Phylogenetic reconstruction and evolution of the Rab GTPase gene family in Amoebozoa

Alfredo L. Porfírio-Sousa, Alexander K. Tice, Matthew W. Brown & Daniel J. G. Lahr

View supplementary material ⎘

Published online: 29 Mar 2021.

Submit your article to this journal ⎘

View related articles ⎘

View Crossmark data ⎘

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Check for updates

Phylogenetic reconstruction and evolution of the Rab GTPase gene family in Amoebozoa

Alfredo L. Porfírio-Sousa[a], Alexander K. Tice[b,c], Matthew W. Brown[b,c,d], and Daniel J. G. Lahr [a]

aDepartment of Zoology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil; bDepartment of Biological Sciences, Mississippi State University, Starkville, Mississippi, USA; cInstitute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, Mississippi, USA; dInstitute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

ABSTRACT

Rab GTPase is a paralog-rich gene family that controls the maintenance of the eukaryotic cell compartmentalization system. Diverse eukaryotes have varying numbers of Rab paralogs. Currently, little is known about the evolutionary pattern of Rab GTPase in most major eukaryotic 'supergroups'. Here, we present a comprehensive phylogenetic reconstruction of the Rab GTPase gene family in the eukaryotic 'supergroup' Amoebozoa, a diverse lineage represented by unicellular and multicellular organisms. We demonstrate that Amoebozoa conserved 20 of the 23 ancestral Rab GTPases predicted to be present in the last eukaryotic common ancestor and massively expanded several 'novel' in-paralogs. Due to these 'novel' in-paralogs, the Rab family composition dramatically varies between the members of Amoebozoa; as a consequence, 'supergroup'-based studies may significantly change our current understanding of the evolution and diversity of this gene family. The high diversity of the Rab GTPase gene family in Amoebozoa makes this 'supergroup' a key lineage to study and advance our knowledge of the evolution of Rab in Eukaryotes.

Introduction

Cell compartmentalization is a crucial characteristic of Eukaryotes, and the Rab GTPase gene family is a central controller of these compartments [1,2]. Rab GTPases comprise a paralog-rich family that regulates all stages of membrane trafficking [1,3,4]. Rab proteins control from vesicle budding, cargo sorting and transportation, to vesicle tethering and fusion [2,3,5]. Through these processes, Rabs perform several cellular roles, such as maintaining the communication between the cell compartments and membrane, endocytic and exocytic pathways, and intraflagellar transport [1,2].

The Rab GTPase family composition varies among diverse eukaryotic lineages. While the human genome has over 60 Rab paralogs [1,2,6], *Saccharomyces cerevisiae* has 11 [1], and *Arabidopsis thaliana* has 57 [7–10]. Along with conserved orthologs present in these three lineages, several Rabs are lineage-specific. This composition specificity is related to multiple Rab GTPase family radiations, and such radiations give rise to 'novel' in-paralogs that may perform lineage-specific roles [9].

The complement of Rab GTPases has been studied for several eukaryotic lineages [11]. Fungi have between eight and 12 Rabs [12]; the ciliate *Tetrahymena thermophila* has around 70 [13,14]; *Plasmodium falciparum* (Apicomplexa) has 11 [15,16]; *Toxoplasma gondii* (Apicomplexa) has 15 [17]; *Trypanosoma brucei* (Kinetoplastida) has 16 [18,19]; *Naegleria gruberi* (Heterolobosea) has around 30 [20]; *Trichomonas vaginalis* (Metamonada) incredibly has around 300 [21,22]. A comparative study has shown that the last eukaryotic common ancestor (LECA) had up to 23 Rab paralogs [23]. Contrastively, the evolution of Rabs in a whole 'supergroup' has been less studied. Recently, it has been shown that the evolution of Rabs in most Archaeplastida was characterized by the conservation of the majority ancestral eukaryotic Rabs, except for the rhodophytes (red algae), and rare gene duplication [24].

Amoebozoa are a very diverse eukaryotic 'supergroup' and have diverse cell forms, life cycles, and ecologies [25]. Currently, the Rab GTPase family has been annotated in members of two of the three major lineages of amoebozoans: Evosea, represented by *Dictyostelium discoideum* with around 56 annotated Rabs, *Mastigamoeba balamuthi* with around 25 Rabs, and *Entamoeba histolytica* with over 90 Rabs; Discosea, represented by *Acanthamoeba castellanii*, with 93 Rabs

[23,26]; no Tubulinea has been sampled for Rabs. Thus, the investigation of the repertoire of the Rab GTPase gene family in Amoebozoa has been restricted to few species. Recently, several deeply sequenced transcriptomes of amoebozoans have been generated [25,27], enabling a broader study of the Rab GTPases in this eukaryotic 'supergroup'.

Here we present a comprehensive phylogenetic study of the Rab GTPase family in Amoebozoa. We considered genomes and deeply sequenced transcriptomes of 44 Amoebozoa lineages and a comprehensive eukaryotic Rab GTPase dataset previously available [23]; we also included representatives of breviates and apusomonads, two lineages that with Opisthokonta represent Obazoa, the sister group of Amoebozoa. We focused on a broad perspective of amoebozoan diversity, aiming to identify the general pattern of evolution of robust Rab GTPase subfamilies in a eukaryotic 'supergroup', rather than a comprehensive identification and annotation of all Rabs in all Amoebozoa. Our phylogenetic reconstruction put in an evolutionary perspective the Rabs previously annotated in the genomes of some amoebozoans and the new paralogs identified in the available transcriptomic data, comparing with the paralogs present in diverse eukaryotic lineages. We demonstrate that the three major lineages of Amoebozoa conserve most of the ancestral paralogs present in the Last Eukaryotic Common Ancestor (LECA) and have undergone a massive expansion of the Rab GTPase gene family through the origin of 'novel' in-paralogs (i.e., new paralogs of a given Rab subfamily originated through gene duplication of ancestral paralogs). By sampling several flagellated amoebozoans, we identified one ancestral paralog that has not been previously found in Amoebozoa. Our study demonstrates that no single amoebozoan lineage represents the diversity of Rab GTPase in Amoebozoa and corroborates that sampling diverse eukaryotic lineages in a 'supergroup' perspective may significantly improve our knowledge of the Rab GTPase gene family diversity and evolution.

Results and discussion

We considered a dataset of 44 Amoebozoa lineages, *Pygsuia biforma* (breviates), and *Thecamonas trahens* (apusomonads) composed of genomes and deeply sequenced transcriptomes (**Supplementary Table 1**). This dataset represents the three major lineages of Amoebozoa and their subclades, with seven representatives of Tubulinea (Corycidia + Echinamoebidia + Elardia subclades), 28 representatives of Evosea (Cutosea + Archamoebae + Eumycetozoa + Variosea subclades), and nine representatives of Discosea (Flabellimia +

Centramoebia subclades) [**Supplementary Table 1;** see 25, for Amoebozoa phylogeny]. The more significant number of Evosea representatives is due to the availability of several genomes for this major group. Also, we considered several evosean flagellated species in our analysis since a Rab paralog involved with the flagellum (IFT27/ RabL4) had not been previously identified in Amoebozoa. Additionally, we considered *P. biforma* and *T. trahens*, that compose amoebozoan sister-group Obazoa, and have not been sampled for Rabs.

We identified Rab sequences from the genomes and transcriptomes of amoebozoans, *P. biforma*, and *T. trahens* through similarity search (BLAST). For that, we compiled a comprehensive dataset of Rab sequences to serve as our query dataset. We initially considered as potential Rabs the sequences of amoebozoans, *P. biforma*, and *T. trahens* significantly similar to the sequences of the query Rab dataset (considering a BlastP E-value ≥ *1e-4*). The BLAST similarity searches did not enable us to easily assign several Rabs of Amoebozoa, *P. biforma*, and *T. trahens* to one of the ancestral Rab subfamilies predicted to be present in the LECA or identify sequences that represent other families of the Ras superfamily; this was already expected given the high diversity and divergence of some Rab paralogs and the sequence similarity between Rab and other members of the Ras superfamily [28–30]. Thus, we further analysed the sequences identified by BLAST through phylogenetic reconstructions (not shown) and excluded non-Rab sequences (*i.e.*, sequences representing other Ras subfamilies) to create a curated amoebozoan, breviate, and apusomonad Rab dataset.

