

Early stages of functional diversification in the Rab GTPase gene family revealed by genomic and localization studies in *Paramecium* species

Lydia J. Bright^{a,b,*}, Jean-Francois Gout^a, and Michael Lynch^a

^aDepartment of Biology, Indiana University, Bloomington, IN 47405; ^bDepartment of Biology, State University of New York at New Paltz, New Paltz, NY 12561

ABSTRACT New gene functions arise within existing gene families as a result of gene duplication and subsequent diversification. To gain insight into the steps that led to the functional diversification of paralogues, we tracked duplicate retention patterns, expression-level divergence, and subcellular markers of functional diversification in the Rab GTPase gene family in three *Paramecium aurelia* species. After whole-genome duplication, Rab GTPase duplicates are more highly retained than other genes in the genome but appear to be diverging more rapidly in expression levels, consistent with early steps in functional diversification. However, by localizing specific Rab proteins in *Paramecium* cells, we found that paralogues from the two most recent whole-genome duplications had virtually identical localization patterns, and that less closely related paralogues showed evidence of both conservation and diversification. The functionally conserved paralogues appear to target to compartments associated with both endocytic and phagocytic recycling functions, confirming evolutionary and functional links between the two pathways in a divergent eukaryotic lineage. Because the functionally diversifying paralogues are still closely related to and derived from a clade of functionally conserved Rab11 genes, we were able to pinpoint three specific amino acid residues that may be driving the change in the localization and thus the function in these proteins.

Monitoring Editor

Patrick J. Brennwald
University of North Carolina

Received: Jun 6, 2016

Revised: Feb 21, 2017

Accepted: Feb 22, 2017

INTRODUCTION

Gene families expand and their members diversify in function over time via the process of retention of individual gene duplicates and subsequent sequence changes (regulatory or protein coding) in sequence, leading to functional changes. This process is subject to both genomic and cell physiological constraints and is a major contributor to the diversification of cellular functions in different lineages (De Bodt *et al.*, 2005; Qian and Zhang, 2014). However, we have little understanding of how evolutionary forces interact with constraints at the subcellular level to determine whether a duplicate

will be retained and, if it is, whether and how its function will change over time.

Gene duplication events occur in genomes at rates approximating the mutation rate per nucleotide site (Lynch and Conery, 2000). However, most gene duplicates are lost through nonfunctionalizing mutations (Lynch and Conery, 2000). Duplicate genes that are retained may experience several functional fates, including conservation of function, subfunctionalization (division of the ancestral functions between the two duplicates), or neofunctionalization (one copy takes on a new function; Force *et al.*, 1999; Conant and Wolfe, 2008; Hahn, 2009). In addition, recent whole-genome studies in *Paramecium* and yeast species support a model in which duplicates are often retained due to constraints on changes in dosage; this retention may produce a milieu in which functional change can occur (Qian and Zhang, 2014; Gout and Lynch, 2015). To better understand these intermediate steps that lead to functional change between duplicates, we examined closely related duplicates produced in *Paramecium aurelia* genomes as the result of whole-genome duplication (WGD), as well as less closely related paralogues within the same subfamily of genes.

This article was published online ahead of print in MBc in Press (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E16-06-0361>) on March 1, 2017.

*Address correspondence to: Lydia J. Bright (lydjbright@gmail.com, bright@newpaltz.edu).

Abbreviations used: LECA, last eukaryotic common ancestor; WGD, whole-genome duplication.

© 2017 Bright *et al.* This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®,” “The American Society for Cell Biology®,” and “Molecular Biology of the Cell®” are registered trademarks of The American Society for Cell Biology.

Rab GTPases comprise the largest gene family of membrane-trafficking determinants, that is, proteins that control membrane-trafficking steps within eukaryotic cells (Behnia and Munro, 2005). Duplication and functional diversification of paralogues within these determinant families is now believed to be the driving force of the increasing number and complexity of membrane-trafficking pathways in different eukaryotic lineages (Schlacht et al., 2014). Rabs in particular, as the most specific determinants in terms of localization patterns (Stenmark, 2009), are believed to be important drivers of this process. This may be due to particular biochemical and evolutionary properties of Rab GTPases.

When activated by GTP binding, Rabs connect upstream vesicle-budding events to downstream motility and fusion with specific target compartments by binding to various effectors, including motor proteins, tethers, and other target membrane-bound proteins (Angers and Merz, 2011; Cherfils and Zeghouf, 2013; Khan and Ménétrey, 2013). Thus the role of Rab proteins as central regulators of trafficking steps depends on their ability to rapidly cycle on and off membranes and to have transient but specific interactions with many different protein partners, which often have overlapping binding sites (Biou and Cherfils, 2004; Khan and Ménétrey, 2013). The site of activity and thus specificity of each Rab protein are determined by interactions with specific key effectors and switch proteins (Aivazian et al., 2006; Cherfils and Zeghouf, 2013; Khan and Ménétrey, 2013). When bound to different effectors and exchange factors, Rab conformations change rapidly throughout their activation cycle in the cell. There is also indirect evidence that Rab conformation may change rapidly over evolutionary time between paralogues, but this is still a poorly understood mechanism (Diekmann et al., 2011; Cherfils and Zeghouf, 2013; Khan and Ménétrey, 2013). This ability to shift rapidly in structure and binding partner(s), and consequently functional location in the cell, appears to be accompanied by changes in a small number of key residues at or surrounding effector binding sites (Merithew et al., 2001; Stein et al., 2012); these qualities may play a role in the evolutionary plasticity of Rabs (Biou and Cherfils, 2004; Bright et al., 2010).

However, for a large subset of Rab GTPases, structure and function are very conserved across eukaryotic lineages, enabling their study and particular use in tracking and reconstructing the ancestral set of pathways and compartments in the last eukaryotic common ancestor (LECA). The LECA is believed to have had 20 Rabs or more, with the attendant complexity in membrane-trafficking pathways (Diekmann et al., 2011; Elias et al., 2012; Klöpffer et al., 2012). In addition, both the expansion and loss of Rab paralogues, in parallel with similar processes in other determinant gene families, appear to have had a major role in sculpting the endomembrane system in different eukaryotic lineages (Elias et al., 2012; Schlacht et al., 2014).

Species in the *Paramecium aurelia* complex of unicellular ciliates present a unique model system for studying the genomic and functional effects of gene duplication. The 15-species *P. aurelia* complex emerged after two WGDs. On average, 50% of duplicates have been retained in each species from the most recent WGD, WGD1 (between 42% for *Paramecium sexaurelia* and 52% for *Paramecium biaurelia*), and on average, 24% of duplicates have been retained from WGD2, the more ancient WGD, with a slow process of ongoing loss likely still occurring (McGrath et al., 2014a). In addition, the *P. aurelia* complex shares an even more ancient WGD, WGD3, with the out-group species *Paramecium caudatum* (McGrath et al., 2014b), from which 8% of duplicates are retained in *P. tetraurelia* (Aury et al., 2006). Thus, unlike other model systems, such as *Saccharomyces cerevisiae*, in which most duplicates have been lost (Wolfe and Shields, 1997; Scannell et al., 2007), a large fraction of *P. aurelia* genes from WGD1

and WGD2 are extant, providing a large data set of closely related paralogues to study (Aury et al., 2006; McGrath et al., 2014a,b). This high number of retained duplicate genes from *P. aurelia*-specific WGDs (and, possibly, the older WGDs) within the species complex represents a host of candidate intermediates on the path to functional change. There is also evidence that at least some WGD1 paralogues in the *P. aurelia* complex have diverged in function (Schilde et al., 2010; Sehring et al., 2010; Osiriska et al., 2011).

In this study, we performed a comprehensive genomic analysis of entire Rab gene families in each of three *Paramecium aurelia* species (*P. biaurelia*, *P. sexaurelia*, and *P. tetraurelia*). To understand genomic constraints on Rabs in the period after duplication, we tracked patterns of Rab gene retention and loss, and divergence in expression levels between WGD1 (recent) Rab duplicates. To track functional fates of *P. aurelia* Rab orthologues and paralogues at the subcellular level, we focused on the expanded Rab11 subfamily, localizing seven key family members in their respective cells. We found evidence of functional diversification in the form of a change in localization from the phagocytic recycling pathway to contractile vacuole tubules. Because the paralogues are still highly similar in amino acid sequence, we were able to identify several amino acid residues that may be driving the localization change.

