trypanosoma brucei*. *Mol Biochem Parasitol* 2005;141:89–97.
- 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;301:1077–1087.
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretailade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 2001;414:450–453.
- Slamovits CH, Fast NM, Law JS, Keeling PJ. Genome compaction and stability in microsporidian intracellular parasites. *Curr Biol* 2004;14:891–896.
- Lazar T, Gotte M, Gallwitz D. Vesicular transport: how many Ypt/Rab-GTPases make a eukaryotic cell? *Trends Biochem Sci* 1997;22:468–472.
- Zerial M, McBride H. Rab proteins as membrane organizers. *Nat Rev Mol Cell Biol* 2001;2:107–117.
- Stenmark H, Olkkonen VM. The Rab GTPase family. *Genome Biol* 2001;2:REVIEWS3007.

26. Furuta N, Fujimura-Kamada K, Saito K, Yamamoto T, Tanaka K. Endocytic recycling in yeast is regulated by putative phospholipid translocases and the Ypt31p/32p-Rcy1p pathway. *Mol Biol Cell* 2007; 18:295–312.
27. Benli M, Doring F, Robinson DG, Yang X, Gallwitz D. Two GTPase isoforms, Ypt31p and Ypt32p, are essential for Golgi function in yeast. *EMBO J* 1996;15:6460–6475.
28. Field MC, Gabernet-Castello C, Dacks JB. Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. In: Jékely G, editor. *Eukaryotic Membranes and Cytoskeleton: Origins and Evolution*. Landes Bioscience Publishing; 2007, pp. 84–110.
29. Segev N. Ypt and Rab GTPases: insight into functions through novel interactions. *Curr Opin Cell Biol* 2001;13:500–511.
30. Louvet O, Roumanie O, Barthe C, Peypouquet MF, Schaeffer J, Doignon F, Crouzet M. Characterization of the ORF YBR264c in *Saccharomyces cerevisiae*, which encodes a new yeast Ypt that is degraded by a proteasome-dependent mechanism. *Mol Gen Genet* 1999;261:589–600.
31. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 2005;123:507–519.
32. Calero M, Collins RN. *Saccharomyces cerevisiae* Pra1p/Yip3p interacts with Yip1p and Rab proteins. *Biochem Biophys Res Commun* 2002; 290:676–681.
33. Otte S, Belden WJ, Heidman M, Liu J, Jensen ON, Barlowe C. Erv41p and Erv46p: new components of COPII vesicles involved in transport between the ER and Golgi complex. *J Cell Biol* 2001;152:503–518.
34. Itoh T, Watabe A, Toh EA, Matsui Y. Complex formation with Ypt11p, a rab-type small GTPase, is essential to facilitate the function of Myo2p, a class V myosin, in mitochondrial distribution in *Saccharomyces cerevisiae*. *Mol Cell Biol* 2002;22:7744–7757.
35. Boldogh IR, Ramcharan SL, Yang HC, Pon LA. A type V myosin (Myo2p) and a Rab-like G-protein (Ypt11p) are required for retention of newly inherited mitochondria in yeast cells during cell division. *Mol Biol Cell* 2004;15:3994–4002.
36. Roberts M, Barry S, Woods A, van der Sluijs P, Norman J. PDGF-regulated rab4-dependent recycling of alpha5beta3 integrin from early endosomes is necessary for cell adhesion and spreading. *Curr Biol* 2001;11:1392–1402.
37. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 2004;428:617–624.
38. Field H, Farjah M, Pal A, Gull K, Field MC. Complexity of trypanosomatid endocytosis pathways revealed by Rab4 and Rab5 isoforms in *Trypanosoma brucei*. *J Biol Chem* 1998;273:32102–32110.
39. Barbieri MA, Roberts RL, Gumusboga A, Highfield H, Alvarez-Dominguez C, Wells A, Stahl PD. Epidermal growth factor and membrane trafficking. EGF receptor activation of endocytosis requires Rab5a. *J Cell Biol* 2000;151:539–550.
40. Alvarez-Dominguez C, Stahl PD. Increased expression of Rab5a correlates directly with accelerated maturation of *Listeria monocytogenes* phagosomes. *J Biol Chem* 1999;274:11459–11462.
41. Alvarez-Dominguez C, Stahl PD. Interferon-gamma selectively induces Rab5a synthesis and processing in mononuclear cells. *J Biol Chem* 1998;273:33901–33904.
42. 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: 9529–9539.
43. Lapierre LA, Dorn MC, Zimmerman CF, Navarre J, Burnette JO, Goldenring JR. Rab11b resides in a vesicular compartment distinct from Rab11a in parietal cells and other epithelial cells. *Exp Cell Res* 2003;290:322–331.
44. Singer-Kruger B, Stenmark H, Dusterhoft A, Philippsen P, Yoo JS, Gallwitz D, Zerial M. Role of three rab5-like GTPases, Ypt51p, Ypt52p, and Ypt53p, in the endocytic and vacuolar protein sorting pathways of yeast. *J Cell Biol* 1994;125:283–298.
45. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002;415:180–183.
46. Gough J, Chothia C. SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res* 2002;30:268–272.
47. Wilson D, Madera M, Vogel C, Chothia C, Gough J. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res* 2007; 35:D308–D313.
48. Bernards A. GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and *Drosophila*. *Biochim Biophys Acta* 2003;1603:47–82.
49. Albert S, Gallwitz D. Two new members of a family of Ypt/Rab GTPase activating proteins. Promiscuity of substrate recognition. *J Biol Chem* 1999;274:33186–33189.