We performed multiple phylogenetic reconstructions to assign the amoebozoan, breviate, and apusomonad Rab sequences to the Rab GTPase subfamilies predicted to be present in LECA (**Supplementary Figure 1 A – B**). First, we generated a master phylogenetic reconstruction considering the curated Rab dataset of the 44 Amoebozoa species, *P. biforma, T. trahens* (**Supplementary Table 2**), and the dataset curated by 23, (**Supplementary Figure 1A** and **Supplementary Figure 2;** all the sequences considered in the present study are available in FASTA format as **Supplementary Material 1**). Although the master phylogeny has several regions with low resolution, especially at deep branching levels, it enabled us to recover and identify highly supported clades (*i.e.*, ultrafast bootstrap branch support ≥95% as suggested by IQ-TREE documentation; **Supplementary Figure 1C**) of most Rab subfamilies present in Amoebozoa (**Supplementary Figure 2**). Interestingly, six of the seven Rab subfamilies recovered in clades of lower support (ultrafast bootstrap branch

support between 80% and 94%) are those that expanded in Amoebozoa and have several 'novel' in-paralogs, as shown below. Regions of low resolution have been consistently identified as a characteristic of the phylogenetic reconstruction of the Rab GTPase family given the evolutionary complexity of this gene family [23,24]. To further analyse specific subfamilies and the 'novel' in-paralogs that compose the amoebozoan Rab repertoire, we generated multiple phylogenetic reconstructions considering subsets of our master reconstruction (**Supplementary Figure 1B and Supplementary Figure 3–9**). We applied the Rabifier automated annotation to cross-validate the assignments of Rab sequences to Rab subfamilies made based on the phylogenetic reconstructions (**Supplementary Table 2**), enabling us to unambiguously identify the Rab subfamilies that were conserved in the last amoebozoan common ancestor and the extant amoebozoans, as well as the 'novel' Rab in-paralogs that appeared during the evolution of Amoebozoa.

Amoebozoa conserves most of the Rab GTPases subfamilies present in LECA

Amoebozoa conserves 20 Rab subfamilies of the 23 predicted to be present in LECA (Figure 1). We identified these 20 subfamilies in all major groups of Amoebozoa, except for IFT27 (RabL4), RTW (RabL2), and Rab23, that we were able to find exclusively in the major group Evosea, and Rab34 that was identified only in few amoebozoans sampled (Figures 2 and 3). By sampling several flagellated amoebozoans, we identified for the first time the paralog IFT27 (RabL4) in Amoebozoa, a paralog known to be involved with intraflagellar transport in diverse Eukaryotes [31]. These findings demonstrate that most of the Rab GTPase paralogs present in LECA have been conserved in the last amoebozoan common ancestor (LACA) and are present in extant Amoebozoa lineages (Figure 3).

We consistently found IFT27 (RabL4), RTW (RabL2), and Rab23 (Rab paralogs functionally associated with flagellar apparatus [31–39]) in flagellated Evosea of diverse subclades (Figure 2; see 25and 40 for the classification of flagellated evoseans), for instance, *Mastigamoeba balamuthi* and *Rhizomastix elongata*, (Archamoebae), *Echinostelium minutum*, *Ceratiomyxa fruticulosa* (Eumycetozoa), and *Idionectes vortex* (Cutosea). These findings indicate that the intraflagellar transport control in Amoebozoa may be a conserved process, involving these ancestral Rabs. We also found these paralogs in evoseans closely related to flagellated lineages, such as *Planoprotostelium fungivorum* (Variosea) and *Echinosteliopsis oligospora* (Eumycetozoa) [41] (Figure

2). The presence/absence pattern of these paralogs in other eukaryotic flagellated or non-flagellated lineages is complex [24]; for instance, the paralogs IFT27 and RTW are absent in some lineages that have flagella, while Rab23 is present in lineages that have not been observed to have flagella [24]. Previously, the absence of these paralogs has been associated with eukaryotic groups that lost the flagellar apparatus or have only a transient flagellum [24]. Here we show that even amoebozoan species that have transient flagella (i.e., *Idionectes vortex*, *Echinostelium minutum*, and *Protosporangium articulatum*; [25, 40]) or in which flagellum have not been observed (i.e., *Echinosteliopsis oligospora*; 41, 42]) maintained IFT27 (RabL4), RTW (RabL2), Rab23, or even all these three paralogs (Figure 2). We did not identify the paralogs IFT27 (RabL4), RTW (RabL2), and Rab23 present as a cohort in several flagellated amoebozoans considered (*e.g.*, *Idionectes vortex*, *Ceratiomyxa fruticulosa*, and *Rhizomastix elongata*); since these species have only transcriptomic data available, we cannot assess whether these paralogs are actually missing in the genomes of these flagellated evoseans or in the genomes of the other amoebozoans.

The 20 paralogs predicted to be present in LACA have different conservation and potential loss patterns throughout eukaryotic groups. Several Amorphea (Amoebozoa + Obazoa) have conserved most of these paralogs, except for *Pygsuia biforma* (breviate), *Thecamonas trahens* (apusomonads), and Fungi, where there are marked potential losses (Figure 3). Several of these paralogs have not been found in members of the other 'supergroups', such as Excavata, Archaeplastida, and SAR [23,24]. While Rab 1, 2, 5, 6, 7, 8, 11, and 18 have been conserved in most eukaryotes examined for Rabs, Rab 4, 14, 20, 21, 23, 24, 32, 34, 50, and Titan have been constantly lost in diverse lineages of Fungi, SAR, Excavata, and Archaeplastida [24]. Thus, while Fungi, SAR, Excavata, and some Archaeplastida can be characterized by a pattern of a massive reduction of these Rabs [23,24], Amorphea (except for Fungi, *P. biforma*, and *T. trahens*) have a pattern of conservation of most of the Rab paralogs present in the LECA, including consistent conservation in the three major lineages of Amoebozoa as shown here (Figure 3).

Potentially 3 Rab paralogs (Rabs 20, 22, and 28) of the 23 predicted to be present in the LECA were absent in the LACA. These paralogs are absent in the genomes or transcriptomes of the 44 amoebozoans considered in this study. Although most amoebozoans transcriptomes and genomes are not complete (**Supplementary Table 1**), thus not being informative about the absence of a given Rab paralog, currently, we have no evidence
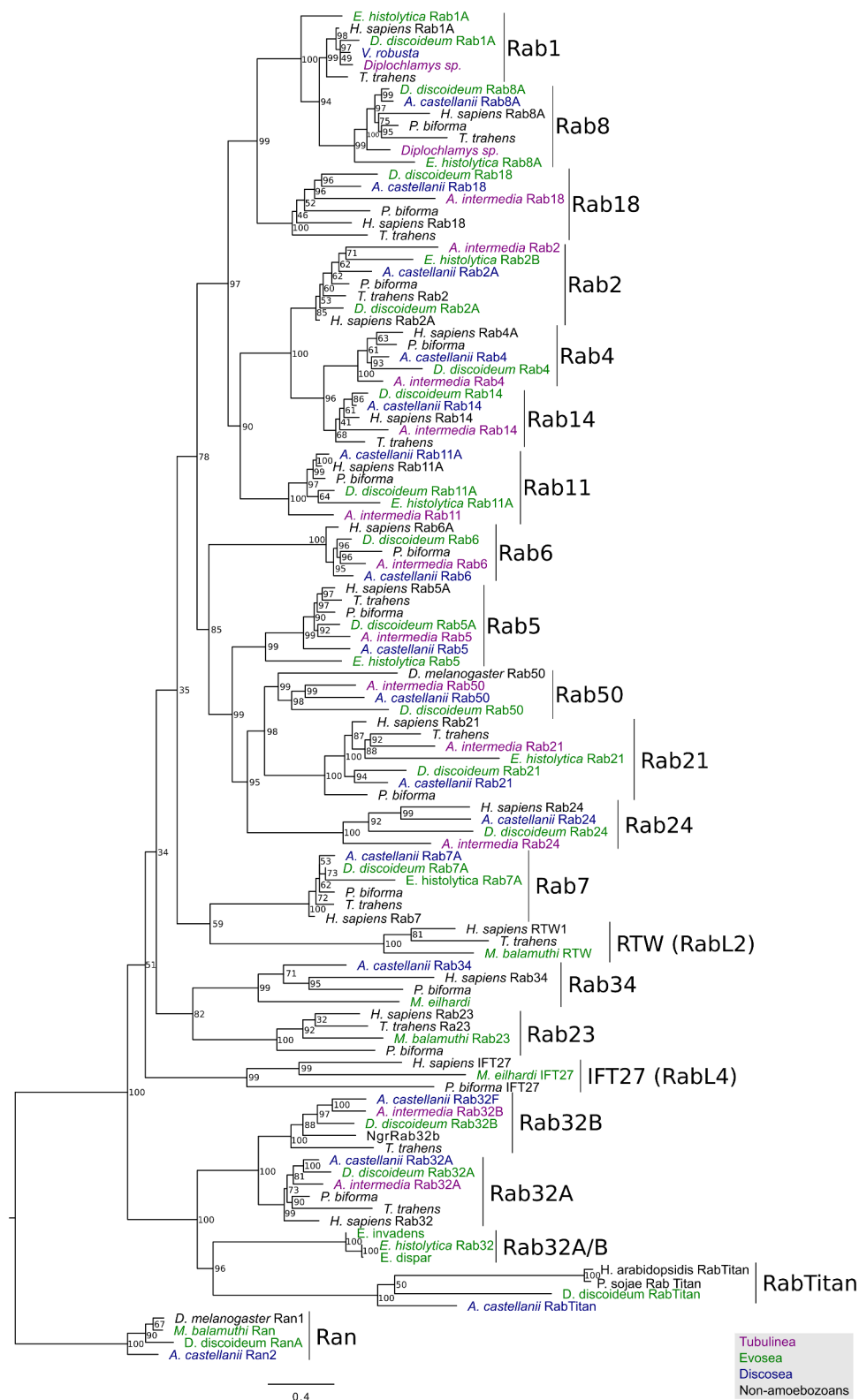
Figure 1. Phylogenetic reconstruction of the 20 ancestral Rab subfamilies predicted to be present in the Last Amoebozoa Common Ancestor (LACA). Maximum likelihood (ML) tree of a subset of the master dataset represented in **Supplementary Figure 2**, focusing on the 20 Rab subfamilies predicted to have been conserved in LACA, as well as in few selected amoebozoan and non-amoebozoan taxa. We considered Ran as the outgroup. Representatives of the three major groups of Amoebozoa are highlighted in purple (Tubulinea), green (Evosea), and blue (Discosea). Vertical bars indicate the Rab subfamilies clades. Values at nodes are ML bootstrap (BS) (1,000 ultrafast BS rep, IQ-TREE LG+I+ G4). Note that the subfamilies Rab1, 2, 4, and 32 were recovered in lower supported (ultrafast BS<95%) or paraphyletic clades. This observation is consistent to what previous studies have found; Rab 1 and Rab2 have