RESULTS

P. aurelia genomes contain very large Rab gene families

We identified the full complement of Rab homologues in each of three *Paramecium aurelia* species, *P. biaurelia*, *P. sexaurelia*, and *P. tetraurelia*, and one out-group, *P. caudatum*, through a combination of BLAST analysis of de novo-assembled genomes and verification of the presence or absence of WGD1 paralogues by examination of paralogous segments of the assembly (McGrath et al., 2014b). We defined genes as Rabs based on the presence of five Rab-specific domains, or RabFs, defined previously (Pereira-Leal and Seabra, 2001). We also used Rabifier (Diekmann et al., 2011) to confirm our Rab-calling; the designations were largely in agreement, with some manual removal of divergent Rab-like genes and apparently misannotated genes (Supplemental Table S1).

In this way, we uncovered Rab gene families in *Paramecium* genomes that are larger than those in other well-studied eukaryote genomes (Table 1). Part of the gene family expansion in the *P. aurelia* complex is the result of the two WGDs; however, the preduplication out-group species *P. caudatum* contains 90 Rabs, more than most other eukaryotes (Diekmann et al., 2011). In addition, there is evidence that Rab family size is at least partially correlated with genome size; an earlier study found that ~60% of the variance in Rab family size between organisms can be explained by genome size alone (Diekmann et al., 2011). However, this may be biased by the available genomes, which are not sampled evenly across the eukaryotic tree. The previous study and others also found multiple independent eukaryotic phyla with large lineage-specific expansions and sequence diversification of Rabs (Rutherford and Moore, 2002; Lal et al., 2005; Saito-Nakano et al., 2005, 2010; Bright et al., 2010). These expansions in Rab families appear to correspond to an increase in the number of trafficking pathways, which is consistent with evidence that both *Tetrahymena thermophila* and *Paramecium* species have a large number of subcellular trafficking pathways (Plattner, 2010), particularly from a localization study in *T. thermophila*, in which a majority of Rabs were found to have distinct specialized patterns (Bright et al., 2010). Significant expansion of Rab gene families in different lineages is also believed to accompany tissue specificity (in multicellular organisms; Gurkan et al., 2005; Diekmann et al., 2011) or life-stage specificity (Quevillon et al., 2003; Bright

Species	Number of predicted Rabs
<i>Homo sapiens</i>	63
<i>Arabidopsis thaliana</i>	57
<i>Saccharomyces cerevisiae</i>	12
<i>Plasmodium falciparum</i>	11
<i>Dictyostelium discoideum</i>	54
<i>Entamoeba histolytica</i>	91
<i>Tetrahymena thermophila</i>	60
<i>Paramecium caudatum</i>	90
<i>Paramecium tetraurelia</i>	172
<i>Paramecium biaurelia</i>	146
<i>Paramecium sexaurelia</i>	134

Data for other species are from *H. sapiens*, *A. thaliana*, and *S. cerevisiae* (Pereira-Leal and Seabra, 2001); *P. falciparum* (Quevillon et al., 2003); *D. discoideum* (Lal et al., 2005); *E. histolytica* (Saito-Nakano et al., 2005); and *T. thermophila* (Bright et al., 2010).

TABLE 1: Number of Rabs in genomes of well-studied organisms.

et al., 2010; Saito-Nakano et al., 2010). It is not known what overall pattern of innovation and conservation *Paramecium* Rabs follow, but there is evidence that many *Paramecium* paralogues are retained as the result of constraints on gene dosage (Gout et al., 2010; Gout and Lynch, 2015).

P. aurelia Rabs have higher expression levels and paralogue retention than other genes

To make inferences about evolutionary forces acting at the genomic level in *Paramecium* Rab GTPase families, we tracked retention and expression patterns within full Rab gene families in *P. biaurelia*, *P. sexaurelia*, and *P. tetraurelia*. By tracing ohnologons (an ohnologon is a pair of duplicates from a WGD, or the single remaining gene in the case of gene loss) produced from the most recent WGD, WGD1, we found that sets of *P. aurelia* Rab duplicates from WGD1 are retained at between 61 and 70%, which is significantly higher than the genome-wide retention rate in all three *P. aurelia* species (Table 2). However, we also found that Rabs are, on the whole, more highly expressed than other *P. aurelia* genes (Table 2), and previous work showed that highly expressed genes are more likely to be retained in duplicate than genes with low expression (Gout et al., 2010). Therefore, to ascertain whether the higher overall expression levels of Rabs completely explain the higher paralogue retention rate, we randomly selected sets of duplicate genes from the genome with similar expression levels to those of Rab genes and compared their retention rates to those of Rab duplicate pairs. Each set comprised the same number of genes as the number of Rabs. We then repeated this random drawing process 1000

times to achieve a measure of statistical significance. We found that the average retention rate of Rab genes was still significantly greater than that of similarly highly expressed genes (Table 2). Thus the higher average expression level of Rabs does not solely explain their higher retention rates in the *P. aurelia* genomes. Another parameter that strongly influences the probability of post-WGD paralogue retention is dosage-balance constraint, which favors the retention of genes encoding proteins that are part of stable complexes with other proteins (Papp et al., 2003; Aury et al., 2006). Although at least some Rabs are homodimeric, particularly when GDP bound (Pasqualato et al., 2004), this conformation does not bear on dosage-balance constraints with other genes. For the most part, Rabs are not part of heteromeric stable protein complexes, instead interacting transiently with many other protein partners (Biou and Cherfils, 2004); thus the high retention rate of Rabs in *P. aurelia* genomes is even more striking.

Rab duplicates produced as a result of the most recent WGD are more divergent in expression level than other duplicate pairs

A divergence in expression levels between two paralogues can be a signature of functional diversification (Lynch and Conery, 2000; Qian and Zhang, 2008; Gout and Lynch, 2015). Because all regulatory regions are duplicated along with the coding sequences, paralogues from a WGD are born with identical expression patterns. Subsequent changes in regulatory regions, however, can lead to divergence in expression between paralogues. To assay the degree of divergence in expression levels between *P. aurelia* Rab duplicates, we calculated the Pearson's *r* of mRNA expression levels between WGD1 pairs of *P. aurelia* Rabs (McGrath et al., 2014b). Of importance, this is a comparison between duplicates that are all the same age, and thus differences in expression levels cannot be caused by different ages of duplicates. Because expression levels are more conserved in pairs of highly expressed than in lower-expressing genes (Gout et al., 2010), we compared the average *r* of all Rab WGD1-retained ohnologons to that of randomly selected ohnologons with similar average expression levels. We found that, in all three *P. aurelia* species examined, Rab WGD1 ohnologons have a lower correlation of expression level than others with similar average expression levels (Table 3), although this difference was not significant in *P. tetraurelia*, possibly because the lower amount of RNA sequencing data available in this species results in a noisier gene-expression level estimation (*Materials and Methods*). Because microarray analysis across different conditions has been conducted in *P. tetraurelia* (Arnaiz et al., 2010), we were able to calculate correlations of expression level across different life stages and conditions for each pair of WGD1 duplicates in *P. tetraurelia*. We found an average *r* = 0.38 for the *P. tetraurelia* Rab paralogues versus 0.49 for similarly expressed, randomly drawn genes in the genome (*p* < 0.001; based on 1000 random drawings), suggesting that expression patterns of *P. tetraurelia* Rab paralogues across life

Species	WGD1 retention rate, Rabs	WGD1 retention rate, genome-wide ^a	Average expression level, Rabs	Average expression level, genome-wide
<i>P. biaurelia</i>	0.70	0.62	4.3	1.2
<i>P. tetraurelia</i>	0.68	0.53	3.3	1.9
<i>P. sexaurelia</i>	0.61	0.48	4.5	2.2

^aRetention rates were calculated for pools of similarly expressed genes. Paralogue retention rates and expression levels are derived from McGrath et al. (2014b). Expression units are in log-transformed fragments per kilobase per million (FPKM).

