Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins

Carme Gabernet-Castello[a,b], Amanda J. O'Reilly[a], Joel B. Dacks[b], and Mark C. Field[a]
[a]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom; [b]Department of Cell Biology, University of Alberta, Edmonton, AB T6G 2H7, Canada

ABSTRACT Rab GTPases serve as major control elements in the coordination and definition of specific trafficking steps and intracellular compartments. Rab activity is modulated in part by GTPase-activating proteins (GAPs), and many RabGAPs share a Tre-2/Bub2/Cdc16 (TBC)–domain architecture, although the majority of TBC proteins are poorly characterized. We reconstruct the evolutionary history of the TBC family using ScrollSaw, a method for the phylogenetic analysis of pan-eukaryotic data sets, and find a sophisticated, ancient TBC complement of at least 10 members. Significantly, the TBC complement is nearly always smaller than the Rab cohort in any individual genome but also suggests Rab/TBC coevolution. Further, TBC-domain architecture has been well conserved in modern eukaryotes. The reconstruction also shows conservation of ancestral TBC subfamilies, continuing evolution of new TBCs, and frequent secondary losses. These patterns give additional insights into the sculpting of the endomembrane system.

INTRODUCTION

Although internal compartmentalization has been described in several prokaryotic lineages, the endomembrane features of eukaryotic cells have attained a much higher level of diversity, representing a major evolutionary transition (Stanier, 1970). The compartments encompass both endosymbiont-derived organelles, specifically chloroplasts and mitochondria, and endogenously arising compartments, including the endoplasmic reticulum (ER), Golgi complex, and endosomes. Endomembrane specializations are associated with adaptation to distinct circumstances, including environment, parasitism, and differentiation of function in multicellular organisms.

Complete genome data are available from a substantial range of eukaryotes, which allows reconstruction of multiple evolutionary processes underpinning derivation of the endomembrane system. These mechanisms include an ancient, surprisingly sophisticated core system in the last eukaryotic common ancestor (LECA), together with lineage-specific innovations/expansions and very frequent secondary losses (Dacks and Field, 2007; Diekmann et al., 2011; Elias et al., 2012). Endomembrane compartments are dynamic, and despite a steady-state composition, this state is generated via flux through specific pathways and interactions with cytoskeletal and other structural organizers. Vesicle transport is mediated by collaborations between cohorts of paralogous families, including the soluble N-ethylmaleimide–sensitive factor attachment protein receptors, protocoatomer coats, tethering complexes, and small GTPases of the ARF and Rab families. All of these families contribute to defining organelles, as well as to controlling specificity and rate of transport through individual pathways (Cai et al., 2007). Rab proteins are central in these processes, acting as both signaling and switching molecules (Brighouse et al., 2010). The propensity of Rabs to interact with a wide range of partners provides a major component to the integration between individual transport steps; this promiscuity/integration is extended by the many proteins interacting with multiple Rabs. For example, SAND1/Mon1 coordinates late endosomal trafficking by modulating interactions of both Rab5 and Rab7 (Poteryaev et al., 2010).

The specific control of small GTPase activity is largely mediated by GTPase-activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs). GAPs serve to increase GTPase activity by contributing

residues to the active site (Albert *et al.*, 1999; Rak *et al.*, 2000; Pan *et al.*, 2006), promoting conversion of the GTP to GDP form. The resultant conformational change alters the ability to bind downstream effectors. For Rab proteins, the vast majority of identified GAPs contain a Tre-2/Bub2/Cdc16 (TBC) Rab-binding domain (Richardson and Zon, 1995; Neuwald 1997), although frequently the TBC is associated with additional domains, presumably serving to add functional diversification to the family. The TBC family has been poorly studied, and those few members of the family that have been functionally characterized in any detail have been studied in either Metazoa or *Saccharomyces cerevisiae*. Several studies analyzed TBC interactions with putative Rab partners by high-throughput and targeted approaches, but these are also restricted to Metazoa and Fungi (Itoh *et al.*, 2006; Brett *et al.*, 2008; Costanzo *et al.*, 2010; Will and Gallwitz, 2001). Many disease-related processes, including responses to infection and proliferative or degenerative disorders, are associated with defective trafficking pathways, and multiple TBC RabGAP mutations have emerged as important contributors to pathological states. For example, mutations in *Homo sapiens* TBC1D23 are associated with cancer (El-Bchiri *et al.*, 2008; De Arras *et al.*, 2012), and a TBC1D20 mutation is associated with Parkinson's disease (Cooper *et al.*, 2006; Yeger-Lotem *et al.*, 2009). The situation is further complicated by the fact that not all TBCs may have a primary role as a Rab GAP, whereas, conversely, there are potentially Rab GAPs that do not contain a TBC domain. The absence of robust systematic analysis or data beyond animals and Fungi is also challenging.

Rab proteins are exploited widely for comparative cell biology on account of their high levels of specificity in subcellular localization. When orthologous Rabs have been localized in distantly related organisms their locations, and by inference functions, are frequently maintained; for example, Rab5 is targeted to early endosomes, and Rab1 is involved in early exocytic steps at the ER/ERGIC in mammals, plants, protists, and amoebae (Field *et al.*, 1998; Dhir *et al.*, 2004; Lee *et al.*, 2004; Khurana *et al.*, 2005; Pinheiro *et al.*, 2009). This conservation provides both experimental possibilities for pathway-specific manipulation and with genome-based predictions of presence, absence, or complexity for a given pathway based on the Rab complement within individual organisms. More recently, two molecular evolutionary approaches for reconstructing Rab evolutionary history have been described (Diekmann *et al.*, 2011; Elias *et al.*, 2012). Both demonstrate that the LECA was highly complex and that secondary losses, paralogous expansions, and evolution of novel Rab isoforms accompanied subsequent evolution. We also introduced ScrollSaw, a strategy for phylogenetic reconstruction of highly paralogous protein families (Elias *et al.*, 2012).

Here we seek to further understanding of TBC diversity and evolution. We perform a detailed phylogenetic analysis of the TBC family using a variant of ScrollSaw with the aim of providing a systematic nomenclature and evolutionary reconstruction of the TBC family. We systematically classify TBC family members across the eukaryotes, allowing comparisons of experimental data for TBCs between organisms. The analysis also facilitates identification of novel TBCs in various eukaryotic lineages and establishes a TBC complement in the LECA. In addition, we determine TBC essentiality in trypanosomes, a highly divergent model organism.

RESULTS
Identification of TBC domain–containing open reading frames in representative eukaryotic genomes
We generated a nonredundant data set of 591 predicted protein sequences containing the TBC domain from an array of 26 eukary-

otic lineages, spanning the taxonomic breadth of eukaryotes (Supplemental Table S1).

We initially compared representation of TBC-containing proteins with Rab GTPases for 21 species also sampled in a recent study (Elias *et al.*, 2012; Figure 1A). The TBC repertoire is expanded in several taxa, including metazoans, vascular plants, and several unicellular organisms, which correlates with an expanded Rab complement, suggesting major innovations to these trafficking systems. As genome size increases, the number of TBC and Rab open reading frames (ORFs) also tends to increase. Superimposed upon this are several clear examples of expansions (e.g., *Entamoeba histolytica*, *Trichomonas vaginalis*) and losses (e.g., *Chlamydomonas reinhardtii*) of the Rab repertoire. Therefore increased Rab and TBC ORF frequency per genome correlates with, but is not entirely explained by, genome coding complexity.