been consistently recovered as paraphyletic or lower supported clades due to Rab8 and Rabs4/14 respectively [23,24]. The Rab32 subfamily is recovered as a paraphyletic clade due to the branching pattern of *Entamoeba*'s sequences classified as Rab32.
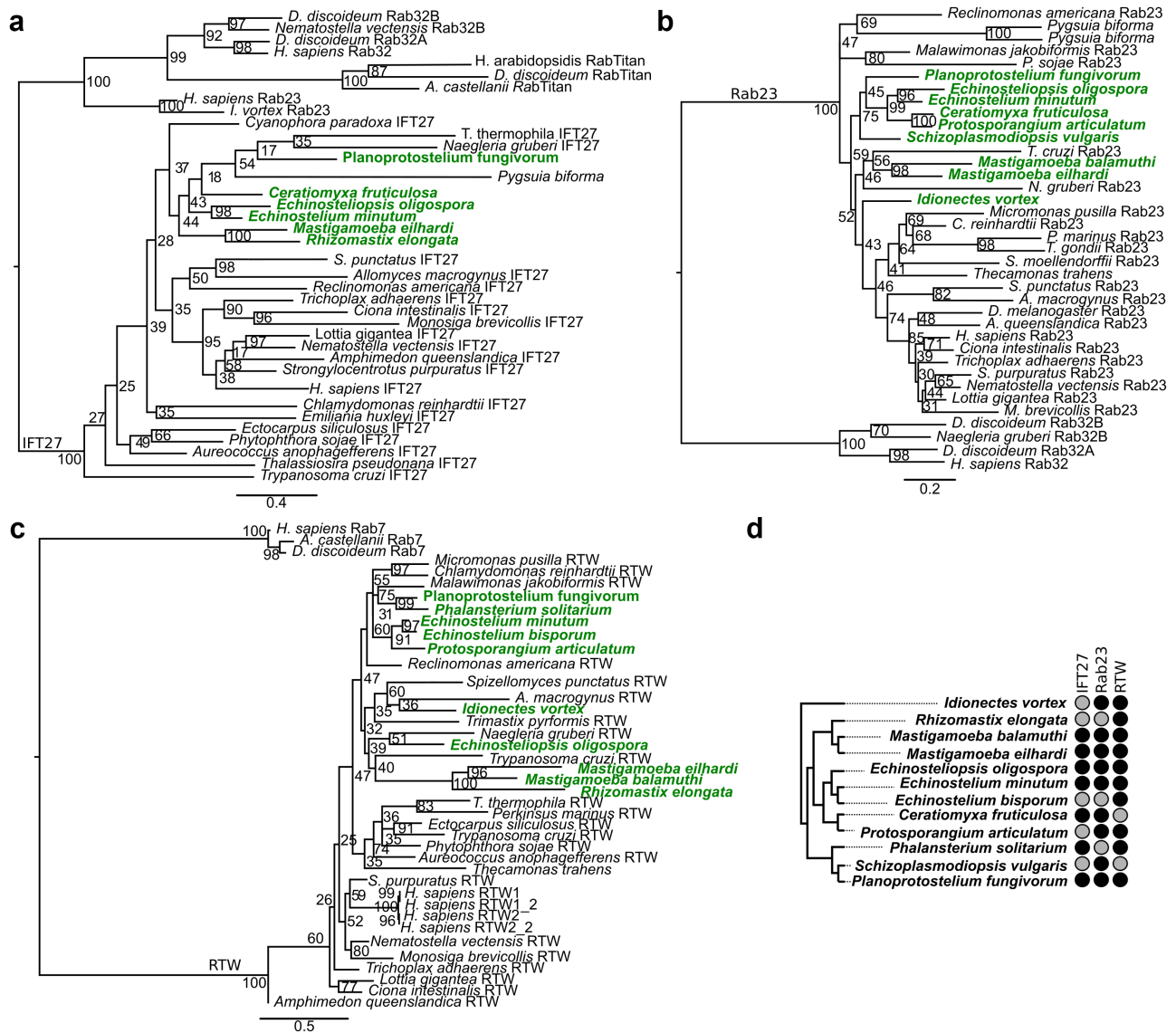

Figure 2. Phylogenetic reconstructions of Rab subfamilies IFT27 (RabL4), Rab23, and RTW (RabL2) and their presence profile in Amoebozoa. A. Maximum likelihood (ML) tree of subsets of the master dataset represented in **Supplementary Figure 2**, focusing on the IFT27 (RabL4) subfamily and the amoebozoans (highlighted in green) that have this Rab paralog. We included some non-amoebozoans and considered Rab23, Rab32A, Rab32B, and RabTitan clade as the outgroup. Values at nodes are ML bootstrap (BS) (1,000 ultrafast BS rep, IQ-TREE LG+I+ G4). **B**. Maximum likelihood (ML) tree of subsets of the master dataset represented in **Supplementary Figure 2**, focusing on the Rab23 subfamily and the amoebozoans (highlighted in green) that have this Rab paralog. We included some non-amoebozoans and considered Rab32A and Rab32B clade as the outgroup. Values at nodes are ML bootstrap (BS) (1,000 ultrafast BS rep, IQ-TREE LG+G4). **C**. Maximum likelihood (ML) tree of subsets of the master dataset represented in **Supplementary Figure 2**, focusing on the RTW (RabL2) subfamily and the amoebozoans (highlighted in green) that have this Rab paralog. We included some non-amoebozoans and considered Rab7 as the outgroup. Values at nodes are ML bootstrap (BS) (1,000 ultrafast BS rep, IQ-TREE LG+I+ G4). **D**. Presence profile of IFT27, Rab23, and RTW (RabL2) in Amoebozoa. These Rab subfamilies were identified exclusively in representatives of the Evosea group. Black circles represent that the paralog is present and grey circles represent that the paralog is absencet in the transcriptome observed. The phylogenetic representation of the evosean species was based on 25 and 40.
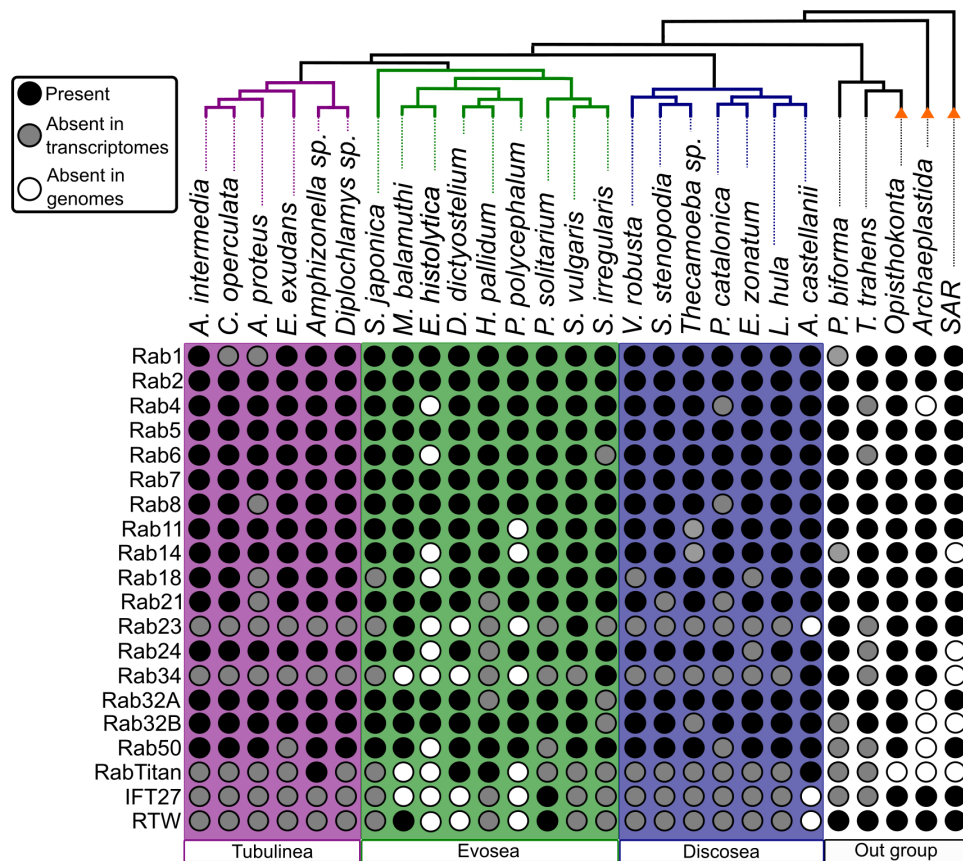
Figure 3. Presence profile in Amoebozoa of the 20 Rab subfamilies predicted to be present in the Last Amoebozoa Common Ancestor (LACA). We present the list of the 20 Rab GTPase subfamilies (left) identified in Amoebozoa considering the 44 Amoebozoa lineages sampled in the present study. For clarity in this figure, we only show 22 amoebozoans and few non-amoebozoan groups [top phylogeny based on 25, 65]. The figure is based on **Supplementary Figure 2**, the Rabifier annotation cross-validation (**Supplementary Table 2**), and 23. Orange triangles identify the phylogenetic tree leaves represented by several species, which we considered to generate the plot. The three ancestral paralogs not found in Amoebozoa [Rabs 20, 22, and 28] are not shown in the figure.