TABLE 2: Average retention rates and expression levels of Rabs in three *Paramecium* genomes.

Species	Average <i>r</i> of expression, WGD1 Rabs	Average <i>r</i> of expression, WGD1, genome-wide ^a	<i>p</i>
<i>P. biaurelia</i>	0.49	0.67	0.05
<i>P. tetraurelia</i>	0.50	0.58	0.27
<i>P. sexaurelia</i>	0.29	0.55	0.03

^aCorrelation coefficients represent the genome-wide value of similarly expressed genes. *p* values are for the difference between Rab and genome-wide correlation coefficients.

TABLE 3: Expression-level correlation between Rab duplicates in three *Paramecium* genomes.

stages are, on average, more divergent than that of other gene families.

Finally, to ascertain whether these differences might be caused by a higher proportion of genomic rearrangements surrounding or near Rab genes, we tested whether there were, overall, more rearrangements around Rabs. We found that if we look specifically at Rabs, 95% of annotated Rabs are in syntenic blocks (uninterrupted by rearrangements), whereas 90% of genes in the genome on the whole are in such syntenic blocks. This observation indicates that most, if not all, gene expression changes between Rab duplicates are not caused by genomic rearrangements and supports a model in which differences in expression level come from small-scale mutations (base substitutions, small insertions, and deletions) in the regulatory regions of Rab genes as opposed to large-scale mutations.

Specific Rab subfamilies are expanded in the *Paramecium* lineage

Phylogenetically, the Rab GTPase family can be divided into six ancient, deep-branching clades, which originated long before the last common eukaryotic ancestor (LECA; Elias *et al.*, 2012). These six ancient clades contain conserved members representing the breadth

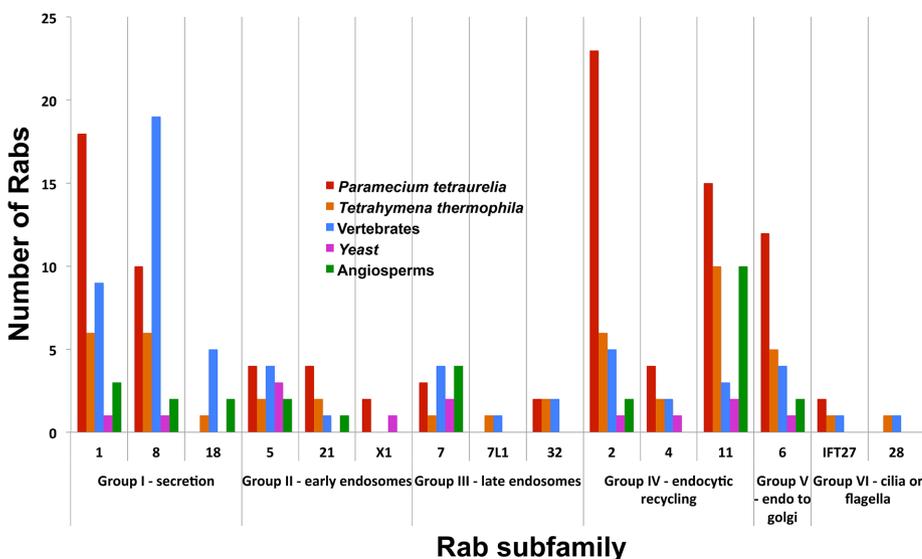


FIGURE 1: *Paramecium* genomes contain lineage-specific expansions of Rab subfamilies. Each numbered subfamily represents one Rab present in the LECA and retained in the ciliate lineage in either *T. thermophila* or *P. aurelia* complex. Subfamily sizes from several well-studied eukaryotic lineages are also included. Numbers of Rabs in vertebrate, yeast, and angiosperm lineages (not specific genomes) are derived from Klöpper *et al.* (2012); numbers of Rabs in *T. thermophila* genome are from Bright *et al.* (2010).

of endomembrane trafficking pathways, which can be divided roughly into those involved in endocytic traffic (inward in the cell) or exocytic traffic (outward toward the plasma membrane). To analyze more precisely the expansions and contractions in the *Paramecium* Rab subfamilies, we divided the most conserved *P. tetraurelia* Rabs (approximately two-thirds of the *P. tetraurelia* Rab family members) into subfamilies, using phylogenetic methods (see *Materials and Methods*). We then compared these numbers to subfamily sizes in well-studied lineages (for which functional data on the roles of Rabs within the subfamilies are available; Diekmann *et al.*, 2011; Elias *et al.*, 2012; Klöpper *et al.*, 2012). Note that “loss” from a subfamily may simply designate high sequence divergence and thus would place a given gene in the less-conserved one-third of Rabs. We found that *Paramecium* genomes contain recognizable descendants of 12 of ~23 proposed LECA Rabs, representing all six major clades (Figure 1 and Supplemental Table S1). We found five clearly and highly expanded Rab subfamilies in *Paramecium* genomes, including Rab1, 2, 6, 8, and 11 (Figure 1). Both *Tetrahymena* and *Paramecium* share all of these expansions, although they are magnified in *P. aurelia* species (particularly the Rab1 and 2 expansions), in large part due to the two successive WGDs. Many Rab clades that were present in the last eukaryotic ancestor appear to have been lost or become highly divergent in sequence in the SAR lineage (including Rab14, 24, 32B, and 34), the alveolate lineage (Rab20 and 50 and RabTitan), and the ciliate lineage (including Rab22 and 23) or in one ciliate lineage and not another (including Rab18, 28, and 7L1, lost in *Paramecium* and not *Tetrahymena*, and RabX1, lost in *Tetrahymena* and not *Paramecium*; Figure 1). These losses have occurred across five of the six larger ancestral clades of Rabs.

Localization analysis reveals functional diversification between paralogues in the expanded Rab11 subfamily in *Paramecium*

The Rab11 subfamily is expanded in both *Paramecium* and *Tetrahymena* lineages; in addition, this subfamily is also highly expanded in the plant *Arabidopsis thaliana*, encompassing a large portion of its total Rab count (Rutherford and Moore, 2002). Previous work in both *Tetrahymena* (Bright *et al.*, 2010) and *Arabidopsis* (Choi *et al.*, 2013) showed that functional diversification accompanied this expansion.

The ancestrally conserved function of the Rab11 protein is in trafficking to and from the recycling endosome. This compartment’s central role in membrane trafficking appears to predispose the subfamily to functional diversification; Rab11 paralogues direct traffic between the recycling endosome and the plasma membrane, between sorting endosomes and the recycling endosome, and between the recycling endosome and the *trans*-Golgi network (Kelly *et al.*, 2012). As a result of their roles in trafficking to this centralized and highly interconnected compartment, in addition to key roles in receptor and membrane recycling, Rab11 subfamily members are involved in many eukaryotic cell processes, including delivery of material to the cleavage furrow during cytokinesis (Horgan and McCaffrey, 2012), phagocytosis (Cox *et al.*, 2000), long-term potentiation (Wang *et al.*, 2008; Kelly *et al.*, 2011), cell

migration (Yoon *et al.*, 2005; Caswell *et al.*, 2008), and pathogen resistance (Guichard *et al.*, 2014).