We next considered the ratios between Rab and TBC ORFs in each species (Figure 1B). The number of TBC genes is most often lower than that of the Rab genes, suggesting that either some TBCs act on more than one Rab or a non-TBC GAP is operating. This finding is further represented by the Rab/TBC ORF ratio, where two broad categories were apparent. A ratio <2.0 was restricted to Opisthokonta, unicellular Archaeplastida, and representatives of the Stramenopile, Alveolate, and Rhizarian (SAR) clade, plus kinetoplastids. The Amoebozoa, multicellular plants, and excavates *T. vaginalis* and *Naegleria gruberi* have Rab/TBC ratios >2.0. Each of these last-named sets is known to possess highly expanded Rab complements (Rutherford and Moore, 2002; Lal *et al.*, 2005; Saito-Nakano *et al.*, 2005; Carlton *et al.*, 2007; Fritz-Laylin *et al.*, 2010). Nonetheless, the general correlation between TBC and Rab frequencies suggests that as the Rab family expands, the TBC family is under pressure to follow this expansion. Both variation in total number of TBC genes in a genome and the ratio of Rab to TBC genes within a supergroup suggest that TBC family evolution is highly dynamic.

Evolutionary relationships of the TBC complement of humans and yeast
To understand the evolution of the TBC family, we undertook a phylogenetic analysis. Initial reconstructions using standard phylogenetic strategies resulted in very little resolution between and among the TBC genes (data not shown). To overcome this, we implemented ScrollSaw, a strategy we recently introduced for analysis of Rab genes, which similarly cannot be resolved by existing approaches (Elias *et al.*, 2012). Initially, we defined the TBC complement in *H. sapiens* and *S. cerevisiae* in phylogenetic terms, that is, to assign clades, as a basis for classification of TBCs, as most functional information is derived from these organisms. We reasoned that robustly related sequences represent versions of the same TBC subclass, whereas sequences that failed to resolve into clades represent either members of TBC subclasses missing from humans and yeast or are unique TBC sequences. From our analysis (Figure 2), we identified seven robust and two moderately supported clades. In total, we found 21 putative TBC subclasses (as defined in *Materials and Methods*). Because our reconstruction used only the TBC domain itself, we retrieved domain information for the entire TBC-containing ORF from the Pfam database and looked for shared domain architecture as an independent assessment of the quality of the reconstruction. We found multiple examples of retention of such shared architectures, suggesting that both the reconstruction based on TBC domains alone was accurate and, moreover, that overall TBC protein architecture is retained in these species.
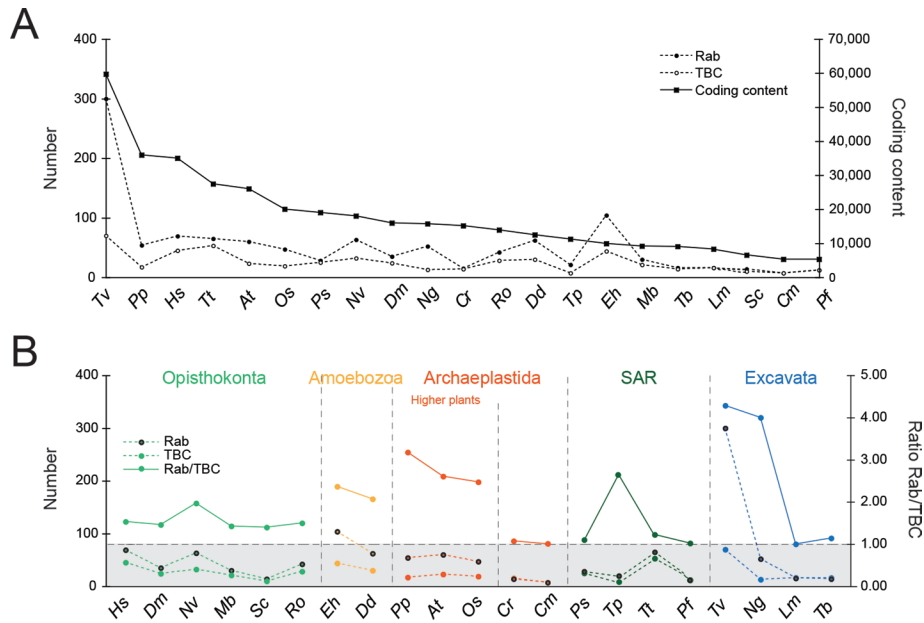
FIGURE 1: Representation of Rab and TBC coding sequences in selected eukaryotic genomes. (A) Counts of Rab and TBC coding sequences in 21 completed genomes ranked according to descending number of predicted open reading frames. Dashed lines represent Rab (black symbols) and TBC (white symbols) gene counts per taxon (reads are on the *y*-axis, left); the solid line indicates coding content as measured by number of genes in a genome (reads are on the *y*-axis, right). Note that as genome size increases, the number of Rab ORFs also tends to increase but with clear outliers. Further, the number of Rab ORFs nearly always exceeds the number of TBC ORFs per genome. (B) Rab and TBC ORF numbers broken out by eukaryotic supergroup. Dashed lines represent Rab (black symbols) and TBC (colored symbols) ORF counts (reads are on the left *y*-axis) for representatives of the Opisthokonta (light green), Amoebozoa (yellow), Archaeplastida (red), SAR (dark green), and Excavata (blue). The solid line represents the ratio of Rab to TBC ORFs (right *y*-axis, right). Shaded region corresponds to a Rab/TBC ratio of <1.0. Two-letter abbreviations of Linnaean names are defined in Supplemental Table S3.

Reconstructed complement of ancestral and lineage-specific TBCs in extant eukaryotes supports a complex LECA and subsequent innovation

TBC sequences of representative taxa for each supergroup were assembled and analyzed by Bayesian, PhyML, and RAxML methods (Supplemental Figures S1–S5). The number of well-supported clades was assessed for each supergroup-specific data set, with the opisthokonts exhibiting the highest number and archaeplastids the lowest (Supplemental Table S2). TBCs for each supergroup were provisionally assigned to one of the 21 subclasses defined earlier.

To test the subclass assignments from the supergroup-specific analyses and therefore assign TBCs as ancient or lineage specific, we reconstructed a phylogeny using the least divergent representative of each reconstructed TBC subclass clade in each supergroup-specific data set, identified by shortest branch length or, when possible, the most significant *E*-value to the *H. sapiens* or *S. cerevisiae* orthologue. The resulting topology gave moderate to robust reconstruction of 14 TBC subclasses at supergroup level by all methods (Figure 3). Given that placement of the eukaryotic root is unclear (Roger and Simpson, 2009) but the relationship of the major eukaryotic supergroups is relatively resolved (Burki *et al.*, 2008; Hampl *et al.*, 2009), we opted to assign evolutionary origins for TBC clades based on their presence in each of the five sampled supergroups. TBC subclasses B, D, F, M, and Q were identified in four or more supergroups by phylogenetics (Figure 3). TBC subclasses E, I, and N were identified and confirmed by phylogenetics to be present in three supergroups (Figure 3), but a putative representative was identified in each case in a fourth supergroup either based on BLAST

or by less stringent node support criteria (Figure 4A). Finally, TBC-RootA (absent from Opisthokonta and the Archaeplastida), and TBC-L (missing in the Amoebozoa and the Archaeplastida) were identified in three groups, with broad taxonomic spread. Consequently, these 10 subclasses are all presumed as ancient, that is, present in LECA, but with deduced losses in various taxa. In addition, TBC-K only just failed to be supported in Figure 3 but in other analyses obtained sufficient support values (0.99/60/55; data not shown) uniting taxa from four supergroups; consequently, we also tentatively consider this as an ancient subclass. TBC-G is found in the Opisthokonta, Amoebozoa, and Excavata and thus could be counted as ancient, pending the root of the eukaryotic tree. Furthermore, TBC-H was found in Holozoa and Kinetoplastida, with a single putative homologue in the SAR representative *P. sojae* also identified, based on BLAST criteria only (data not shown). This TBC might represent an ancient subclass or would need to be explained by some other mechanism (e.g., horizontal gene transfer between hosts and parasites). Thus these three TBC subclasses may well have also been present in the LECA, but deductions regarding their origins are more speculative pending further data.