for the presence of Rabs 20, 22, and 28 in any of the three amoebozoan major groups. These three paralogs have been consistently lost in several eukaryotic groups. For example, most Fungi, Excavata, Archaeplastida, and SAR have none of these paralogs [23,24]. Conversely, choanoflagellates, Metazoa, and *P. biforma* have retained Rabs 20, 22, and 28 (**Supplementary Figure 2**). We identified Rab28 in the *T. trahens* genome (**Supplementary Figure 2**), a paralogs also conserved in the kinetoplastids *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major* [18,43,44].

In light of our phylogenetic reconstructions and the previous discussion, we show that the evolution of Rab family in Amoebozoa is characterized by the conservation of most of the ancestral Rab

representatives predicted to be present in LECA. This has been previously identified for single amoebozoan lineages with available genomes at that time (*D. discoideum*, *A. castellanii*, and *M. balamuthi*); here we show that this pattern of conservation of ancestral Rab subfamilies is robustly observed throughout the three Amoebozoa major lineages. This pattern of ancestral Rab conservation in all Amoebozoa major lineages contrasts to some eukaryotic lineages. For example, the whole rhodophyte red algae group (Archaeplastida 'supergroup') shows a pattern of massive loss of ancestral paralogs having only 6 of the 17 Rab paralogs presumably present in the ancestor of Archaeplastida [24]. Furthermore, most of the ancestral subfamilies of Rab that composed the last common ancestor of Amoebozoa have

expanded over the evolution and diversification of Amoebozoa.

Rab GTPase family has expanded in all amoebozoan major lineages

The 20 Rab subfamilies predicted to be present in LACA represent the base for innovations of Rab GTPase in Amoebozoa. From the 'prototypical' sequences of these subfamilies, many 'novel' in-paralogs originated through gene duplication across the evolution of Amoebozoa (for the approach behind the identification of 'novel' in-paralogs, please check (**Supplementary Figure 1 D – E**). We unambiguously identified 'novel' in-paralogs in seven of the 20 Rab subfamilies that we predicted to compose the last amoebozoan common ancestor; these are Rabs 1, 2, 5, 8, 7, 11, and 32A/B (**Supplementary Figures 3–10**).

Rab1 expansions

Amoebozoa have at least 17 'novel' in-paralogs identified as members of the Rab1 subfamily (EvoRab1B, EvoRab1C, EvoRab1D, EvoRabG1, EvoRabG2, EvoRab1E, DdiRab1E DdiRabA, DdiRabF1, MbaRab1C, MbaRab1E, EntRab1B, DisRab1B, DisRab1D, DisRab1E, DisRab1G, and AcaRab1F – **Supplementary Figures 3** and **10**). The 'novel' in-paralogs EvoRab1B, EvoRab1C, EvoRab1D, and EvoRabG1/G2, previously annotated in *Dictyostelium discoideum*, were also identified in other evoseans; while EvoRab1B, EvoRab1C, and EvoRabG1/G2 (duplicated in *D. discoideum*) are present in diverse Eumycetozoa, EvoRab1D was identified in members of the four groups of Evosea (Cutosea, Archamoebae, Eumycetozoa, and Variosea) (**Supplementary Figures 3** and **10**). EvoRab1E, a 'novel' in-paralog that was first identified here, was exclusively found in species of the Variosea (Evosea) group. The 'novel' in-paralogs DdiRab1E, DdiRabA, and DdiRabF1, previously annotated in *D. discoideum*, were not identified in other amoebozoan species sampled in the present study (**Supplementary Figures 3** and **10**). Similarly, the 'novel' in-paralogs MbaRab1C and MbaRab1E were identified exclusively in *Mastigamoeba*, while EntRab1B was identified exclusively in species of the *Entamoeba* genus. 'Novel' in-paralogs of the Rab1 subfamily were also identified in Discosea. The in-paralogs DisRab1B, DisRab1D, and DisRab1E, previously annotated in *Acanthamoeba castellanii*, are also present in other Centramoebida (Discosea) (**Supplementary Figures 3** and **10**). DisRab1G, also previously known exclusively in *A. castellanii*, was identified in multiple members of Flabellinia and Centramoebia (Discosea).

Conversely, we found the in-paralog AcaRab1F exclusively in *A. castellanii* (**Supplementary Figures 3** and **10**).

Rab2 expansions

Amoebozoa have at least five 'novel' in-paralogs of Rab2 (AmoRab2AC, TubRab2B, EvoRabQ, EvoRab2C, and DisRab2B – **Supplementary Figures 4** and **10**). The 'novel' in-paralog AmoRab2AC, previously annotated in *D. discoideum* (DdiRab2B), *M. balamuthi* (MbaRab2AC), and *A. castellanii* (AcaRab2AC), was also identified in members of the Tubulinea group and several other members of Evosea and Discosea (**Supplementary Figures 4** and **10**). Interestingly, two *E. histolytica* Rabs (EhRab2B and EhRab2C) branch as a member of the AmoRab2AC clade (**Supplementary Figures 4** and **10**), indicating this in-paralog is duplicated in *E. histolytica*. Also, we identified other 'novel' Rab2 in-paralogs exclusively to one of the three major groups of Amoebozoa. For the first time, we identified the 'novel' in-paralog TubRab2B, an in-paralog exclusively found in Tubulinea (**Supplementary Figures 4** and **10**). EvoRabQ, previously annotated in *D. discoideum*, was also identified in other members of Eumycetozoa (Evosea) (**Supplementary Figures 4** and **10**). Conversely, EvoRab2C, firstly identified here, was identified exclusively in Variosea (Evosea) (**Supplementary Figures 4** and **10**). The 'novel' in-paralog DisRab2B, previously annotated in *A. castellanii*, was also identified in other members of Centramoebia (Discosea) (**Supplementary Figures 4** and **10**).

Rab5 expansions

Amoebozoa have at least eight 'novel' in-paralogs of Rab5 (TubRab5B, EvoRab5B, EvoRb5C, EvoRab5D, AcaRab5B, AcaRab5C, AcaRab5L, and AcaRab5L2-**Supplementary Figures 5** and **10**). TubRab5B, an in-paralog first identified here, was exclusively identified in Tubulinea (**Supplementary Figures 5** and **10**). Three 'novel' Rab5 in-paralogs characterize the Evosea group; EvoRab5B was identified exclusively in the Variosea group species, EvoRab5C was identified in members of Eumycetozoa and Variosea, and EvoRab5D was identified in members of Eumycetozoa and Archamoebae (**Supplementary Figures 5** and **10**). The AcaRab5B, AcaRab5C, AcaRab5L, and AcaRab5L2, were found exclusively in *A. castellanii*; exceptionally, Rab5L2 were not annotated as a Rab5 by Rabifier, being annotated as RabX (**Supplementary Table 2**), although it branches as a member of the Rab5 clade and may represent a divergent member of this subfamily.