50. Albert S, Gallwitz D. Msb4p, a protein involved in Cdc42p-dependent organization of the actin cytoskeleton, is a Ypt/Rab-specific GAP. *Biol Chem* 2000;381:453–456.
51. Albert S, Will E, Gallwitz D. Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J* 1999;18:5216–5225.
52. De Antoni A, Schmitzova J, Trepte HH, Gallwitz D, Albert S. Significance of GTP hydrolysis in Ypt1p-regulated endoplasmic reticulum to Golgi transport revealed by the analysis of two novel Ypt1-GAPs. *J Biol Chem* 2002;277:41023–41031.
53. Will E, Albert S, Gallwitz D. Expression, purification, and biochemical properties of Ypt/Rab GTPase-activating proteins of Gyp family. *Methods Enzymol* 2001;329:50–58.
54. Strom M, Vollmer P, Tan TJ, Gallwitz D. A yeast GTPase-activating protein that interacts specifically with a member of the Ypt/Rab family. *Nature* 1993;361:736–739.
55. Jahn R, Scheller RH. SNAREs—engines for membrane fusion. *Nat Rev Mol Cell Biol* 2006;7:631–643.
56. Besteiro S, Coombs GH, Mottram JC. The SNARE protein family of *Leishmania major*. *BMC Genomics* 2006;7:250.
57. Sanderfoot A. Increases in the number of SNARE genes parallels the rise of multicellularity among the green plants. *Plant Physiol* 2007; 144:6–17.
58. Ayong L, Pagnotti G, Tobon AB, Chakrabarti D. Identification of *Plasmodium falciparum* family of SNAREs. *Mol Biochem Parasitol* 2007; 152:113–122.
59. Gupta GD, Brent Heath I. Predicting the distribution, conservation, and functions of SNAREs and related proteins in fungi. *Fungal Genet Biol* 2002;36:1–21.
60. Fasshauer D, Sutton RB, Brunger AT, Jahn R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc Natl Acad Sci U S A* 1998;95:15781–15786.
61. Lowe M. Membrane transport: tethers and TRAPPs. *Curr Biol* 2000; 10:R407–R409.
62. Sacher M, Barrowman J, Schieltz D, Yates JR 111, Ferro-Novick S. Identification and characterization of five new subunits of TRAPP. *Eur J Cell Biol* 2000;79:71–80.
63. Boehm M, Bonifacino JS. Adaptins: the final recount. *Mol Biol Cell* 2001;12:2907–2920.

64. Wakeham DE, Abi-Rached L, Towler MC, Wilbur JD, Parham P, Brodsky FM. Clathrin heavy and light chain isoforms originated by independent mechanisms of gene duplication during chordate evolution. *Proc Natl Acad Sci U S A* 2005;102:7209–7214.
65. Sokac AM, Schietroma C, Gundersen CB, Bement WM. Myosin-1c couples assembling actin to membranes to drive compensatory endocytosis. *Dev Cell* 2006;11:629–640.
66. Sun Y, Martin AC, Drubin DG. Endocytic internalization in budding yeast requires coordinated actin nucleation and myosin motor activity. *Dev Cell* 2006;11:33–46.
67. Kunin V, Ouzounis CA. The balance of driving forces during genome evolution in prokaryotes. *Genome Res* 2003;13:1589–1594.
68. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* 2006;38:953–956.
69. Chanfreau G. Conservation of RNase III processing pathways and specificity in hemiascomycetes. *Eukaryot Cell* 2003;2:901–909.
70. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 2006;443:818–822.
71. Liu YJ, Hodson MC, Hall BD. Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of kingdom Fungi inferred from RNA polymerase II subunit genes. *BMC Evol Biol* 2006;6:74.
72. Wootton JC, Federhen S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* 1996;266:554–571.
73. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
74. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 2002;18:502–504.
75. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. *Curr Biol* 2004;14:R693–R696.
76. Hedges SB. The origin and evolution of model organisms. *Nat Rev Genet* 2002;3:838–849.
77. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* 2005;15:1620–1631.
78. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulsen R et al. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 2005;438:1151–1156.
79. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA, Allen JE, Bosdet IE, Brent MR, Chiu R, Doering TL et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 2005;307:1321–1324.
80. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scaccocchio C, Farman M, Butler J et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 2005;438:1105–1115.
81. Martinez D, Larrondo LF, Putnam N, Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F, Coutinho PM, Henrissat B, Berka R, Cullen D, Rokhsar D. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* 2004;22:695–700.
82. Broad-Institute. Fusarium Graminearum Sequencing Project. 2006. http://www.broad.mit.edu/annotation/genome/fusarium_graminearum/Home.html
83. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 2005;434:980–986.
84. Broad-Institute. Stagonospora Nodorum Sequencing Project. 2006. http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum/Info.html
85. Joint-Genome-Institute. *Trichoderma Reesei* Draft Genome v0.1. 2006. <http://genome.jgi-psf.org/Trire2/Trire2.home.html>
86. Broad-Institute. *Ustilago Maydis* Sequencing Project. 2006. http://www.broad.mit.edu/annotation/genome/ustilago_maydis/Home.html
87. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S. The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci U S A* 2004;101:7329–7334.
88. Dujon B, Sherman D, Fischer G, Durrrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthonard V, Babour A et al. Genome evolution in yeasts. *Nature* 2004;430:35–44.
89. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD, Philippsen P. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 2004;304:304–307.
90. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 2003;423:241–254.
91. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature* 2002;415:871–880.
92. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, Weng S, Botstein D. SGD: *saccharomyces* genome database. *Nucleic Acids Res* 1998;26:73–79.
93. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* 2003;301:71–76.
94. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 1997;387:708–713.