TBC-A, O, P, and T, detected in both the Opisthokonta and the Amoebozoa, are likely to have appeared with the division of the Unikonta (Stechmann and Cavalier-Smith, 2003; Roger and Simpson, 2009), also recently renamed Amorphea (Adl *et al.*, 2012). Those TBC subclasses found in a single supergroup, as verified by their lack of affinity with any other TBC sequences in the pan-eukaryotic phylogeny, are presumed to be lineage specific. The results of these analyses are summarized in Figure 4A.
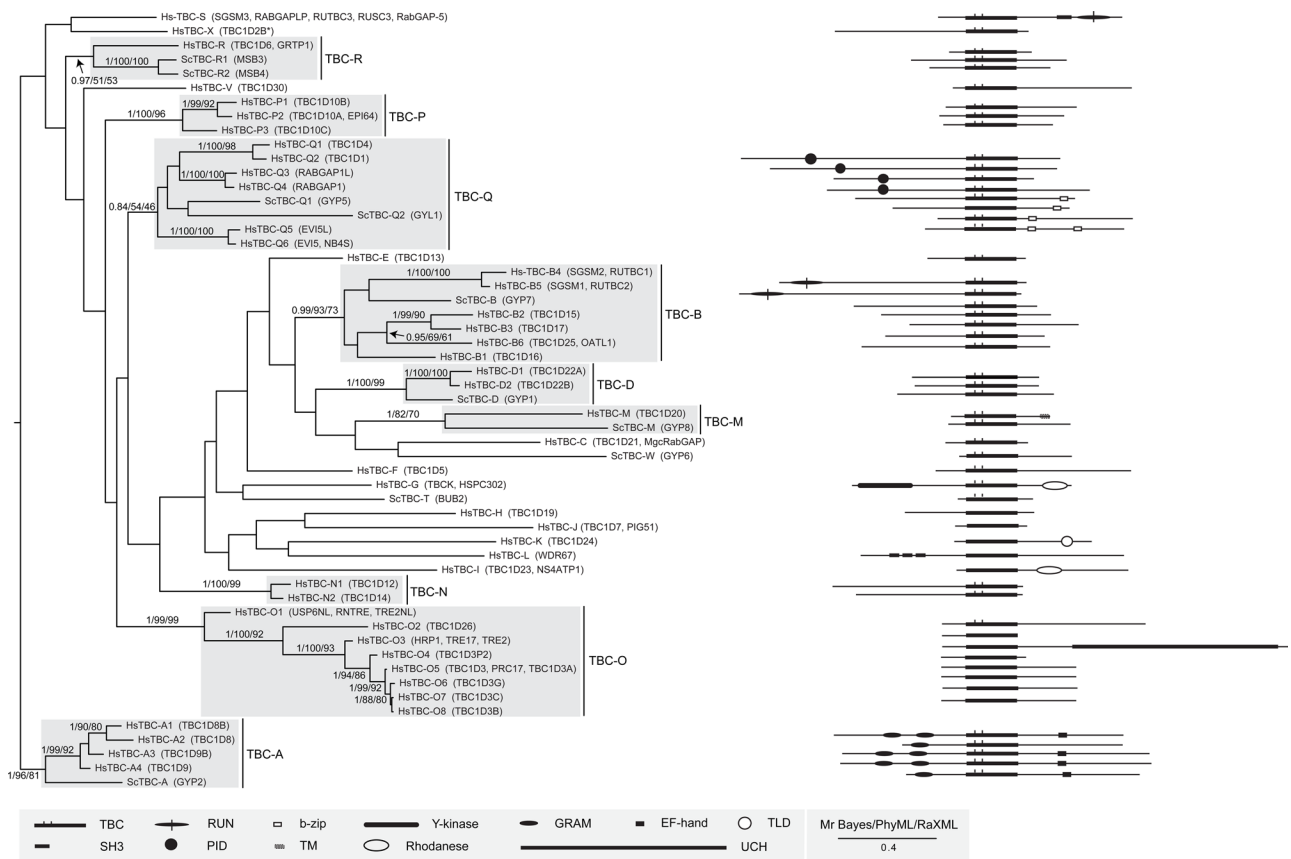
FIGURE 2: The TBC RabGAP complement in *H. sapiens* and *S. cerevisiae*. A phylogenetic tree was generated from the shared TBC domains from the *H. sapiens* and *S. cerevisiae* genomes (see *Materials and Methods*), and the best Bayesian topology shown. In this phylogeny and all subsequent analyses, support values are given for any nodes reconstructed with values of >0.8 posterior probability (MrBayes) and 50% by one of the two maximum-likelihood methods (PhyML or RAxML). TBC clade names are given using a useful convention for new taxa. This connects with clades and preexisting naming and has consequences for domain architecture; existing names in humans and yeast are in parentheses. The asterisk denotes the fact that, although only the TBC1D2B sequence was analyzed, preliminary calculations (data not shown) robustly place the TBC1D2A variant in the TBC-X clade as well. Gray boxes highlight clades that are reconstructed with robust or moderate statistical support. Scale bar, 0.4 change on average per site. Right, domain architecture of RabGAPs represented schematically; only domains with significant *E*-values recognized by Pfam are shown. For clarity the TBC domains are aligned (filled rectangles), and catalytic residues (R and Q fingers) are indicated as small tick marks where present.

For most TBC classes the R and Q fingers and surrounding amino acids appear to be conserved across supergroups (Supplemental Figure S6). However, some TBC domains lack this conventional catalytic amino acid pair (Albert *et al.*, 1999; Rak *et al.*, 2000; Pan *et al.*, 2006). Significantly, this variance in R and Q conservation also appears conserved between supergroups, suggesting that the identity of the amino acids corresponding to the R and Q positions occurred early in evolution. Specifically, TBC-L (WDR67), TBC-K (TBC1D24), TBC-H (TBC1D19), TBC-I (TBC1D23), and TBC-J (TBC1D7) are missing one or both of the R or Q residues in the majority of supergroups. However, in other cases, substitution of R or Q was restricted to some species or paralogues of a TBC class, for example, some paralogues of the TBC-D class in the Opisthokonta and the Excavata. This suggests that the status of these residues is evolutionarily flexible in some cases. In other examples domain architecture is not conserved; for example the β-zip domain is present in TBC-Q from opisthokonts only. However, TBC-G homologues in opisthokonts, excavates, and Amoebozoa all share a conserved tyrosine kinase domain N-terminal to their TBC domain, whereas TBC-M sequences from all supergroups

share a C-terminal *trans*-membrane domain. In addition, TBC-I sequences from amoebozoans, opisthokonts, and excavates share a rhodanese domain downstream of the TBC domain, and the TBC-K homologues in these supergroups with SAR taxa share a C-terminal Toll-like domain. Finally, TBC-L sequences, found only in the excavates and opisthokonts, share EF-hand domains. These instances of architectural conservation suggest conserved functions between orthologues and provide additional confidence in the phylogenetic reconstruction. Further, these observations suggest stability of overall domain architecture, at least for a subpopulation of TBCs.