Ciliates, including *Tetrahymena* and *Paramecium*—as befits voracious swimming bacterivores—have elaborate, yet highly efficient phagocytic pathways requiring the recycling of massive amounts of membrane (Allen *et al.*, 1993). As *Paramecium* cells swim, specialized cilia sweep bacteria into a specialized oral apparatus, from which large food vacuoles rapidly and continually form in the presence of food particles. Unlike mammalian (Steinman *et al.*, 1983) or amoeboid (Bowers, 1980) phagocytic cells, in which food vacuoles are formed from the plasma membrane, *Paramecium* food vacuoles are composed of recycled intracellular membrane (Allen and Wolf, 1974). Membrane is recycled back to the base of the oral apparatus from several points in the phagocytic pathway. Early phagosomes quickly become acidified phagolysosomes through the fusion of smaller, endocytic-derived, low-pH acidosomes with the larger vacuole; these acidosomes soon pinch off as discoidal or flattened vesicles to return back to the oral apparatus (Allen *et al.*, 1993). Later-stage phagosomes encounter secondary lysosomes, which fuse and then pinch back off to rejoin the lysosomal pool of vesicles (Allen and Fok, 1984). Membrane recycling also occurs from the cytoproct (Allen and Wolf, 1974, 1979). Each of these membrane-recycling steps must involve specialized molecular components. In addition, *T. thermophila* Rab11 paralogues have diverse and non-overlapping localization patterns, targeting to the oral apparatus (TtRab11B), endocytic recycling compartments (TtRab11A, D), or plasma membrane (TtRab11E, G; Bright *et al.*, 2010). Analysis of *T. thermophila* Rab11 localization patterns did not elucidate a clear ancestral link between these compartments. We hypothesized that localization analysis of Rab11 proteins in the *P. aurelia* complex might shed more light on the evolution of phagocytic membrane recycling in these cells and others, as well as conservation and diversification of Rab11 paralogues.

We localized seven Rab11 proteins representing three conserved Rab11 clades as N-terminally tagged GFP-Rab proteins (Figure 2 and Supplemental Figure S1). Of a possible 14 paralogues in these three clades, we injected 11 and were able to detect GFP signal from 7 Rab11-GFP fusion proteins in cells. Subcellular localization patterns were determined in live cells using confocal microscopy and verified both by the highly characteristic morphology of structures, such as the oral apparatus and contractile vacuoles in *Paramecium* cells (Allen, 2000; Sehring *et al.*, 2010), and colabeling with a subcellular marker dye, LysoTracker (Figure S5 and Supplemental Video S2). The identification of structures by characteristic morphology was conducted using simultaneously captured differential interference contrast images (see Supplemental Video S1 for a representative movie). In *Paramecium* cells, acidic structures include acidified phagolysosomes, phagoacidosomes (structures that have been shown to form a link between the endocytic and phagosomal pathways), and lysosomes (Allen *et al.*, 1993). We analyzed the specific compartments to which the Rab11-GFP fusions targeted and whether they shared similar localization patterns to paralogues within and between Rab11 clades (Figure 2, B and C). All constructs were expressed and analyzed in cells from the native species, unless otherwise indicated.

Within the I clade, which, along with one orthologue from each of the two preduplication species, contains two *P. biaurelia* WGD1 paralogues, two *P. tetraurelia* WGD1 paralogues, and just one *P. sexaurelia* orthologue, we were able to localize the two *P. biaurelia* paralogues, one of the *P. tetraurelia* paralogues, and the *P. sexaurelia* orthologue (Figure 2A). Of interest, all four of these proteins shared a nearly identical localization pattern. These proteins all label the

base of the oral apparatus (with a very strong signal), filaments (likely to be vesicles associated with actin filaments) leading to the oral apparatus and to acidified structures that include vesicles and larger, more tubulated structures that may represent phagoacidosomes, or another structure that overlaps with the endolysosomal pathway (Figure 2B). Time-lapse imaging of cells exhibiting this pattern (Supplemental Video S2) shows that this is likely to be a membrane-recycling pathway of spent food vacuolar membrane to the base of the oral apparatus (Allen *et al.*, 1993). In summary, the entirety of clade I appears to share this pattern, and we predict that both the one unlocalized in-group orthoparologue (t11i1) and the preduplication out-group orthologues share this pattern, as it appears to be ancestral to the group. This assessment was strengthened when one of the clade A proteins was localized, as it shares this very specific localization pattern (Figure 2B). However, when two members of the C clade were labeled, their localization pattern was divergent from the others. Although they share some targeting to the base of the oral apparatus with the other two clades examined, members of the C clade appear to have much more signal in the cytoplasm (*P. tetraurelia* 11c1-GFP) and to label the contractile vacuole complex (CVC) and the nuclear envelope (*P. biaurelia* 11c2-GFP; Figure 2 and Supplemental Figure S1). It remains to be determined whether other members of the C clade, including the preduplication orthologues, share these pattern(s) or harbor intermediate localization patterns with the other clades. Of importance, in this study, we observed no instances of diversification between duplicates produced as the result of WGD1 or 2. We uncovered evidence of diversification only of older paralogues.

Because the I and C clades are more closely related to one another than to the out-group A clade (Figure 2), we made the evolutionary inference that the targeting of the A and I clade Rab11 paralogues to filaments and putative phagoacidosomes is the ancestral pattern, whereas the localization to CVCs and the nucleus seen in the C clade is derived. In addition, the localization pattern at the base of the oral apparatus is shared by all Rab11 paralogues examined thus far, indicating that some ancestral function is retained.

Functional diversification between Rab11 paralogues traces to changes in three key amino acids

Earlier sequence-based analysis identified specific domains in Rab proteins believed to determine the specificity of different subfamilies (Pereira-Leal and Seabra, 2000). Subsequent combined sequence and structural analysis further pinpointed specific residues that correlate with conformational changes of key binding pockets in Rab proteins (Merithew *et al.*, 2001). Because the binding partners that interact with these pockets are integral both for linking Rab activity to downstream motor and fusion functions and directing Rab localization and activity to specific membranes, these specific amino acids may themselves be determinants of specificity and, therefore, localization. Of interest, the discovery of one or a small number of amino acids that cause changes in function is not without precedent; earlier work identified, via a very carefully targeted small interfering RNA knockdown study, different functional effects from targeting two different splice variants, or isoforms, of human Rab6A that differ by only two amino acids (Del Nery *et al.*, 2006).

Even though the A, C, and I clades are related by ancient gene duplication events that appear to predate even WGD3, the protein sequences are still closely related enough, particularly in key functional domains, that we were able to pinpoint specific amino acid changes that may be driving the functional diversification