Rab7 expansions

Amoebozoa have at least 21 'novel' in-paralogs of Rab7 (AmoRab7B, TubRab7B, EvoRab7B, EvoRab7C, EvoRab7D, EntRab7B, EntRab7C, EntRab7E, EntRab7F, EntRab7G, EntRab7H, EntRab7I, DisRab7D, DisRab7D2, DisRab7F, AcaRab7B, AcaRab7C, AcaRabC2, AcaRab7E, AcaRab7H, and AcaRab7L – **Supplementary Figures 6 and 10**). AmoRab7B, previously annotated in *A. castellanii* (AcaRab7B), was identified in several species representing all the three major Amoebozoa groups (**Supplementary Figures 6 and 10**). TubRab7B is a 'novel' in-paralog of Rab7 that was exclusively identified in Tubulinea (**Supplementary Figures 6 and 10**). Evosea have three 'novel' Rab7 in-paralogs, EvoRab7B, EvoRab7C, and EvoRab7D. While EvoRab7B and EvoRab7D were exclusively identified in Eumycetozoa, EvoRab7C is represented by members of Eumycetozoa and Variosea (**Supplementary Figures 6 and 10**). EvoRab7B, EvoRab7C, EvoRab7D, previously annotated in *A. castellanii*, were also identified in other members of Centramoebia (Discosea) (**Supplementary Figures 6 and 10**). Conversely, the remaining 'novel' Rab7 in-paralog annotated in *A. castellanii* (AcaRab7B, AcaRab7C, AcaRabC2, AcaRab7E, AcaRab7H, and AcaRab7L) were exclusively identified in this species (**Supplementary Figures 6 and 10**).

Rab8 expansions

Amoebozoa have at least eight 'novel' in-paralogs of Rab8 (DdiRab8B, EntRab8B, EntRab8C, AcaRab8B, AcaRab8C, AcaRab8D, AcaRab8L, and AcaRab8L2 – **Supplementary Figures 7 and 10**). DdiRab8B, previously annotated in *D. discoideum*, was identified exclusively in this lineage (**Supplementary Figure 7**). EntRab8B and EntRab8C are 'novel' Rab8 in-paralogs identified exclusively in *Entamoeba* genus (**Supplementary Figures 7 and 10**). AcaRab8B, AcaRab8D, AcaRab8L, and AcaRab8L2, previously annotated in *Acanthamoeba castellanii*, were identified exclusively in this lineage (**Supplementary Figures 7 and 10**).

Rab11 expansions

Amoebozoa have at least 12 'novel' in-paralogs of Rab11 (TubRab11B, TubRab11 C, EvoRab11 C, EvoRab11D, DdiRab11B, MbaRab11B, MbaRab11 C, EntRab11B, EntRab11 C, EntRab11D, DisRab11B, AcaRab11 C – **Supplementary Figures 8 and 10**). TubRab11B and TubRab11 C, in-paralogs first identified here, were exclusively found in Tubulinea (**Supplementary Figures 8 and 10**). EvoRab11C and EvoRab11D were exclusively identified in Evosea, while members of Eumycetozoa have EvoRab11C (previously annotated in *D.discoideum*), EvoRab11D was identified exclusively in Variosea

(**Supplementary Figures 8 and 10**). Conversely, DdiRab11B is present exclusively in *D.discoideum* while MbaRab11B and MbaRab11C are present in *M. balamuthi* (**Supplementary Figures 8 and 10**). The 'novel' in-paralogs EntRab11B, EntRab11C, EntRab11D, previously annotated in *E. histolytica*, were identified exclusively in the *Entamoeba* genus (**Supplementary Figures 8 and 10**). DisRab11B was exclusively identified in Centramoebia, while AcaRab11C was identified exclusively in *A. castellanii* (**Supplementary Figures 8 and 10**).

Rab32A and Rab32B expansions

Amoebozoa have at least 12 'novel' in-paralogs of Rab32 (TubRab32AB, EvoRab32AB, EvoRab32 C, EvoRab32D, EvoRab32E, EntRab32AB2, AcaRab32B, AcaRab32 C, AcaRab32D, AcaRab32E, AcaRab32G, and AcaRab32H – **Supplementary Figures 9 and 10**). The 'novel' in-paralog TubRab32AB was identified exclusively in Tubulinea (**Supplementary Figures 9 and 10**). Evosea has five "novel" in-paralogs of Rab32, while EvoRab32AB was identified in members of Cutosea, Eumycetozoa, and Variosea, EvoRab32 C, EvoRab32D, EvoRab32E were identified exclusively in Eumycetozoa (**Supplementary Figures 9 and 10**). EntRab32AB2, newly identified here as a member of Rab32, was identified in several *Entamoeba* species sampled (**Supplementary Figures 9 and 10**). The in-paralogs AcaRab32B, AcaRab32 C, AcaRab32D, AcaRab32E, AcaRab32G, and AcaRab32H, previously identified in *A. castellanii*, were exclusively identified in this species (**Supplementary Figures 9 and 10**). Given the branching pattern of these 'novel' in-paralogs, we are not able to unambiguously assign them to Rab32A or Rab32B since they may be either a divergent 'novel' in-paralog of Rab32A or a divergent 'novel' in-paralog of Rab32B.

Altogether, the expansions observed in the subfamilies Rab1, Rab2, Rab5, Rab8, Rab7, Rab11, and Rab32A/B account for the total of 83 'novel' Rab in-paralogs currently identified only in Amoebozoa and that are present in at least one of its major lineages (**Supplementary Figures 3–10**). Most of these in-paralogs, mostly analysed in few amoebozoan lineages previously studied, are present in several species of Amoebozoa. Based on the pattern of presence observed for these 'novel' in-paralogs among the representatives of Amoebozoa sampled, we can presume in which ancestral these 'novel' in-paralog were already present (Figure 4). The current evidence indicates that independent Rab duplications leading to 'novel' in-paralogs may have occurred early in the evolution of each of the Amoebozoa major groups, for instance, TubRab2B (Tubulinea), EvoRab1D (Evosea), and DisRab1G (Discosea) (Figure 4). Some other in-paralogs may have appeared during the evolution of more inclusive
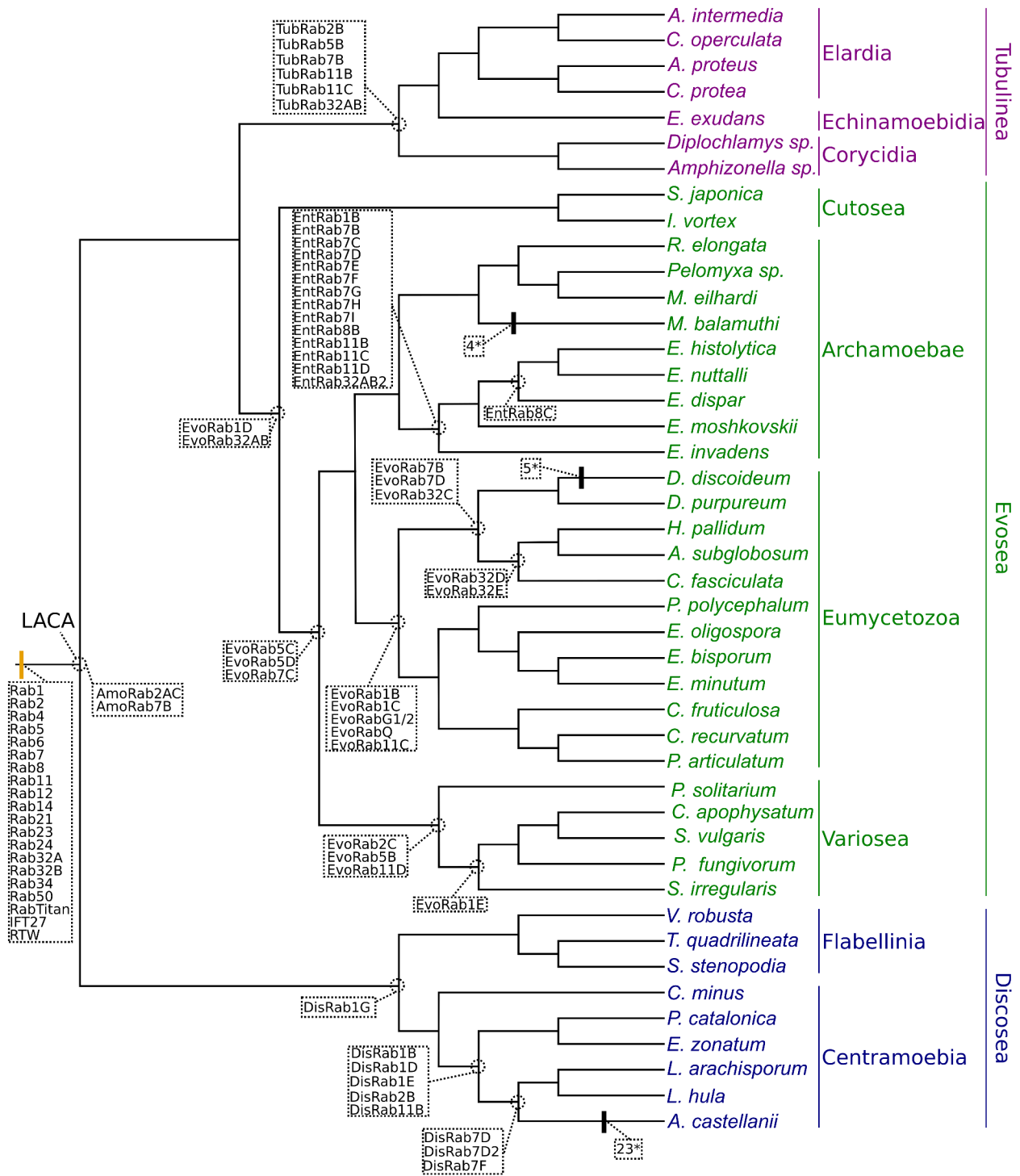
Figure 4. Representation of the presumed presence of the 'novel' Rab in-paralogs in ancestors of amoebozoan groups. The dashed boxes list the 'novel' in-paralogs presumably present in each ancestral indicated by dashed lines and circles. The numbers indicated with * represent the number of 'novel' in-paralogs exclusively identified in a single species. The orange bar represent the conservation in Amoebozoa of the 20 Rab predicted to have been present in the Last Eukaryotic Common Ancestor. The phylogenetic reconstruction representation was based on [25, 40, 66, and 67]. We named the 'novel' Rab in-paralogs based on the lineages they were identified and subfamily they compose. The Rab subfamilies are indicated by numbers and the members of the same subfamily are differentiated by letters. TubRab represents in-paralogs identified in multiple Tubulinea lineages, EvoRab represents in-paralogs identified in multiple Evosea lineages, DisRab represents in-paralogs identified in multiple Discosea, and EntRab represents in-paralogs identified in multiple Entamoeba lineages. AmoRab represents in-paralogs identified in at least one member of each of the three amoebozoan major lineages.