Figure 4B shows the percentage of predicted TBC proteins in each of the sampled organisms as ancient, lineage-specific, or singleton, that is, sequences that failed to resolve into clades. Two trends are observed. Those organisms that have been traditionally difficult to analyze by molecular phylogenetic methods have, not surprisingly, a high proportion of singletons; this is almost certainly due to sequence divergence, leading to difficulties with placing the TBC sequences within reconstructed clades with any degree of support. However, in comparing the proportion of
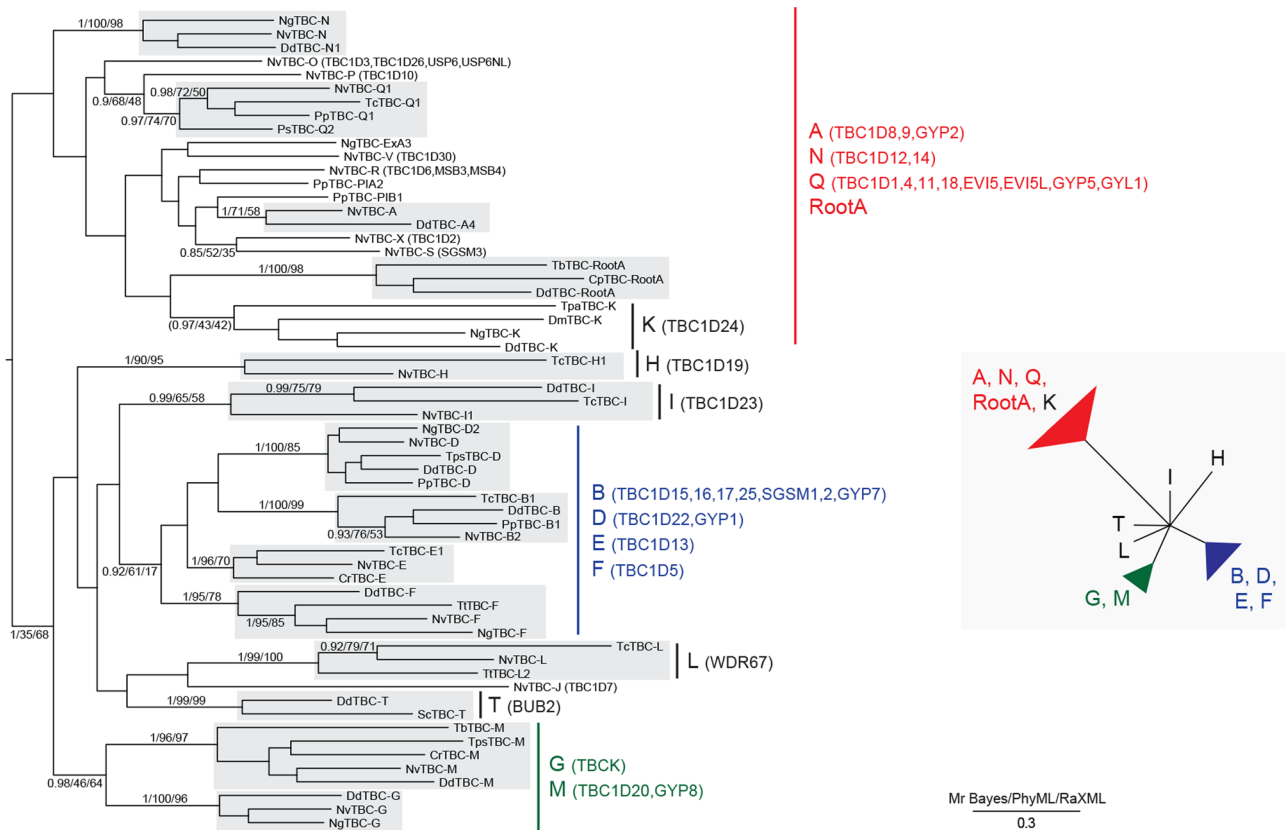
FIGURE 3: Phylogenetic reconstruction of TBC genes across the eukaryotes. A phylogenetic tree was generated from the least divergent representative from each supergroup for each putative ancestral TBC clade and putative lineage-specific TBC clades (see *Materials and Methods* and Supplemental Figures S1–S5). Next to the clade membership signifier are current names for the human and yeast orthologues in parentheses. Gray boxes highlight clades reconstructed with statistical support, and all values of >0.8 posterior probability and 50% bootstrap support are given. Scale bar, 0.3 change on average per site. The phylogeny indicates a sophisticated complement of TBCs in the ancestral eukaryote. The support values for base node of the TBC-K clade (shown in brackets) did not meet the criteria for inclusion; in other data sets this clade was reconstructed (data not shown). Inset shows schematically the interclade relationships reconstructed for the various TBC subfamilies.

ancestral to lineage-specific TBCs, the opisthokonts and archae-plastids stand out as possessing a larger proportion of lineage-specific TBCs as compared with the other supergroups. We also note, in the supergroup-specific analyses, that TBC-B and TBC-Q were expanded in these organisms. This additional complexity potentially opens up opportunities for evolution of increased trafficking complexity.

Our phylogenetic analysis (Figure 3) also provides some resolution between TBC subclasses, with one clade uniting TBC-G and M, a second containing TBC-B, D, E, and F, and a third encompassing ancient subclasses TBC-A, K, N, Q, and RootA together with many putative supergroup-specific TBCs. Although these three clades were only weakly supported by maximum-likelihood (ML) methods, they were consistently reconstructed and obtained high Bayesian posterior probability support. The resolution in the tree backbone is encouraging, but biological interpretation of these data awaits better functional definition of the encompassed TBC clades.

Analysis of interactions between TBCs and Rabs in trypanosomes, a model eukaryote distantly related to yeast and humans

To gain some functional insights into the importance of TBCs in divergent taxa, we analyzed TBC interactions and functions in *Trypanosoma brucei*, a member of the Excavata (Supplemental

Figures S7 and S8). The trypanosome proteins lack many of the accessory domains of animal and fungal TBCs. A yeast two-hybrid screen of all trypanosome Rab/Rab-related genes (Ackers *et al.*, 2005) and TBC proteins was performed; to increase interaction affinity, the predicted catalytic Q was mutated to A for TBCs (Pan *et al.*, 2006), and the GTPases were mutagenized to the predicted GTP-locked (Q-to-A) forms. The screen identified 23 potential interactions out of 196 pairs, that is, ~12%, and several Rabs (TbRab5B, 6, 21, 23, and X1) and TBCs (TbTBC-D3, E, L, ExC, and RootA) failed to interact, suggesting that this interactome was significantly under-populated (Supplemental Figure S7). Further, compared with *S. cerevisiae* or *H. sapiens* (Itoh *et al.*, 2006), many interactions were inconsistent between orthologue pairs. We also assessed expression levels for all trypanosome TBC and Rab mRNAs by quantitative real-time PCR to identify possible correlations between life cycle–dependent expression and interaction (Supplemental Figure S7). There was sparse evidence for coexpression of detected interactions; only TbTBC-M was up-regulated in one life stage, along with predicted interactors TbRab1B, 7, and 11. Overall, these data suggested that the yeast two-hybrid analysis fails to capture the full Rab-TBC interactome, at least for trypanosomes.