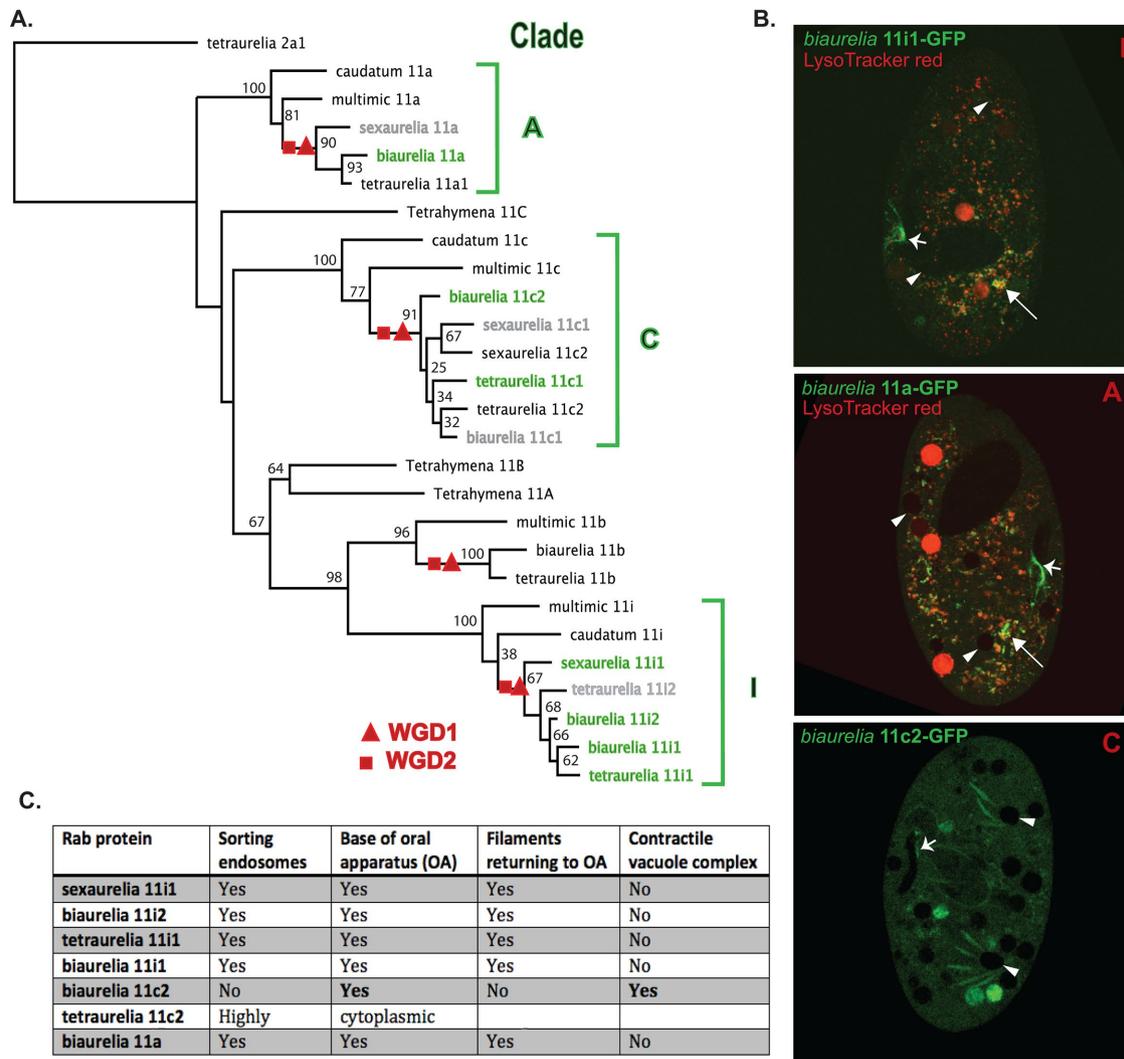


FIGURE 2: Functional diversification has occurred in one clade of the *Paramecium* Rab11 subfamily. (A) Maximum-likelihood phylogeny of four closely related Rab11 clades; see *Materials and Methods* for details of tree construction. Rabs localized in this study are in green letters. Rab genes for which injection and localization attempts were made are in gray. Gene names denote *Paramecium* species (except for *Tetrahymena* genes) and Rab ID. Red triangles and red circles mark approximate times of WGD1 and 2, respectively. (B) Representative localization patterns for each clade. Green signal is that of the protein-GFP fusion as expressed in the native species. Red signal is LysoTracker Red; areas of overlap are yellow. In each cell, arrows mark the base of the oral apparatus, arrowheads mark contractile vacuole complexes, and flat-headed arrows mark putative phagoacidosomes (except in the clade C cell). Each representative cell represents >10 identical localization patterns recorded for each fusion. The *P. biaurelia* Rab11c cell image does not include the LysoTracker Red channel to better show the cytoplasmic vs. organellar GFP signal (see Supplemental Figure S1 for individual channels and merged images for each GFP fusion made). (C) Chart exemplifying the presence and absence of key components of the localization patterns.

(Figure 3). Only three amino acid changes in conserved domains in the C clade of Rab11 paralogues; these residues are located adjacent to or within key effector binding sites, and conformational changes in these sites are correlated with dramatic changes in effector binding partners (Merithew *et al.*, 2001; Biou and Cherif, 2004).

DISCUSSION

In this study, we examined how early stages in the evolution of Rab duplicates have unfolded in a group of closely related *P. aurelia* species. It appears that, after a whole-genome duplication in the *P. aurelia* lineage, Rab duplicates were retained at a higher rate than other genes in the genome and were able to diverge in expression

level. This combination of events might have set the stage for later functional diversification to occur, which is indeed what we observe in older pre-WGD1 and -2 paralogues. This retention and later diversification may be particularly true in *P. aurelia* genomes, which lose duplicates at a relatively slow rate (McGrath *et al.*, 2014b). It may also be true for Rab proteins specifically, as Rabs appear to be particularly nimble in space and time, cycling rapidly through activation and inactivation steps, on and off membranes, and with many different effector and exchange factor proteins. These combined characteristics of Rabs and *P. aurelia* genomes may have led or be leading to a high amount of functional diversification in *P. aurelia* Rabs and contributing to diverse roles in membrane trafficking. For this reason, the Rab GTPase family in *P. aurelia* is a particularly good system

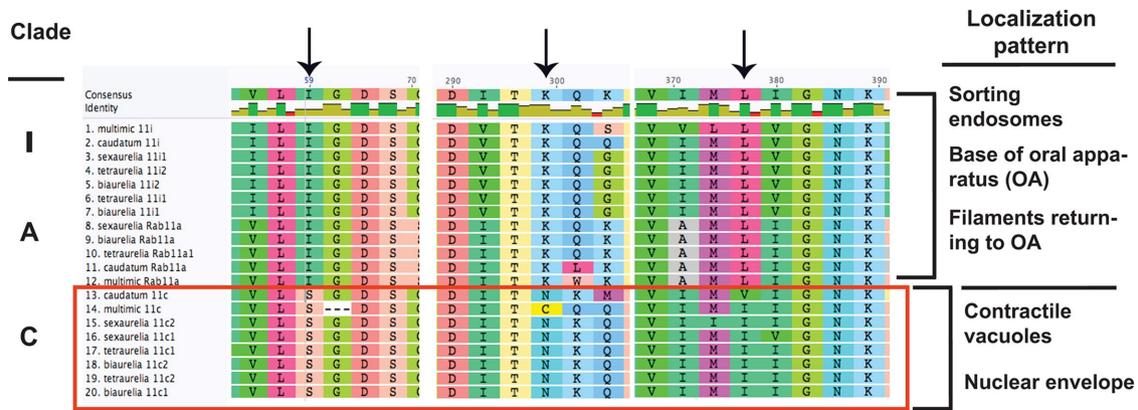


FIGURE 3: Portions of the amino acid alignment of Rab11 genes from the A, C, and I clades in the Figure 2 phylogeny. Only parts of the alignment with changes from A/I clades to C clade are shown. Red box surrounds the C clade of functionally diversifying Rab11s. Arrows denote the three amino acid changes correlated with functional diversification. The representative localization patterns of the two sets of genes (those in either the I/A or C clade) are listed on the right.

in which to observe duplicate diversification at both the genomic and subcellular levels.

Five of the seven proteins localized in this study had nearly identical and broad localization patterns, encompassing filaments returning to the base of the oral apparatus, putative phagoacidosomes, and sites on the plasma membrane. Strikingly, this pattern is very different from that observed in the *T. thermophila* Rab11 subfamily, in which each Rab11 paralogue has a specific, unique localization pattern (Bright *et al.*, 2010). However, whereas most or all of the closely related *P. aurelia* paralogues are believed to have arisen from WGDs (Aury *et al.*, 2006), the *T. thermophila* Rab paralogues derive from more ancient tandem duplications (Eisen *et al.*, 2006). Tandem duplicates and WGD-derived paralogues may have different evolutionary trajectories after duplication due to either constraints imposed by linkage (Lynch and Conery, 2000) or dosage, with WGD-derived paralogues more likely to be retained initially for dosage-balance reasons (Papp *et al.*, 2003) and tandem duplicates more likely to acquire new, beneficial functions (Maere *et al.*, 2005). However, even in the small set of proteins that we localized in this study, we see functional diversification, in the form of localization changes, that have occurred and may still be occurring, nested within conserved-localization clades, supporting a model of functional change occurring after retention in a functionally conserved state for long periods of time.

In future work, it will be important to determine whether other members of the C clade, particularly the preduplication orthologues, share the diversified (and possibly less specific) pattern that *P. biaurelia* 11c2 and *P. tetraurelia* 11c1 display. This may illuminate intermediate steps and shed light on the timing of this functional change.

The broadly shared and presumably ancestral localization pattern of a pair of Rab11 clades uncovered by this study (a subset of all Rab11 clades in *Paramecium* genomes, it must be noted) strengthens the case for the conservation of extensive interactions between the phagocytic and endocytic pathways (Yutin *et al.*, 2009; Boulais *et al.*, 2010). Phagocytosis, considered to be a specialization of endocytic pathways, was probably used by the ancestral eukaryote in a nutritive capacity, just as it is employed by free-swimming heterotrophic protists today. However, phagocytic functions have been lost in many eukaryotic lineages and have evolved into immune-related functions in macrophages. Although

it is known that these parallel versions of phagocytosis intersect the endolysosomal pathway at several points (Schmid *et al.*, 2014), it is not known how many or in what way the molecular components are conserved or subject to innovation and whether that innovation involves recruitment of new types of genes or duplication and diversification within existing gene families. However, it is hypothesized that the latter is a major part of this process (Wideman *et al.*, 2014).