groups, such as EvoRab1B (Eumycetozoa), EvoRab2C (Variosea), EvoRab7B (Dictyostelia: Eumycetozoa), and DisRab7D (Acanthamoebidae: Centramoebia) or even in a single genus, for example, EntRab1B (*Entamoeba*), DdiRab11B (*Dictyostelium*), and AcaRab8B (*Acanthamoeba*) (Figure 4). Interestingly, our analyses indicate that two 'novel' Rab in-paralogs (AmoRab2AC and AmoRab7B), previously identified in few species, may have appeared early in the evolution of Amoebozoa and have been conserved in extant members of the three amoebozoan major groups (Figure 4). This finding indicates that LACA had at least 22 Rab paralogs, represented by 20 that were already present in LECA and 2 (AmoRab2AC and AmoRab7B) exclusively identified in Amoebozoa.

These results demonstrate that Rab GTPases have independently expanded in all amoebozoan major lineages. We highlight the massive expansion of robust subfamilies observed in Evosea and Discosea. However, it is worth noting that Tubulinea, the major group of Amoebozoa with the least expressive evidence of Rab expansions, has no genome available to date. Interestingly, most of the 'novel' in-paralogs exclusive to Amoebozoa are assigned as Rab 1, 7, 11, or Rab32. This finding corroborates the observation of recurrent duplications of specific paralogs in diverse lineages [11,24]. For example, diverse lineages of Archaeplastida have duplicated Rabs 1 and 11 multiple times [24], while several eukaryotic lineages have independently duplicated Rab 5 [11]. Recurrent gene expansions in Amoebozoa is not restricted to the Rab GTPase gene family. The genome of *D. discoideum* is characterized by the presence of ~2770 genes that have originated through recent gene duplications [45], *E. histolytica* have several gene families expanded, such as Arf, Rho GTPases, receptor Ser/Thr kinases, and cysteine proteases [46–48], while *M. balamuthi* have expanded kinase, cathepsin, guanylate cyclases, and cGMP-dependent phosphodiesterases gene families [48]. The patterns and correlation of these gene family expansions are yet to be elucidated based on sequencing and analyses of amoebozoan genomes. For instance, analyses of *E. histolytica* have demonstrated the link between the expansions of some families (e.g., Hsp70) with transportable elements, while tandem duplication, local inversion, and interchromosomal exchange account for most of the gene duplication identified in *D. discoideum* [45,47].

Among these 'novel' in-paralogs, several seem highly divergent from their 'prototypical' in-paralogs, based on their relatively longer branches and distribution pattern in our phylogenetic reconstructions. As proposed by other authors [*e.g.*, 24], relatively more divergent Rab in-paralogs may suggest the occurrence of neofunctionalization. In accordance, studies have successfully demonstrated some roles of Rabs that are characteristic to Amoebozoa [49,50]. For instance, EntRab11B, an 'novel' in-paralog of Rab11 subfamily identified in all *Entamoeba* species considered, is involved in the process of cysteine proteases secretion in *E. histolytica* and has a role in the pathogenicity of this species [49]. Interestingly, even some 'prototypical' in-paralogs (*i.e.*, conserved amoebozoan Rab in-paralogs that represent orthologs shared by diverse eukaryotic lineages) have unique cellular roles in Amoebozoa; for example, the 'prototypical' EntRab11A, other member of the Rab11 subfamily present in *Entamoeba*, may be involved in the encystation process of these organisms [51], while the 'prototypical' in-paralogs Rab7A and Rab5 of *E. histolytica* are involved with the function and biogenesis of the prephagosomal vacuole, a cellular structure characteristic to this species [52,53]. It is worth noting the massive expansion of the Rab7 subfamily in Amoebozoa, a subfamily involved with phagocytosis [53–55], that raises the question whether this expansion can be linked to a diversification of specialized phagocytosis in Amoebozoa. Thus, the diversity of Rabs identified in Amoebozoa, given the conservation and expansion of many Rab subfamilies, may underlie role innovations of this gene family in Amoebozoa that can be elucidated base on further studies of Rab functions in these organisms.

'Orphan' in-paralogs

In addition to the 'novel' in-paralogs assigned to one of the Rab subfamilies, several Rab GTPases of Amoebozoa are highly divergent and cannot be assigned to a Rab subfamily. These in-paralogs does not consistently branch as a member of one of the Rab clades analysed and, accordingly, are annotated as RabX by Rabifier (**Supplementary Table 2**). The amoebozoan 'Orphan' Rab in-paralogs annotated as RabX are spread along our master phylogenetic reconstruction and represent several paralogs identified in the genomes of *D. discoideum* (at least 31 RabX), *E. histolytca* (at least 61 RabX), *E. invadens* (at least 59 RabX), *E. moshkovskii* (at least 41 RabX), *E. dispar* (at least 46 RabX), *E. nuttalli* (at least 21 RabX), and *A. castellanii* (at least 36 RabX) (**Supplementary Table 2**); while some of the RabX identified in *Entamoeba* are shared between different species of this genus, most of the RabX paralogs identified in amoebozoans are exclusive to single species and are not present in the other amoebozoans and eukaryotes considered in this study. These observations corroborate the notion that the high

diversity of the Rab GTPase gene family in Amoebozoa impairs the unambiguous assignment of various Rab paralogs to a given subfamily or even identify the complete Rab repertoire of a given lineage, as noted by some previous studies [28,29]. Moreover, the vast repertoire of 'orphan' Rab in-paralogs present in Amoebozoa may represent a vast functional innovation and pseudogene origination of Rabs in this diverse eukaryotic group.

The quantitative disparity of RabX repertoire identified in lineages represented by sequenced genomes and those lineages represented by transcriptomes demonstrate the relevance of genomes to comprehensively assess the Rab GTPase gene family diversity. The abundance of divergent RabX is not exclusive to amoebozoans, some other groups deeply studied for Rab GTPases have diverse repertoires of divergent Rabs, for instance, *Trichomonas vaginalis* that has at least 51 divergent Rabs and *Tetrahymena thermophila* that has at least 42 divergent RabX [13,22]. It is worth noting the diversity and divergence of Rabs that compose the Rab repertoire of the parasitic amoebae *Entamoeba histolytica* **Supplementary Figure 2**; 26, 29). Besides having a vast number of Rab in-paralogs, most of them currently assigned as RabX, even in-paralogs successfully assigned to one of the known Rab subfamilies (*e.g., E. histolytica* Rabs 1 and 32A/B) seems to be highly diverging sequences based on their relatively longer and divergent branches (Figure 1 and **Supplementary Figure 2**). The identification of a large Rab family have been reported for other Eukaryotes [28], including for parasitic lineages [21,22]. Beyond underlying diversification of the eukaryotic cells, large and diverging Rab GTPase repertoires account for the potential of targeting Rabs to treat diseases caused by parasitic organisms, such as the parasitic amoebae *E. histolytica* [22,50,56].