We selected a cohort of trypanosome TBCs for RNA interference (RNAi) knockdown (Supplemental Figure S8); TbTBC-Q1 and Q2 with differing predicted connectivity (TbTBC-Q1, six; TbTBC-Q2,
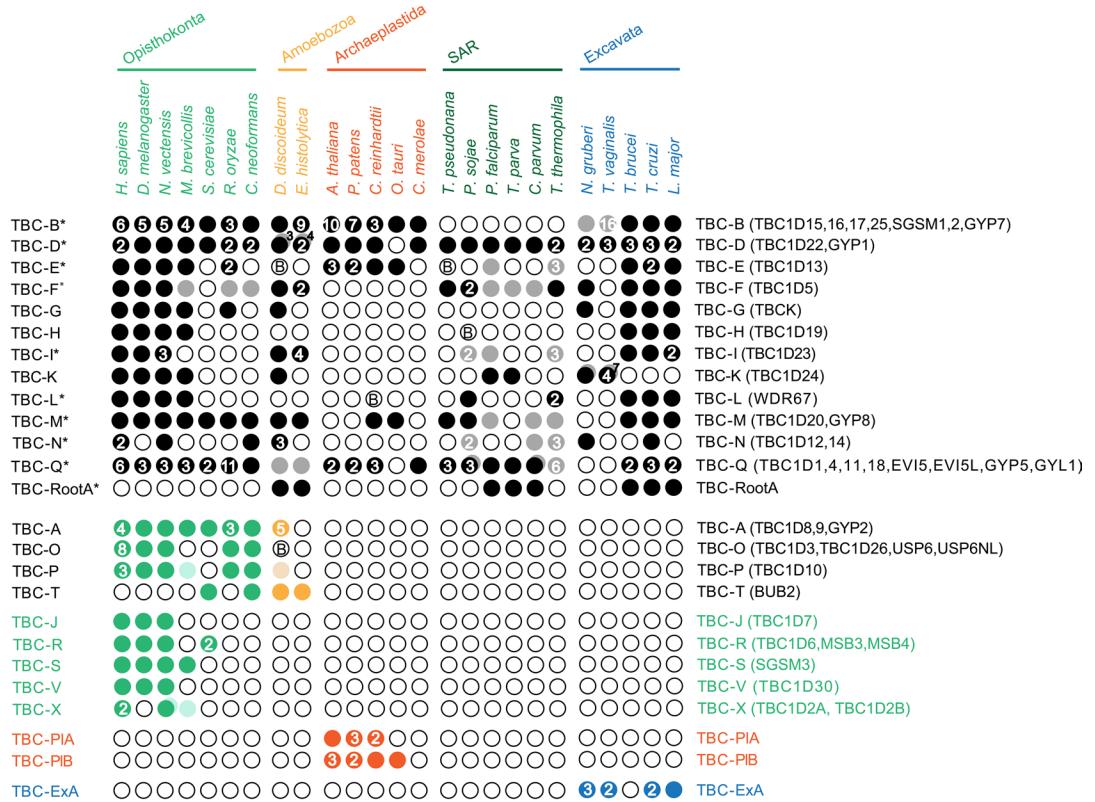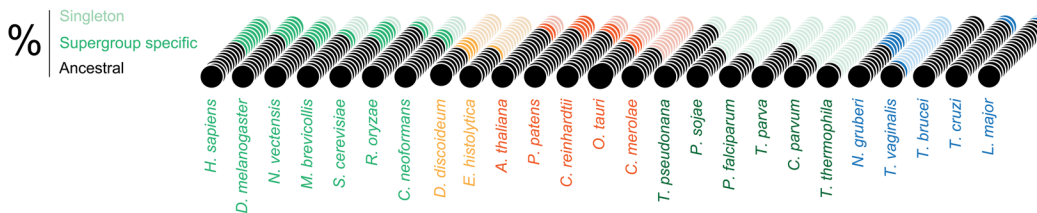
FIGURE 4: Ancestral and lineage-specific TBCs across the Eukaryota. (A) Distribution of TBC domain–containing genes by taxon and clade, shown as a dot plot in which filled circles in rows depict presence of a TBC type and columns represent the taxa analyzed. Colors indicate supergroup-specific entries, and black indicates pan-eukaryotic entries. Those designated with an asterisk denote the 10 TBC subclasses confidently reconstructed as present in the LECA. On the right next to the clade membership signifier are current names for the human and yeast orthologues in parentheses. Decreased-opacity circles represent presence of ancestral or supergroup TBCs based on the same supergroup phylogenies with less stringent criterion (cut-off value 0.8 posterior probability). Circles with B denote sequences that we classified based on retrieval of the human or yeast homologue with *E*-values of >E-50. Empty circles depict absence, based on the supergroup phylogenies (cut-off value 0.8 posterior probability and 50% bootstrap support). Numbers within circles indicate paralogue counts for a given TBC subfamily in a genome if more than one paralogue was recovered. Superscripts denote the presence of additional paralogues assigned by lower evidence. (B) For each taxon included in the analysis the proportion of ancestral lineage-specific and singleton (i.e., unassigned) TBCs as resolved by the phylogenetic analysis are shown. Dots are colorized, with black indicating ancestral, colored indicating supergroup specific, and faded color indicating singleton.

one), and developmental expression (TbTBC-Q1 and TBC-B constitutive, and TbTBC-Q2 down-regulated in bloodstream stage). This selection also addressed potential redundancy between TbTBC-Q1 and TbTBC-Q2, which are paralogues. RNAi against TbTBC-Q1 and TbTBC-Q2 produced significant proliferative defects, with essentially no effect from TbTBC-B. However, the effect of these knockdowns was inconsistent with their predicted interactions from the yeast two-hybrid analysis. The data do indicate that

TbTBC-Q1 and TbTBC-Q2 are individually required for normal cell functions and that, due to the additive effect of RNAi against both genes, are likely nonredundant.

DISCUSSION

TBC Rab-GAPs are essential regulators of cellular functions, specifically in modulating Rab activity and hence intracellular transport. Clear indications that TBCs are central participants in membrane
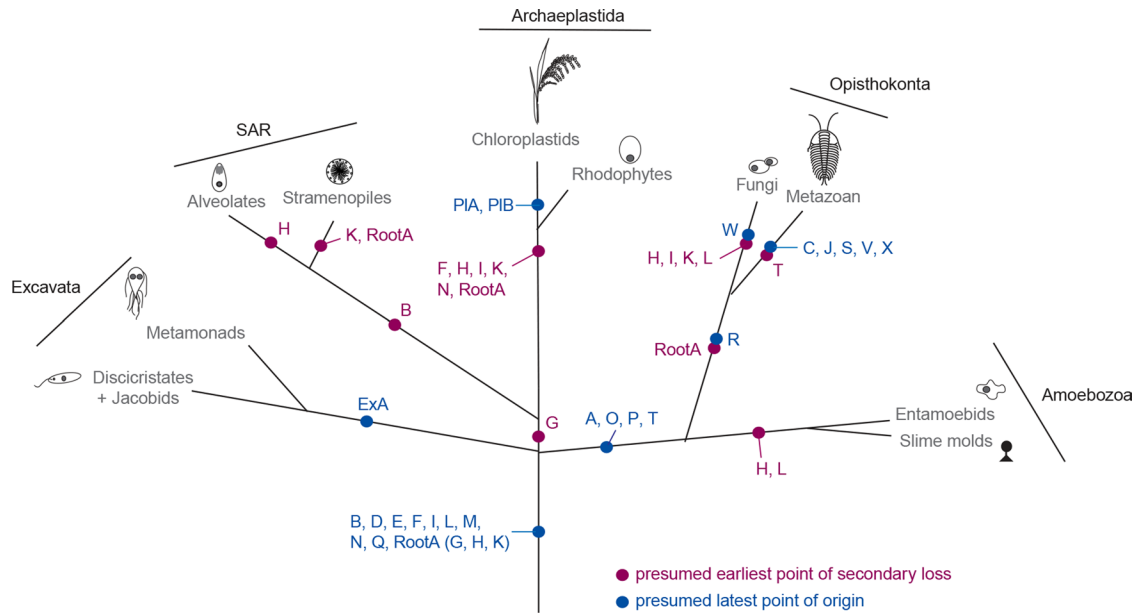
FIGURE 5: Evolutionary history and functions of TBC domain–family proteins. Schematic eukaryotic taxonomy drawn to emphasize the five sampled supergroups. Positions of proposed origins and losses of TBC clades are shown in blue and magenta, respectively. A loss is scored if all taxa above the internode lack the TBC clade, and an origin is recovered based on the assignment of novel families shown in Figure 4. Losses are only scored if two or more taxa demonstrate the loss, and therefore in more terminal branches many potential losses have likely been omitted due to lack of data. The TBC classes in brackets are designated as more speculatively deduced as present in the LECA.

transport, signal transduction, and developmental programs have emerged, and although Rabs are obviously central to TBC activity, the TBC family has been found to have extensive connections with additional GTPases and other proteins (Barr and Lambright, 2010; Fukuda, 2011; Frasa *et al.*, 2012; Popovic *et al.*, 2012). It remains less established how these functions are retained between organisms, in part due to somewhat complex phenotypes and the lack of an established evolutionary framework facilitating comparison and unification of data from multiple organisms. By contrast, evolutionary studies on Rab protein functions have been extremely valuable, allowing such assimilation of information.