With regard to the switching of function from a canonical or conserved endocytic/phagocytic recycling pattern to a contractile vacuole recycling pattern in the C clade, there is evidence that in both trypanosomes (which are excavates, a different eukaryotic supergroup than alveolates, of which ciliates are part; Jeffries *et al.*, 2001; Ulrich *et al.*, 2011) and *Dictyostelium* (which are in the Amoebozoa phylum; Harris *et al.*, 2001), Rab11 proteins are involved in both endocytic recycling and contractile vacuole functions. Because all three of these eukaryotic lineages are highly divergent from one another, these organelles may be ancestrally related, possibly through ancient lysosomal-related organelles (Docampo, 2015). This ancient relatedness may mean that the two pathways share similar components, facilitating functional switching of a Rab11 paralogue from one compartment to the other. The foregoing example shows clearly that as we amass functional data on gene-family members from disparate branches of the eukaryotic tree and observe the repeated cooption of duplicates from one molecular pathway to another, we gain key information about the evolutionary history of these compartments.

On a shorter evolutionary time scale, this work provides a clear case in which the study of multiple sets of retained duplicates across multiple species can elucidate parallel evolution and, in the process, provide insight into a small number of amino acid changes that may be driving functional (and, potentially, structural) change. In addition, it provides support for a model in which whole-genome duplicates are retained over long periods of time, possibly for dosage reasons, setting the stage for functional diversification to occur.

MATERIALS AND METHODS

Rab identification and phylogenetic comparisons

Using known *T. thermophila* Rabs, we identified a set of putative Rabs by taking the top five hits from BLASTP searches of all

predicted proteins in each *Paramecium* genome (e-value cutoff of 1×10^{-5} , but all results had e-values $< 1 \times 10^{-40}$; McGrath et al., 2014a,b). From this large initial set, we analyzed all translated putative Rab sequences for the presence of five Rab-specific motifs: IGVDV, KLQIW, RFRSIT, YYRGA, and LVDIT (Pereira-Leal and Seabra, 2000). Genes with a majority of motifs that resembled other Ras-like GTPases and those without C-terminal prenylation motifs (Rab-like proteins) were eliminated from the Rab set. The full *Paramecium* Rab tree (Supplemental Figure S1) was made by aligning the full (untrimmed) Rab amino acid sequences using the Muscle program. Some minimal manual adjustment was made to the alignment to align known conserved Rab domains; Supplemental Figure S2 gives the full amino acid alignment. The neighbor-joining (NJ) tree was made using the Jukes–Cantor substitution model and 100 bootstrapped replicates to construct the consensus tree shown in Supplemental Figure S1. The out-group used was *P. tetraurelia* Ran GTPase 1a. The larger phylogeny of the entire Rab gene family was then used to make preliminary groupings of *Paramecium* Rabs into subfamilies (Supplemental Table S1). Subfamilies and orthology (with specific *T. thermophila* Rabs) were confirmed by reciprocal tblastn of coding sequences with e-value cutoffs of -10 used for the reciprocal BLASTs; however, the phylogenetic criteria were always more stringent (i.e., a *Tetrahymena* or human landmark sequence might come up as a first-hit homologue using BLAST but would not robustly group with the *Paramecium* set of orthoparalogues in the smaller NJ phylogeny). Subfamily alignments and trees are available upon request. Rabs were numbered by subfamily (Rab 1, 2, etc.) according to their closest human and *T. thermophila* orthologues; paralogues within a subfamily were then named a, b, c, and so on, until all paralogues in a subfamily were lettered. WGD1 paralogues were then numbered 1 and 2 if both were retained.

For the Rab11 subfamily tree shown in Figure 2, both NJ and maximum-likelihood (ML) phylogenies were constructed from nucleotide alignments of the coding sequences, trimmed to remove the faster-evolving N- and C-terminal regions (see Supplemental Figures S3 and S4 for the trimmed nucleotide and translated amino acid alignments used, respectively). The ML tree was constructed using the PhyML program (Guindon et al., 2010) and rooted with the PtRab2a1 sequence; the TN93 substitution matrix was used, and the Subtree Pruning and Regrafting topological moves algorithm (Hordijk and Gascuel, 2005) was used to search the tree space and determine the best topology. The topologies of the trees made by both NJ and ML methods were similar, but the *Tetrahymena* TtRab11A and C orthologues grouped with different clades in each tree, with low bootstrap support in each case. Phylogenies were constructed using the Geneious software package (www.geneious.com).

Duplicate retention, expression level, and expression correlation analysis

Whole-genome sequencing, annotation, alignment of paralogous chromosome segments and RNA sequence analysis of *P. biaurelia*, *P. sexaurelia*, *P. tetraurelia*, and *P. caudatum* species were performed and published previously (McGrath et al., 2014a,b). Expression levels were calculated from whole-genome sequencing of RNA from vegetative cells: RNA sequencing reads were mapped to the genome using Bowtie/TopHat (Langmead et al., 2009; Kim et al., 2013), and transcripts were predicted using Cufflinks (Trapnell et al., 2012) from a logarithm transformation of the values of fragments per kilobase of exon per million fragments mapped (FPKM) reported by Cufflinks; before the log transformation, we added 0.1 to

the FPKM value of all genes to avoid log transformation of a null value (McGrath et al., 2014b).

To measure retention rates for similarly expressed genes, we randomly drew sets of WGD1 ohnologs with the same expression level as the set of Rab WGD1 ohnologs (86, 83, and 102 ohnologs for *P. biaurelia*, *P. sexaurelia*, and *P. tetraurelia*, respectively) from the genome and calculated their overall retention rate. The random drawing was repeated 1000 times for each species; *p* values are calculated as the fraction of times out of the 1000 drawings when the randomly drawn ohnologs have equal or higher retention rates than Rab ohnologs.

To calculate correlation of expression levels between duplicates, we analyzed two sets of expression data. First, because RNA sequencing data from vegetative cells were available for vegetative growth for all four species studied (McGrath et al., 2014a,b), we were able to calculate the Pearson's *r* of expression level between retained WGD1 Rab paralogues in *P. biaurelia*, *P. sexaurelia*, and *P. tetraurelia*. Because highly expressed paralogues tend to have less divergence in expression level (Gout et al., 2010; Gout and Lynch, 2015), we compared the *r* to that of WGD1 paralogues with similar average expression level. We made 1000 random drawings of the same number of non-Rab WGD1 pairs (60, 51, and 70 WGD1 paralogue pairs from the *P. biaurelia*, *P. sexaurelia*, and *P. tetraurelia* genomes, respectively) and computed *r* across the sets of randomly drawn non-Rab pairs after each drawing. We calculated *p* values as the number of random drawings with *r* equal to or less than that of Rab WGD1 pairs. Second, we used microarray data available in *P. tetraurelia* across several conditions and life stages (Arnaiz et al., 2010) to calculate *r* across those stages for Rab and non-Rab WGD1 pairs in that species. Again, we made 1000 random drawings of 70 non-Rab WGD1 pairs and computed *r* values across sampled microarray conditions for each pair. Numbers reported in the text represent averages of *r* values for each pair, and the *p* value was calculated again as the number of random drawings with *r* equal to or less than that of the *P. tetraurelia* Rab WGD1 pairs.

Rab-GFP expression in live cells

We attempted cloning and localization of the majority of the *P. aurelia* Rab11 genes in these three clades (11 of 14); however, of 11 cloned and injected, we were able to detect GFP signal for seven genes. Included in the set to which we did not successfully express a GFP fusion are *P. tetraurelia* 11i2, *P. biaurelia* 11c1, *P. sexaurelia* 11c1, and *P. sexaurelia* 11a, which would potentially have provided valuable information about how genes within the 11c and 11a clades are or are not diversifying from one another.