Conclusions

Here, we present a comprehensive phylogenetic reconstruction and annotation of the Rab GTPase gene family in the 'supergroup' Amoebozoa. We demonstrate both the conservation of ancestral Rab paralogs in the extant representatives of Amoebozoa and the independent origin of 'novel' in-paralogs that occurred early in the evolution of Amoebozoa and in its three major lineages. From an amoebozoan ancestor with at least 22 Rab paralogs, each Amoebozoa major lineage diverged with different 'novel' in-paralogs. Several paralogs may even be restricted to more inclusive lineages (i.e., species, genus, or family). Our findings highlight that while key model organisms are useful as

a starting point for understanding biological phenomena, taking into account the phylogenetic diversity is crucial. Also, we identified a consistent higher diversity of Rabs in lineages represented by genomes, supporting that the Rab GTPase gene family's repertoire is yet to be revealed once more genomes become available, not only in Amoebozoa but also in other eukaryotic groups. Thus, the diversity and evolution of the Rab GTPases are still underrepresented. The high diversity and evolutionary pattern of Rab in Amoebozoa bring a robust base for future studies aiming to reveal the structure, biochemistry, cellular role, and functional innovations of this gene family that may be responsible for part of the diversity of Amoebozoa. Furthermore, the diversity of Rab repertoire identified in Amoebozoa highlights the potential to target Rabs in therapeutic interventions against parasitic amoebozoans. Finally, Amoebozoa represents a fruitful lineage to advance further the current understanding of the Rab GTPase gene family, taking advantage of the availability of a robust body of knowledge about the diversity and evolution of this 'supergroup'.

Material and methods

Amoebozoa genomes and transcriptomes dataset

We considered a dataset of genomes and deeply sequenced transcriptomes of 44 Amoebozoa lineages (**Supplementary Table 1**). These lineages compose the nine subclades and the three major lineages of Amoebozoa (**Supplementary Table 1**; see [25], for a phylogenomic study of Amoebozoa), constituting a representative sampling of this group's diversity. This amoebozoan dataset is the compilation of deeply sequenced transcriptomes generated by two different studies, [25] (BioProje/Rab GTPase family has expanded in all amoebozoan major lineagesct PRJNA380424) and [27] (BioProject PRJNA513164), as well as available genomes of Amoebozoa (**Supplementary Table 1**). Additionally, we included the transcriptome of *Pygsuia biforma* (BioProject PRJNA185780) and the genome of *Thecamonas trahens* (BioProject PRJNA37929) (**Supplementary Table 1**) to sample a breviate and an apusomonad, respectively, that together with Opisthokonta form Obazoa, the sister group of Amoebozoa. We performed all the analyses of this study based on amino acid sequences; thus, we predicted the ORFs and obtained the amino acid sequences of all genomes and transcriptomes considered through TransDecoder (https://github.com/TransDecoder/TransDecoder). We assessed the completeness of the genomes and transcriptomes through

BUSCO tool suite, searching for 255 single-copy orthologs expected to be present in eukaryotes [57], command used: *busco -i input_genome_transcriptome.faa -o output_file -m protein -l eukaryota_odb10.*

Rab GTPase identification and classification

We identified from the 44 Amoebozoa species, *P. biforma*, and *T. trahens* sequences similar to Rab GTPases through BlastP similarity search [58] implemented on BLAST® Command Line Applications (https://www.ncbi.nlm.nih.gov/books/NBK279690/). For that, we combined the translated transcriptomes and genomes of Amoebozoa, *P. biforma*, and *T. trahens* in a single dataset and created a custom blast database (command used: *makeblastdb – in combined_dataset. faa – dbtype prot*). We assembled a Rab GTPase query dataset compiling Rab sequences identified and annotated in *Homo sapiens*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Dictyostelium discoideum*, and *Entamoeba histolytica* available in online databases (dictybase – http://dicty base.org/; FLYtRAB – http://rablibrary.mpi-cbg.de/; GenBank – https://www.ncbi.nlm.nih.gov/genbank/). We used this query dataset for a BlastP search against the combined dataset of the 44 amoebozoans, *P. biforma*, and *T. trahens* (command used: *blastp -query input_Rab_query.faa -db combined_dataset.faa -evalue 1e-4 -out blastp_output_file*).

The sequences identified through BlastP composed our preliminary Amoebozoa, *P. biforma*, and *T. trahens* Rab GTPase dataset. We combined this dataset and the dataset curated by [23],(see [23], Table S1 for Rab's sequence ID or accession number of 55 different eukaryotic lineages) and built our preliminary master dataset. To curate our preliminary dataset, we performed multiple sequence alignments and phylogenetic tree inference (not shown) using Ran sequences as our outgroup. This approach enabled us to select the sequences representing Rab GTPase members from the sequences identified through BlastP since the similarity search identified Rab sequences and sequences that represent members of the other families that belong to the Ras superfamily [see [30], for a description of the Ras superfamily and its members' relationship]. We also analysed these phylogenetic trees to identified artifactual duplication patterns of the same protein in each lineage. The artifactual pattern consists of a single lineage presenting several slightly different sequences of a single protein (probably due to assembly artefacts or alternative RNA splicing). We manually curated our Amoebozoa, *P. biforma*, and *T. trahens* Rab GTPase dataset, excluding the artifactual duplications identified. Finally, we obtained the master Rab GTPase

dataset used in the present study, composed of 2,998 sequences (**Supplementary Material1**), including the 44 Amoebozoa species, *P. biforma*, *T. trahens*, and the dataset curated by [23]. To cross-validate our identification and assignments of Rab sequences, we classified the Rab sequences of Amoebozoa, *P. biforma*, and *T. trahens* with Rabifier2 [59].

Phylogenetic reconstructions

We performed the Rab GTPases' phylogenetic reconstructions based on multiple sequence alignments and phylogenetic tree inference by maximum likelihood. We performed the multiple sequence alignments with MAFFT [60], command used: *mafft input > output*. We performed automated alignment trimming to exclude and mask in the final alignments poorly conserved N- and C-terminal regions and highly variable internal regions using trimAl [61], command used: *trimal -in input -out output -gt 0.75*. We obtained trimmed Rab alignments composed of ~150 amino acids that we used for the phylogenetic analyses. We inferred all the maximum-likelihood trees using ModelFinder [62] and obtained node supports with the ultrafast bootstrap [63], both implemented in the IQ-TREE software [64], command used: *iqtree -s input -m TEST -bb 1000*. We considered Ran as an outgroup for the phylogenetic reconstructions that included all Rab subfamilies identified in Amoebozoa. Ran is the Ras superfamily member closest to Rab GTPase and has been proposed as the outgroup for phylogenetic studies of Rab GTPase [23,30]. For each phylogenetic reconstruction focusing on specific Rab subfamilies, we considered the closest Rab members as outgroup.

Disclosure statement

The authors declare no competing or financial interests.

Funding

Authors' contributions

ALPS conceived of the study, designed the study, carried out dataset curation, carried out the analyses, and drafted the manuscript; AT designed the study, helped carried out dataset curation and analyses, and critically revised the manuscript; MB designed the study, helped carried out dataset

curation and critically revised the manuscript; DL conceived of the study, designed the study, coordinated the study, participated in data analysis and helped draft the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

ORCID

Daniel J. G. Lahr ⓘ http://orcid.org/0000-0002-1049-0635

References

[1] Stenmark H, Olkkonen VM. The rab gtpase family. Genome Biol. 2001;2(5):reviews3007–1.

[2] Zhen Y, Stenmark H. Cellular functions of Rab GTPases at a glance. J Cell Sci. 2015;128 (17):3171–3176.

[3] Hutagalung AH, Novick PJ. Role of Rab GTPases in membrane traffic and cell physiology. Physiol Rev. 2011;91(1):119–149.

[4] Zerial M, McBride H. Rab proteins as membrane organizers. Nat Rev Mol Cell Biol. 2001;2(2):107–117.

[5] Pfeffer S. A model for Rab GTPase localization. Biochem Soc Trans. 2005;33(4):627–630.

[6] Colicelli J. Human RAS superfamily proteins and related GTPases. Science's STKE. 2004;2004(250):re13–re13.

[7] Asaoka R, Uemura T, Ito J, et al. Arabidopsis RABA1 GTPases are involved in transport between the trans-Golgi network and the plasma membrane, and are required for salinity stress tolerance. Plant J. 2013;73 (2):240–249.

[8] Gurkan C, Lapp H, Alory C, et al. Large-scale profiling of Rab GTPase trafficking networks: the membrome. Mol Biol Cell. 2005;16(8):3847–3864.

[9] Rutherford S, Moore I. The Arabidopsis Rab GTPase family: another enigma variation. Curr Opin Plant Biol. 2002;5(6):518–528.

[10] Vernoud V, Horton AC, Yang Z, et al. Analysis of the small GTPase gene superfamily of Arabidopsis. Plant Physiol. 2003;131(3):1191–1208.

[11] Brighouse A, Dacks JB, Field MC. Rab protein evolution and the history of the eukaryotic endomembrane system. Cell Mol Life Sci. 2010;67(20):3449–3465.

[12] Pereira-Leal JB. The Ypt/Rab family and the evolution of trafficking in fungi. Traffic. 2008;9(1):27–38.

[13] Bright LJ, Kambesis N, Nelson SB, et al. Comprehensive analysis reveals dynamic and evolutionary plasticity of Rab GTPases and membrane traffic in Tetrahymena thermophila. PLoS Genet. 2010;6(10):e1001155.

[14] Eisen JA, Coyne RS, Wu M, et al. Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol. 2006;4(9):e286.