The present work extends past analyses, which focused on the opisthokonts (Gao *et al.*, 2008). Our comparative genomics provides a framework for understanding the evolution, conservation, and diversification of the TBC family across the eukaryotes, together with predictions for functions of TBC subfamilies and orthologues. TBC ORFs were generally at lower copy number than Rabs within a given genome, and the correlation between Rab and TBC ORF copy number suggests a coevolutionary component, that is, that a higher number of Rabs is generally associated with a higher TBC number. This implies both maintenance of a level of specificity between TBCs and their partners and a need for comparative independence in regulating the activity of individual Rab proteins. The data also suggest that a small core of TBCs is insufficient to control a large Rab cohort. Clearly, this is consistent with long-term selective pressure against multiple inputs from TBCs into the GTP hydrolysis part of the Rab cycle. In cases in which Rabs are highly expanded, the number of TBCs may lag behind, consistent with earlier models in which Rabs appear to be one of the gene families whose expansion drives pathway innovation (Dacks and Field, 2007). Moreover, because the number of non-TBC Rab-GAPs that are known is a very small cohort, it is unlikely that expansions or contractions in these families have a substantial effect on the size of the TBC cohort.

We found that yeast two-hybrid analysis of the Rab-TBC interactome was of rather limited value and, perhaps most significantly, that there was an unexpectedly low level of concordance between reconstructed Rab-TBC interactomes from humans, yeast, and trypanosomes (Itoh *et al.*, 2006). Further, predictions made from these interactomes did not appear to correlate with RNAi knockdown–based validation; we suggest that a systems approach of this type does not represent a fruitful path to understanding TBC functions.

Phylogenetic reconstruction revealed additional features of TBC evolution, with clear parallels in the Rab family. First, there is an ancient cohort of TBCs predicted to have been present in the LECA, comprising at least 10 clades found in three or more supergroups (Figure 5). Although fewer than the 23 Rab clades reconstructed in the LECA (Elias *et al.*, 2012), this ancient complement of 10 TBCs does indicate considerable complexity and exceeds the repertoire of many unicellular extant organisms, including *S. cerevisiae*. Second, we observed innovation across the eukaryotes, resulting in 25 TBC subclasses recovered from across the entirety of eukaryotes at various taxonomic depths (Figure 5). However, in common with Rab evolution, there is evidence for substantially greater innovation in opisthokonts than other supergroups (Diekmann *et al.*, 2011; Elias *et al.*, 2012). We also identified novel TBC subclasses in a wide range of lineages, with supergroup-specific subfamilies and an ancient subfamily (TBC-RootA) lost from opisthokonts and plants. Secondary losses are quite common, with most supergroups exhibiting evidence for this, with particularly strong losses in Archaeplastida. Specifically in the Archaeplastids this is consistent with the reduced diversity of Rabs but expansion of certain Rab clades (Rutherford and Moore, 2002; Elias *et al.*, 2012) and more generally supports a paradigm originally offered for Rabs, that sculpting, the removal of specific functions, is an important driver in shaping the evolution of the eukaryotic cell (Elias *et al.*, 2012).

Although implemented slightly differently from our earlier analysis of Rab evolution, the present study also relied on ScrollSaw (Elias *et al.*, 2012). Whereas Rab subclasses are well defined and orthologues relatively easily identifiable (Elias *et al.*, 2012), this is not the case for multidomain TBC proteins, which necessitated a preliminary analysis using two model organisms to provide an initial subfamily classification. This enabled us to sample both evolutionary breadth of conservation of these initially established subfamilies and identification of subfamily innovations among other supergroups. This may be a useful and general initial step in future ScrollSaw analyses. Further, the method for identification of the least divergent representative of each subfamily in each supergroup differed. Previously the approach identified pairs of sequences with the lowest mutual distances in ML-corrected distance matrices (Elias *et al.*, 2012), but here we identified robust subfamilies in each supergroup and selected the least divergent sequence within that node or best BLAST *E*-value to the landmark TBC from the initial phylogeny. We could classify ~60% of the sequences in the data set (Supplemental Table S2). When *T. vaginalis*, *T. thermophila*, and *E. histolytica* sequences are excluded, due to having divergent membrane-trafficking systems, successful assignment increased to 71%. Furthermore, in 12 of the 26 genomes examined, >85% of TBCs were classifiable (Supplemental Table S2). Despite this variability, which may in part be due to some individual genomes possessing larger proportions of singletons, we were successful in identifying TBC group relationships across wide evolutionary distances. Ten TBC subfamilies, common to three or more supergroups spread across eukaryotic diversity and thus presumably present in the ancient eukaryotic ancestor, were detected.

We found evidence that overall architecture of many TBCs is conserved, such that orthologues predicted solely from phylogenetic reconstruction of TBC domains have additional conserved domains. Examples include representatives from TBC-G, K, M, and Q clades and partial retention of TBC-I architecture. Although we did not observe complete retention by all members of TBC clades in most instances, this evidence does suggest that, for many TBCs, interactions between Rabs and other cellular components are an ancient feature, providing a functional context retained for 1 billion years. Of interest, the evolution of a second GTPase family, the ARFs, is somewhat distinct to that of Rabs, where evidence suggests a very small ARF cohort in the LECA and most innovation being lineage specific (Li *et al.*, 2004). A recent analysis also suggests that the architectures of ARF-GAP proteins are likely not conserved (Schlacht *et al.*, 2013). Taken together, these studies suggest that there may be coevolution between both the ARF and Rabs and their respective GAP proteins but with a distinct evolutionary trajectory for Rab and ARF systems. Our results are, however, consistent with a recent analysis of Ras-GAPs showing a set of conserved ancestral domains (van Dam *et al.*, 2011).

There is evidence that analysis of orthologous TBC genes in evolutionarily distant lineages can identify common functions. This is particularly clear when orthologues share both a common evolutionary history and domain architecture. Perhaps the best example is TBC-M, with a hallmark C-terminal *trans*-membrane domain. In TBC1D20 (human TBC-M type), the TMD domain is necessary for localization mainly to the ER, where TBC1D20 blocks exit of secretory cargo by inactivating Rab1 (Haas *et al.*, 2007; Sklan *et al.*, 2007). Both the ER distribution and modulation of ER exit are likely important to the pathogenesis of Parkinson's disease (Cooper *et al.*, 2006; Yeger-Lotem *et al.*, 2009). Of significance, overexpression of the *S. cerevisiae* TBC1D20 orthologue, Gyp8p, aggravates the block in ER-to-Golgi trafficking in a yeast model of Parkinson's disease, possibly by enhancing ER accumulation of misfolded α-synuclein (Cooper *et al.*, 2006; Yeger-Lotem *et al.*, 2009). Furthermore, the yeast TBC-B, Gyp7p, and the *H. sapiens* orthologue both appear to function at the vacuole (Eitzen *et al.*, 2000; Brett *et al.*, 2008) and interact with the vacuolar Rab7. Conversely, this framework can highlight potential divergent mechanisms. For example, a human TBC-O member, RN-tre, has well-documented functions in the control of the early endosomal Rab5 (Lanzetti *et al.*, 2004; Haas *et al.*, 2005), and yet this subfamily is unikont specific. Rab5 activity must therefore be modulated by either other TBC families or distinct factors in bikont lineages. With an evolutionary framework it is anticipated that such comparisons can now be made between more divergent organisms, which will accelerate understanding of this highly important family of regulatory proteins.