Rab-GFP gene fusions were cloned using the Gateway system (ThermoFisher). PCR-amplified Rab genes were TOPO cloned into the pENTR-D-TOPO entry vector. CACC was added to each forward primer to enable directional cloning into pENTR-D. The expression vector for constitutive expression under the calmodulin promoter was obtained from Jean Cohen as pPXV (Hauser et al., 2000) and converted into a Gateway destination vector using the Gateway conversion kit (ThermoFisher). The genes were recombined into this final destination expression vector using the Clonase reaction (ThermoFisher) and screened by diagnostic digest and Sanger sequencing.

Paramecium cells from *P. biaurelia*, *P. sexaurelia*, *P. tetraurelia*, and *P. caudatum* were grown in wheat grass medium infused with *Klebsiella* bacteria (Beisson et al., 2010).

Linearized Rab-GFP expression constructs containing *Paramecium* telomeres on each end were injected into the macronucleus of

the relevant *Paramecium* species by microinjection using a Nikon Diaphot inverted microscope and a Narishige IM300 microinjector. Single cells were immobilized under silicon oil using an Eppendorf CellTram Vario manual microinjector, injected, remobilized, and recovered into single-cell culture. Needles for injection and immobilization were pulled using a Sutter Model P-87 micropipette puller. Each resulting line was screened for either rescue, by injection of an Nd7 gene-containing construct, of a trichocyst release phenotype in the *P. tetraurelia* in an Nd7- mutant (Skouri and Cohen, 1997) and GFP expression or solely for GFP expression (all other *Paramecium* species) on a Zeiss Axioplan wide-field epifluorescence microscope. Several GFP-positive lines for each construct were screened to check for consistency of expression pattern at different expression levels. Cellular organelles, including contractile vacuoles and the oral apparatus, were identified by correlating the bright-field channels (representative movie through the Z-stack) of each Z-series of sections taken through the cell with the green fluorescent channel. Localization patterns were virtually all consistent, although different levels of expression or brightness of GFP signal were observed. GFP-expressing cell lines were grown without allowing starvation because triggering autogamy and the ensuing remaking of the macronucleus would discard the exogenous expression vector. For confocal imaging, cells were dyed with a 5 μ M final concentration of LysoTracker Red acidic dye (LysoTracker Red DND-99; ThermoFisher) for 5 min and immobilized by centrifuging a 500- μ l culture into 30 μ l of a 2% methyl cellulose solution. The supernatant was then removed, and the live cells embedded in the methyl cellulose were transferred onto a Fisher SuperFrost slide under a 22 \times 22 mm no. 1 coverslip, and immediately imaged on either a Leica SP5 or SP8 laser scanning confocal microscope. Postprocessing of images was performed using the public domain ImageJ program (imagej.nih.gov/ij/). Supplemental Video S2 is a time lapse taken on the Leica SP8 laser scanning confocal microscope. Because expression levels of injected constructs in *Paramecium* cells can be affected by the amount of injected DNA, at least three independent lines of injected cells were analyzed to ascertain whether the localization patterns were reproducible and specific. In addition, fixation by paraformaldehyde and glutaraldehyde, respectively, was tried, and because the localization pattern in fixed cells was a small subset of that in live cells and thus did not accurately reflect the live-cell pattern, we continued with live-cell imaging alone for the remaining imaging.

ACKNOWLEDGMENTS

The plasmid pPXV was kindly provided by Jean Cohen (Centre National de la Recherche Scientifique [CNRS], Gif-sur-Yvette, France). For training in *Paramecium* microinjection, we thank Judith Van Houten, Junji Yano (University of Vermont), Jean Cohen, Anne Aubusson-Fleury, and Janine Beisson (CNRS, Gif-sur-Yvette, France), with travel assistance from the Groupement De Recherche Européen *Paramecium* DB. Microinjection was conducted in the *Drosophila* Indiana Genomics Initiative facility. Microscopy was conducted in the Light Microscopy Imaging Facility at Indiana University with assistance from Jim Powers. We thank James Pease for help with large-scale Rab alignments and Parul Johri and Tom Doak for invaluable experimental discussions and aid in editing. This research was supported by a National Institutes of Health Ruth L. Kirschstein National Research Service Award, National Institutes of Health/National Institute of General Medical Sciences Grant 1F32GM101891-02 to L.J.B., and National Science Foundation Award MCB-1050161 to M.L.

REFERENCES

- Aivazian D, Serrano RL, Pfeffer S (2006). TIP47 is a key effector for Rab9 localization. *J Cell Biol* 173, 917–926.
- Allen RD (2000). The contractile vacuole and its membrane dynamics. *Bioessays* 22, 1035–1042.
- Allen RD, Fok AK (1984). Retrieval of lysosomal membrane and acid phosphatase from phagolysosomes of *Paramecium caudatum*. *J Cell Biol* 99, 1955–1959.
- Allen RD, Ma L, Fok AK (1993). Acidosomes: recipients of multiple sources of membrane and cargo during development and maturation. *J Cell Sci* 106, 411–422.
- Allen RD, Wolf RW (1974). The cytoproct of *Paramecium caudatum*: structure and function, microtubules, and fate of food vacuole membranes. *J Cell Sci* 14, 611–631.
- Allen RD, Wolf RW (1979). Membrane recycling at the cytoproct of *Tetrahymena*. *J Cell Sci* 35, 217–227.
- Angers CG, Merz AJ (2011). New links between vesicle coats and Rab-mediated vesicle targeting. *Semin Cell Dev Biol* 22, 18–26.
- Arnaiz O, Gout J-F, Bétermier M, Bouhouche K, Cohen J, Duret L, Kapusta A, Meyer E, Sperling L (2010). Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* 11, 547.
- Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178.
- Behnia R, Munro S (2005). Organelle identity and the signposts for membrane traffic. *Nature* 438, 597–604.
- Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR, et al. (2010). Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb Protoc* 2010, pdb.prot5362.
- Biou V, Cherfils J (2004). Structural principles for the multispecificity of small GTP-binding proteins. *Biochemistry* 43, 6833–6840.
- Boulais J, Trost M, Landry CR, Dieckmann R, Levy ED, Soldati T, Michnick SW, Thibault P, Desjardins M (2010). Molecular characterization of the evolution of phagosomes. *Mol Syst Biol* 6, 423.
- Bowers B (1980). A morphological study of plasma and phagosome membranes during endocytosis in *Acanthamoeba*. *J Cell Biol* 84, 246–260.
- Bright LJ, Kambesis N, Nelson SB, Jeong B, Turkewitz AP (2010). Comprehensive analysis reveals dynamic and evolutionary plasticity of Rab GTPases and membrane traffic in *Tetrahymena thermophila*. *PLoS Genet* 6, e1001155.
- Caswell PT, Chan M, Lindsay AJ, McCaffrey MW, Boettiger D, Norman JC (2008). Rab-coupling protein coordinates recycling of alpha5beta1 integrin and EGFR1 to promote cell migration in 3D microenvironments. *J Cell Biol* 183, 143–155.
- Cherfils J, Zeghouf M (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* 93, 269–309.
- Choi S-W, Tamaki T, Ebine K, Uemura T, Ueda T, Nakano A (2013). RABA members act in distinct steps of subcellular trafficking of the FLAGELLIN SENSING2 receptor. *Plant Cell* 25, 1174–1187.
- Conant GC, Wolfe KH (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* 9, 938–950.
- Cox D, Lee DJ, Dale BM, Calafat J, Greenberg S (2000). A Rab11-containing rapidly recycling compartment in macrophages that promotes phagocytosis. *Proc Natl Acad Sci USA* 97, 680–685.
- De Bodt S, Maere S, Van de Peer Y (2005). Genome duplication and the origin of angiosperms. *Trends Ecol Evol* 20, 591–597.
- Del Nery E, Miserey-Lenkei S, Falguières T, Nizak C, Johannes L, Perez F, Goud B (2006). Rab6A and Rab6A' GTPases play non-overlapping roles in membrane trafficking. *Traffic* 7, 394–407.
- 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.
- Docampo R (2015). The origin and evolution of the acidocalcisome and its interactions with other organelles. *Mol Biochem Parasitol* 209, 3–9.
- Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, et al. (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 4, e286.
- Elias M, Brighthouse 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.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