[15] Ezougou CN, Ben-Rached F, Moss DK, et al. Plasmodium falciparum Rab5B is an N-terminally myristoylated Rab GTPase that is targeted to the parasite's plasma and food vacuole membranes. PloS One. 2014;9(2):e87695.

[16] Quevillon E, Spielmann T, Brahimi K, et al. The Plasmodiumfalciparum family of Rab GTPases. Gene. 2003;306:13–25.

[17] Langsley G, Van Noort V, Carret C, et al. Comparative genomics of the Rab protein family in Apicomplexan parasites. Microbes Infect. 2008;10 (5):462–470.

[18] Ackers JP, Dhir V, Field MC. A bioinformatic analysis of the RAB genes of Trypanosoma brucei. Mol Biochem Parasitol. 2005;141(1):89–97.

[19] Field MC. Signalling the genome: the Ras-like small GTPase family of trypanosomatids. Trends Parasitol. 2005;21(10):447–450.

[20] Fritz-Laylin LK, Prochnik SE, Ginger ML, et al. The genome of Naegleria gruberi illuminates early eukaryotic versatility. Cell. 2010;140(5):631–642.

[21] Carlton JM, Hirt RP, Silva JC, et al. Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science. 2007;315(5809):207–212.

[22] Lal K, Field MC, Carlton JM, et al. Identification of a very large Rab GTPase family in the parasitic protozoan Trichomonas vaginalis. Mol Biochem Parasitol. 2005;143(2):226–235.

[23] Eliáš M, Brighouse A, Gabernet-Castello C, et al. Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J Cell Sci. 2012;125(10):2500–2508.

[24] Petrželková R, Eliáš M. Contrasting patterns in the evolution of the Rab GTPase family in Archaeplastida. Acta Societatis Botanicorum Poloniae. 2014;83:4.

[25] Kang S, Tice AK, Spiegel FW, et al. Between a pod and a hard test: the deep evolution of amoebae. Mol Biol Evol. 2017;34(9):2258–2270.

[26] Nakada-Tsukui K, Saito-Nakano Y, Husain A, et al. Conservation and function of Rab small GTPases in Entamoeba: annotation of E. invadens Rab and its use for the understanding of Entamoeba biology. Exp Parasitol. 2010;126(3):337–347.

[27] Lahr DJ, Kosakyan A, Lara E, et al. Phylogenomics and morphological reconstruction of Arcellinida testate amoebae highlight diversity of microbial eukaryotes in the Neoproterozoic. Curr Biol. 2019;29 (6):991–1001.

[28] Diekmann Y, Seixas E, Gouw M, et al. Thousands of rab GTPases for the cell biologist. PLoS Comput Biol. 2011;7(10):e1002217.

[29] Saito-Nakano Y, Loftus BJ, Hall N, et al. The diversity of Rab GTPases in Entamoeba histolytica. Exp Parasitol. 2005;110(3):244–252.

[30] Wennerberg K, Rossman KL, Der CJ. The Ras superfamily at a glance. J Cell Sci. 2005;118(5):843–846.

[31] Huet D, Blisnick T, Perrot S, et al. The GTPase IFT27 is involved in both anterograde and retrograde intraflagellar transport. Elife. 2014;3:e02419.

[32] Eguether T, San Agustin JT, Keady BT, et al. IFT27 links the BBSome to IFT for maintenance of the ciliary signaling compartment. Dev Cell. 2014;31(3):279–290.

[33] Kanie T, Abbott KL, Mooney NA, et al. The CEP19-RABL2 GTPase complex binds IFT-B to initiate intraflagellar transport at the ciliary base. Dev Cell. 2017;42(1):22–36.

[34] Lim YS, Tang BL. A role for Rab23 in the trafficking of Kif17 to the primary cilium. J Cell Sci. 2015;128 (16):2996–3008.

[35] Lo JC, Jamsai D, O'Connor AE, et al. (2012). RAB-like 2 has an essential role in male fertility, sperm intra-flagellar transport, and tail assembly.

[36] Lumb JH, Field MC. Rab23 is a flagellar protein in Trypanosoma brucei. BMC Res Notes. 2011;4(1):190.

[37] Qin H, Wang Z, Diener D, et al. Intraflagellar transport protein 27 is a small G protein involved in cell-cycle control. Curr Biol. 2007;17(3):193–202.

[38] Wang Y, Ng EL, Tang BL. Rab23: what exactly does it traffic? Traffic. 2006;7(6):746–750.

[39] Yoshimura SI, Egerer J, Fuchs E, et al. Functional dissection of Rab GTPases involved in primary cilium formation. J Cell Biol. 2007;178(3):363–369.

[40] Hess S, Eme L, Roger AJ, et al. A natural toroidal microswimmer with a rotary eukaryotic flagellum. Nat Microbiol. 2019;4(10):1620–1626.

[41] Fiore-Donno AM, Tice AK, Brown MW. A non-flagellated member of the myxogastria and expansion of the echinosteliida. J Eukaryotic Microbiol. 2019;66(4):538–544.

[42] Reinhardt DJ, Olive LS. Echinosteliopsis, a new genus of the Mycetozoa. Mycologia. 1966;58(6):966–970.

[43] Field MC, Carrington M. Intracellular membrane transport systems in Trypanosoma brucei. Traffic. 2004;5(12):905–913.

[44] Field MC, Natesan SKA, Gabernet-Castello C, et al. Intracellular trafficking in the trypanosomatids. Traffic. 2007;8(6):629–639.

[45] Eichinger L, Pachebat JA, Glöckner G, et al. The genome of the social amoeba Dictyostelium discoideum. Nature. 2005;435(7038):43–57.

[46] Loftus B, Anderson I, Davies R, et al. The genome of the protist parasite Entamoeba histolytica. Nature. 2005;433(7028):865–868.

[47] Lorenzi HA, Puiu D, Miller JR, et al. New assembly, reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features and protein content information. PLoS Negl Trop Dis. 2010;4(6):e716.

[48] Žárský V, Klimeš V, Pačes J, et al. (2021). The Mastigamoeba balamuthi genome and the nature of the free-living ancestor of Entamoeba. Molecular biology and evolution, msab020.

[49] Mitra BN, Saito-Nakano Y, Nakada-Tsukui K, et al. Rab11B small GTPase regulates secretion of cysteine proteases in the enteric protozoan parasite Entamoeba histolytica. Cell Microbiol. 2007;9(9):2112–2125.

[50] Verma K, Srivastava VK, Datta S. Rab GTPases take centre stage in understanding Entamoeba histolytica biology. Small GTPases. 2020;11(5):320–333.

[51] McGugan GC, Temesvari LA. Characterization of a Rab11-like GTPase, EhRab11, of entamoeba histolytica. Mol Biochem Parasitol. 2003;129(2):137–146.

[52] Okada M, Nozaki T. New insights into molecular mechanisms of phagocytosis in Entamoeba histolytica by proteomic analysis. Arch Med Res. 2006;37(2):244–251.

[53] Saito-Nakano Y, Yasuda T, Nakada-Tsukui K, et al. Rab5-associated vacuoles play a unique role in phagocytosis of the enteric protozoan parasite Entamoeba histolytica. J Biol Chem. 2004;279(47):49497–49507.

[54] Rupper A, Grove B, Cardelli J. Rab7 regulates phagosome maturation in Dictyostelium. J Cell Sci. 2001;114 (13):2449–2460.

[55] Saito-Nakano Y, Wahyuni R, Nakada-Tsukui K, et al. Rab7D small GTPase is involved in phago-, trogocytosis and cytoskeletal reorganization in the enteric protozoan Entamoeba histolytica. Cell Microbiol. 2021;23 (1):e13267.

[56] Stein MP, Dong J, Wandinger-Ness A. Rab proteins and endocytic trafficking: potential targets for therapeutic intervention. Adv Drug Deliv Rev. 2003;55 (11):1421–1437.

[57] Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Gene prediction. Humana: New York, NY; 2019. p. 227–245.

[58] Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25 (17):3389–3402.

[59] Surkont J, Diekmann Y, Pereira-Leal JB. Rabifier2: an improved bioinformatic classifier of Rab GTPases. Bioinformatics. 2017;33(4):568–570.

[60] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780.

[61] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–1973.

[62] Kalyaanamoorthy S, Minh BQ, Wong TK, et al. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–589.

[63] Hoang DT, Chernomor O, Von Haeseler A, et al. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–522.

[64] Nguyen LT, Schmidt HA, Von Haeseler A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–274.

[65] Adl SM, Bass D, Lane CE, et al. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryotic Microbiol. 2019;66(1):4–119.

[66] Schilde C, Lawal HM, Kin K, et al. A well supported multi gene phylogeny of 52 dictyostelia. Mol Phylogenet Evol. 2019;134:66–73.

[67] Cui Z, Li J, Chen Y, et al. Molecular epidemiology, evolution, and phylogeny of Entamoeba spp. Genetics and Evolution: Infection; 2019. p. 75, 104018.