MATERIALS AND METHODS
Identification and initial subclass assignment of candidate TBC domain–containing ORFs
A panel of 26 predicted proteomes was assembled from the following species: *Arabidopsis thaliana*, *C. reinhardtii*, *Cryptococcus neoformans*, *Cryptosporidium parvum*, *Cyanidioschyzon merolae*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *E. histolytica*, *H. sapiens*, *Leishmania major*, *Monosiga brevicollis*, *N. gruberi*, *Nematostella vectensis*, *Oryza sativa*, *Ostreococcus tauri*, *Plasmodium falciparum*, *P. sojae*, *Physcomitrella patens*, *Rhizopus oryzae*, *S. cerevisiae*, *T. brucei*, *Trypanosoma cruzi*, *Tetrahymena thermophila*, *Thalassiosira pseudonana*, *Theileria parva*, *Toxoplasma gondii*, and *T. vaginalis*. These were selected to represent a broad range of eukaryotes and also to have well-annotated and complete genomes. Accession numbers for sequences used to assemble predicted proteomes are given in Supplemental Table S1. Matches with the TBC domain (Tre-2/Bub2/Cdc16 Rab-binding domain; Richardson and Zon, 1995; Neuwald, 1997) were identified by PSI-BLAST (Altschul *et al.*, 1997) scans of the panel according to the computational procedures we described previously (O'Reilly *et al.*, 2011), using the Pfam (Punta *et al.*, 2012) TBC domain (RabGAP-TBC, PF00566) alignment as query.

Initial subclass assignment for each sequence was obtained by its use as a query in BLASTp searches to the TBC set. *E*-values of <0.05 were considered significant. Assignment as either a tentative member of an opisthokont subclass or not was based on the presence or absence of consistent retrieval of the *H. sapiens* or *S. cerevisiae* TBC subclass rather than on a priori *E*-values, due to the different degrees of conservation seen between the TBC subclasses.

TBC sequences in the pan-eukaryotic data set were initially assigned to a subclass as defined in *H. sapiens* or *S. cerevisiae* if sequences from within a clade consistently retrieved members from the same TBC subclass by BLAST against the *H. sapiens* and *S. cerevisiae* genomes. In reconstructed clades with two or more sequences but where BLAST searches failed to yield consistent retrieval of the same TBC subfamily, the clade was tentatively designated as lineage specific for that supergroup (Supplemental Figures S1–S5 and Supplemental Table S3), pending the phylogenetic analyses described in the next subsection. To minimize false negatives, that is, erroneous exclusion of likely bona fide TBC sequences, 5 of the 591 TBC sequences that were detected in only a single taxon (singleton) in a supergroup but with BLAST *E*-values <*E*-50 were also classified to their respective subclasses (Figure 4). A cut-off value of *E*-50 was determined as the least significant BLAST value that reliably assigned TBCs to a subclass and was also consistent with the results obtained in the phylogenetic analysis.

Phylogenetic analysis

For all phylogenetic analyses the following steps were performed. The TBC domains from relevant sequences were aligned using MUSCLE (Edgar, 2004) and edited by eye. Only regions of unambiguously homologous sequence were included for final analysis. All data sets are available upon request from the authors. The optimal model of sequence evolution for each data set was estimated using Prot-test, version 1.3 (Abascal et al., 2005). The resulting data sets were analyzed using three methods. The optimal tree topology and posterior probability values were obtained using MrBayes, version 3.2 (Ronquist et al., 2012). In all cases two independent runs of four chains were performed for 1×10^6 to 2×10^6 Markov chain Monte Carlo generations, pending convergence of the runs being achieved, as determined by the splits frequency being <0.1. All trees after the graphically determined LnL plateau were included in the consensus. PhyML, version 2.44 (Guindon et al., 2010), and RAxML, version 7.0.0 (Stamatakis, 2006), algorithms were additionally used to produce bootstrap node support from 100 pseudoreplicate data sets.

Supergroup analysis (ScrollSaw)

We previously demonstrated that difficulty with obtaining robust phylogenies can be alleviated by analysis of supergroup-specific data sets, followed by reconstruction of a pan-eukaryote phylogeny using minimal pairwise distance data sets, a procedure we call ScrollSaw (Elias et al., 2012). To obtain a similar analysis here, we first defined the subclades of TBCs via a phylogeny of H. sapiens and S. cerevisiae sequences, with the resulting TBC subclasses assigned either as singletons or clades, with a cut-off for clades reconstructed with statistical support of >0.8 posterior probability from MrBayes and 50% bootstrap support in one of the two maximum-likelihood methods. Clades meeting those criteria were deemed "moderately supported," whereas those meeting the additional criteria of 0.99 posterior probability and 80% bootstrap in one of the two methods were deemed "robustly supported." Phylogenetic analyses were then performed for each supergroup-specific data set to identify and subsequently remove organism-specific duplicates and highly divergent sequences, followed by analyses of the resulting data sets (Supplemental Figures S1–S5). Clades were defined as a well-supported group of at least two sequences from at least two organisms, with statistical support of >0.80 posterior probability and 50% bootstrap support by one ML method (Supplemental Figures S1–S5). The final step of a ScrollSaw analysis is reconstitution of a data set spanning the taxonomic breadth of eukaryotes. Here a representative from each of the clades identified in the supergroup-specific analysis was selected on the basis of having the shortest branch or best BLAST E-value against the H. sapiens or S. cerevisiae genome. Two rounds of phylogeny were necessary. The initial round (data not shown) was needed to identify TBC-C in humans as a singleton; the second is shown in Figure 3. To detect novel/lineage-specific TBCs, the most canonical representative from any novel clade identified in the supergroup phylogenies was also included. Here the representative was the sequence with the shortest branch length, since, by definition, no relevant BLAST value existed against the H. sapiens or S. cerevisiae genome. We assigned TBC sequences from several supergroups as orthologues if recovered in well-supported clades in this final phylogeny (support >0.8 posterior probability/50%/50% bootstrap).

REFERENCES

Abascal F, Zardoya R, Posada D (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Ackers JP, Dhir V, Field MC (2005). A bioinformatic analysis of the RAB genes of Trypanosoma brucei. Mol Biochem Parasitol 141, 89–97.

Adl SM et al. (2012). The revised classification of eukaryotes. J Eukaryot Microbiol 59, 429–493.

Albert S, Will E, Gallwitz D (1999). 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 18, 5216–5225.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

Barr F, Lambright DG (2010). Rab GEFs and GAPs. Curr Opin Cell Biol 22, 461–470.

Brett CL, Plemel RL, Lobingier BT, Vignali M, Fields S, Merz AJ (2008). Efficient termination of vacuolar Rab GTPase signaling requires coordinated action by a GAP and a protein kinase. J Cell Biol 182, 1141–1151.

Brighouse A, Dacks JB, Field MC (2010). Rab protein evolution and the history of the eukaryotic endomembrane system. Cell Mol Life Sci 67, 3449–3465.

Burki F, Shalchian-Tabrizi K, Pawlowski J (2008). Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett 4, 366–369.

Cai H, Reinisch K, Ferro-Novick S (2007). Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. Dev Cell 12, 671–682.

Carlton JM et al. (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science 315, 207–212.

Cooper AA et al. (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. Science 313, 324–328.