- Gout J-F, Kahn D, Duret L, Consortium PP-G (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet* 6, e1000944.
- Gout J-F, Lynch M (2015). Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol Biol Evol* 32, 2141–2148.
- Guichard A, Nizet V, Bier E (2014). Rab11-mediated trafficking in host-pathogen interactions. *Nat Rev Microbiol* 12, 624–634.
- 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.
- Gurkan C, Lapp H, Alory C, Su AI, Hogenesch JB, Balch WE (2005). Large-scale profiling of Rab GTPase trafficking networks: the membrane. *Mol Biol Cell* 16, 3847–3864.
- Hahn MW (2009). Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* 100, 605–617.
- Harris E, Yoshida K, Cardelli J, Bush J (2001). Rab11-like GTPase associates with and regulates the structure and function of the contractile vacuole system in *Dictyostelium*. *J Cell Sci* 114, 3035–3045.
- Hauser K, Haynes W, Kung C, Plattner H, Kissmehl R (2000). Expression of the green fluorescent protein in *Paramecium tetraurelia*. *Eur J Cell Biol* 79, 144–149.
- Hordijk W, Gascuel O (2005). Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics* 21, 4338–4347.
- Horgan CP, McCaffrey MW (2012). Endosomal trafficking in animal cytokinesis. *Front Biosci (Schol Ed)* 4, 547–555.
- Jeffries TR, Morgan GW, Field MC (2001). A developmentally regulated rab11 homologue in *Trypanosoma brucei* is involved in recycling processes. *J Cell Sci* 114, 2617–2626.
- Kelly EE, Horgan CP, McCaffrey MW (2012). Rab11 proteins in health and disease. *Biochem Soc Trans* 40, 1360–1367.
- Kelly EE, Horgan CP, McCaffrey MW, Young P (2011). The role of endosomal-recycling in long-term potentiation. *Cell Mol Life Sci* 68, 185–194.
- Khan AR, Ménétrey J (2013). Structural biology of Arf and Rab GTPases' effector recruitment and specificity. *Structure* 21, 1284–1297.
- Kim D, Perteza G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36.
- Klöpffer TH, Kienle N, Fasshauer D, Munro S (2012). Untangling the evolution of Rab G proteins: implications of a comprehensive genomic analysis. *BMC Biol* 10, 71.
- 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.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Lynch M, Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155.
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005). Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* 102, 5454–5459.
- McGrath CL, Gout J-F, Doak TG, Yanagi A, Lynch M (2014a). Insights into three whole-genome duplications gleaned from the *Paramecium caudatum* genome sequence. *Genetics* 197, 1417–1428.
- McGrath CL, Gout JF, Johri P, Doak TG, Lynch M (2014b). Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res* 24, 1665–1675.
- Merithew E, Hatherly S, Dumas JJ, Lawe DC, Heller-Harrison R, Lambright DG (2001). Structural plasticity of an invariant hydrophobic triad in the switch regions of Rab GTPases is a determinant of effector recognition. *J Biol Chem* 276, 13982–13988.
- Osińska M, Wiejak J, Wypych E, Bilski H, Bartosiewicz R, Wyroba E (2011). Distinct expression, localization and function of two Rab7 proteins encoded by paralogous genes in a free-living model eukaryote. *Acta Biochim Pol* 58, 597–607.
- Papp B, Pál C, Hurst LD (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197.
- Pasqualato S, Senic-Matuglia F, Renault L, Goud B, Salamero J, Cherfils J (2004). The structural GDP/GTP cycle of Rab11 reveals a novel interface involved in the dynamics of recycling endosomes. *J Biol Chem* 279, 11480–11488.
- Pereira-Leal JB, Seabra MC (2000). The mammalian Rab family of small GTPases: definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the Ras superfamily. *J Mol Biol* 301, 1077–1087.
- Pereira-Leal JB, Seabra MC (2001). Evolution of the Rab family of small GTP-binding proteins. *J Mol Biol* 313, 889–901.
- Plattner H (2010). Membrane trafficking in protozoa SNARE proteins, H+-ATPase, actin, and other key players in ciliates. *Int Rev Cell Mol Biol* 280, 79–184.
- Qian W, Zhang J (2008). Gene dosage and gene duplicability. *Genetics* 179, 2319–2324.
- Qian W, Zhang J (2014). Genomic evidence for adaptation by gene duplication. *Genome Res* 24, 1356–1362.
- Quevillon E, Spielmann T, Brahimi K, Chattopadhyay D, Yeramian E, Langley G (2003). The *Plasmodium falciparum* family of Rab GTPases. *Gene* 306, 13–25.
- 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.
- Saito-Nakano Y, Nakahara T, Nakano K, Nozaki T, Numata O (2010). Marked amplification and diversification of products of ras genes from rat brain, Rab GTPases, in the ciliates *Tetrahymena thermophila* and *Paramecium tetraurelia*. *J Eukaryot Microbiol* 57, 389–399.
- Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc Natl Acad Sci USA* 104, 8397–8402.
- Schilde C, Schönemann B, Sehring IM, Plattner H (2010). Distinct subcellular localization of a group of synaptobrevin-like SNAREs in *Paramecium tetraurelia* and effects of silencing SNARE-specific chaperone NSF. *Eukaryotic Cell* 9, 288–305.
- Schlacht A, Herman EK, Klute MJ, Field MC, Dacks JB (2014). Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking system. *Cold Spring Harb Perspect Biol* 6, a016048.
- Schmid SL, Sorkin A, Zerial M (2014). Endocytosis: past, present, and future. *Cold Spring Harb Perspect Biol* 6, a022509.
- Sehring IM, Reiner C, Plattner H (2010). The actin subfamily PtAct4, out of many subfamilies, is differentially localized for specific local functions in *Paramecium tetraurelia* cells. *Eur J Cell Biol* 89, 509–524.
- Skouri F, Cohen J (1997). Genetic approach to regulated exocytosis using functional complementation in *Paramecium*: identification of the ND7 gene required for membrane fusion. *Mol Biol Cell* 8, 1063–1071.
- Stein M, Pilli M, Bernauer S, Habermann BH, Zerial M, Wade RC (2012). The interaction properties of the human Rab GTPase family—comparative analysis reveals determinants of molecular binding selectivity. *PLoS One* 7, e34870.
- Steinman RM, Mellman IS, Muller WA, Cohn ZA (1983). Endocytosis and the recycling of plasma membrane. *J Cell Biol* 96, 1–27.
- Stenmark H (2009). Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol* 10, 513–525.
- Trapnell C, Roberts A, Goff L, Perteza G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7, 562–578.
- Ulrich PN, Jimenez V, Park M, Martins VP, Atwood J, Moles K, Collins D, Rohloff P, Tarleton R, Moreno SNJ, et al. (2011). Identification of contractile vacuole proteins in *Trypanosoma cruzi*. *PLoS One* 6, e18013.
- Wang Z, Edwards JG, Riley N, Provance DW Jr, Karcher R, Li XD, Davison IG, Ikebe M, Mercer JA, Kauer JA, et al. (2008). Myosin Vb mobilizes recycling endosomes and AMPA receptors for postsynaptic plasticity. *Cell* 135, 535–548.
- Wideman JG, Leung KF, Field MC, Dacks JB (2014). The cell biology of the endocytic system from an evolutionary perspective. *Cold Spring Harb Perspect Biol* 6, a016998.
- Wolfe KH, Shields DC (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708–713.
- Yoon SO, Shin S, Mercurio AM (2005). Hypoxia stimulates carcinoma invasion by stabilizing microtubules and promoting the Rab11 trafficking of the $\alpha\beta4$ integrin. *Cancer Res* 65, 2761–2769.
- Yutin N, Wolf MY, Wolf YI, Koonin EV (2009). The origins of phagocytosis and eukaryogenesis. *Biol Direct* 4, 9.