Costanzo M et al. (2010). The genetic landscape of a cell. Science 327, 425–431.

Dacks JB, Field MC (2007). Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. J Cell Sci 120, 2977–2985.

De Arras L, Yang IV, Lackford B, Riches DW, Prekeris R, Freedman JH, Schwartz DA, Alper S (2012). Spatiotemporal inhibition of innate immunity signaling by the Tbc1d23 RAB-GAP. J Immunol 188, 2905–2913.

Dhir V, Goulding D, Field MC (2004). TbRAB1 and TbRAB2 mediate trafficking through the early secretory pathway of Trypanosoma brucei. Mol Biochem Parasitol 137, 253–265.

Diekmann Y, Seixas E, Gouw M, Tavares-Cadete F, Seabra MC, Pereira-Leal JB (2011). Thousands of rab GTPases for the cell biologist. PLoS Comput Biol 7, e1002217.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792–1797.

Eitzen G, Will E, Gallwitz D, Haas A, Wickner W (2000). Sequential action of two GTPases to promote vacuole docking and fusion. EMBO J 19, 6713–6720.

El-Bchiri J et al. (2008). Nonsense-mediated mRNA decay impacts MSI-driven carcinogenesis and anti-tumor immunity in colorectal cancers. PLoS One 3, e2583.

Elias M, Brighouse A, Gabernet-Castello C, Field MC, Dacks JB (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J Cell Sci 125, 2500–2508.

Field H, Farjah M, Pal A, Gull K, Field MC (1998). Complexity of trypanosomatid endocytosis pathways revealed by Rab4 and Rab5 isoforms in Trypanosoma brucei. J Biol Chem 273, 32102–32110.

Frasa MA, Koessmeier KT, Ahmadian MR, Braga VMIlluminating the functional and structural repertoire of human TBC/RABGAPs. Nat Rev Mol Cell Biol 13, 67–73.

Fritz-Laylin LK et al. (2010). The genome of Naegleria gruberi illuminates early eukaryotic versatility. Cell 140, 631–642.

Fukuda M (2011). TBC proteins: GAPs for mammalian small GTPase Rab? Biosci Rep 31, 159–168.

Gao X, Jin C, Xue Y, Yao X (2008). Computational analyses of TBC protein family in eukaryotes. Protein Pept Lett 15, 505–509.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307–321.

Haas AK, Fuchs E, Kopajtich R, Barr FA (2005). A GTPase-activating protein controls Rab5 function in endocytic trafficking. Nat Cell Biol 7, 887–893.

Haas AK, Yoshimura S, Stephens DJ, Preisinger C, Fuchs E, Barr FA (2007). Analysis of GTPase-activating proteins: Rab1 and Rab43 are key Rabs required to maintain a functional Golgi complex in human cells. J Cell Sci 120, 2997–3010.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ (2009). Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc Natl Acad Sci USA106, 3859–3864.

Itoh T, Satoh M, Kanno E, Fukuda M (2006). Screening for target Rabs of TBC (Tre-2/Bub2/Cdc16) domain-containing proteins based on their Rab-binding activity. Genes Cells 11, 1023–1037.

Khurana T, Brzostowski JA, Kimmel AR (2005). A Rab21/LIM-only/CH-LIM complex regulates phagocytosis via both activating and inhibitory mechanisms. EMBO J 24, 2254–2264.

Lal K, Field MC, Carlton JM, Warwicker J, Hirt RP (2005). Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*. Mol Biochem Parasitol 143, 226–235.

Lanzetti L, Palamidessi A, Areces L, Scita G, Di Fiore PP (2004). Rab5 is a signalling GTPase involved in actin remodelling by receptor tyrosine kinases. Nature 429, 309–314.

Lee GJ, Sohn EJ, Lee MH, Hwang I (2004). The *Arabidopsis* rab5 homologs rha1 and ara7 localize to the prevacuolar compartment. Plant Cell Physiol 45, 1211–1220.

Li Y, Kelly WG, Logsdon JMJr, Schurko AM, Harfe BD, Hill-Harfe KL, Kahn RA (2004). Functional genomic analysis of the ADP-ribosylation factor family of GTPases: phylogeny among diverse eukaryotes and function in *C. elegans*. FASEB J 18, 1834–1850.

Neuwald AF (1997). A shared domain between a spindle assembly check-point protein and Ypt/Rab-specific GTPase-activators. Trends Biochem Sci 22, 243–244.

O'Reilly AJ, Dacks JB, Field MC (2011). Evolution of the karyopherin-beta family of nucleocytoplasmic transport factors; ancient origins and continued specialization. PLoS One 6, e19308.

Pan X, Eathiraj S, Munson M, Lambright DG (2006). TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. Nature 442, 303–306.

Pinheiro H, Samalova M, Geldner N, Chory J, Martinez A, Moore I (2009). Genetic evidence that the higher plant Rab-D1 and Rab-D2 GTPases exhibit distinct but overlapping interactions in the early secretory pathway. J Cell Sci 122, 3749–3758.

Popovic D, Akutsu M, Novak I, Harper JW, Behrends C, Dikic I (2012). Rab GTPase-activating proteins in autophagy: regulation of endocytic and autophagy pathways by direct binding to human ATG8 modifiers. Mol Cell Biol 32, 1733–1744.

Poteryaev D, Datta S, Ackema K, Zerial M, Spang A (2010). Identification of the switch in early-to-late endosome transition. Cell 141, 497–508.

Punta M *et al.* (2012). The Pfam protein families database. Nucleic Acids Res 40, D290–D301.

Rak A, Fedorov R, Alexandrov K, Albert S, Goody RS, Gallwitz D, Scheidig AJ (2000). Crystal structure of the GAP domain of Gyp1p: first insights into interaction with Ypt/Rab proteins. EMBO J 19, 5105–5113.

Richardson PM, Zon LI (1995). Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogene and the yeast cell cycle regulators BUB2 and cdc16. Oncogene 11, 1139–1148.

Roger AJ, Simpson AG (2009). Evolution: revisiting the root of the eukaryote tree. Curr Biol 19, R165–R167.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61, 539–542.

Rutherford S, Moore I (2002). The *Arabidopsis* Rab GTPase family: another enigma variation. Curr Opin Plant Biol 5, 518–528.

Saito-Nakano Y, Loftus BJ, Hall N, Nozaki T (2005). The diversity of Rab GTPases in *Entamoeba histolytica*. Exp Parasitol 110, 244–252.

Schlacht AD, Mowbrey K, Elias M, Kahn RA, Dacks JB (2013). Ancient complexity, opisthokont plasticity, and discovery of the 11th subfamily of Arf GAP proteins. Traffic, doi: 10.1111/tra.12063.

Sklan EH, Serrano RL, Einav S, Pfeffer SR, Lambright DG, Glenn JS (2007). TBC1D20 is a Rab1 GTPase-activating protein that mediates hepatitis C virus replication. J Biol Chem 282, 36354–36361.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stanier RY (1970). Some aspects of the biology of cells and their possible evolutionary significance. In: Organization and Control in Prokaryotic and Eukaryotic Cells, Twentieth Symposium of the Society for General Microbiology, ed. HP Charles and BCJG Knight, Cambridge, UK: Cambridge University Press, 1–38.

Stechmann A, Cavalier-Smith T (2003). The root of the eukaryote tree pinpointed. Curr Biol 13, R665–R666.

van Dam TJ, Bos JL, Snel B (2011). Evolution of the Ras-like small GTPases and their regulators. Small GTPases 2, 4–16.

Will E, Gallwitz D (2001). Biochemical characterization of Gyp6p, a Ypt/Rab-specific GTPase-activating protein from yeast. J Biol Chem 276, 12135–12139.

Yeger-Lotem E *et al.* (2009). Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nat Genet 41, 316–323.