Title: Ancient Protein Resurrection of an ancestral Arf GTPase that regulates membrane trafficking uncovers dual cellular localization and unanticipated properties of modern Arf1 proteins.

Marine Alves[1], **Mandeep Sivia**[2], **Kristína Záhonová**[2,3,4,5], **Catherine L. Jackson**[1*], **Joel B. Dacks**[2,3,6*]

[1] Université Paris Cité, CNRS, Institut Jacques Monod, Paris, France

[2] Division of Infectious Diseases, Department of Medicine, and Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

[3] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

[4] Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

[5] Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Vestec, Czech Republic

[6] Centre for Life's Origins and Evolution, Department of Genetics, Evolution, & Environment, University College, London, UK

*to whom correspondence should be addressed CLJ (cathy.jackson@ijm.fr) and JBD (dacks@ualberta.ca)

Abstract: The evolution of eukaryotes, eukaryogenesis, marked a fundamental shift in cellular organisation for life on Earth, with the advent of the nucleus, mitochondria, and other compartmentalization. The organelles and machinery of the endomembrane system, being responsible for material synthesis, intracellular targeting, uptake and secretion, are key eukaryotic hallmark to be explained. The period of eukaryogenesis after the merger of prokaryotic lineages but preceding the Last Eukaryotic Common Ancestor, is inferred to have involved a rise in cellular complexity through expansion of organelle-specific protein machinery, but the steps and stages of organelle emergence is essentially a blackbox. This is particularly problematic since ancestral protein function is not necessarily inferable from modern components. Here we apply Ancient Protein Resurrection (Ancestral Sequence Reconstruction and molecular cell biological characterization) to explore the ancient evolution of the Arf GTPases. Arf1 is canonically understood to act in eukaryotic cells at the Golgi apparatus, while Arf6 acts at endosomes and the plasma membrane. We find that the inferred ancestral Arf1/6 protein acts at both Golgi apparatus and plasma membrane in both mammalian and yeast cells. Unexpectedly, we find that the mechanism of localization to the plasma membrane is due to a C-terminal motif found in Arf1 proteins from the wide diversity of modern eukaryotes, though not the best characterized models of yeast, plants, and humans. Our data suggests that the ancestral Arf protein acted at multiple locations in eukaryotic cells prior to the diversification of modern lineages and Arf1 proteins may do so to a greater extent than appreciated in organisms of medical and ecological importance today.

Keywords: eukaryogenesis, evolutionary cell biology, endomembrane system, Golgi apparatus, plasma membrane, Ancestral Sequence Reconstruction, LECA.

Introduction

The evolution of eukaryotes from their prokaryotic ancestors was an important transition in the history of life, resulting in nearly all complex life on earth today. Understanding of this series of events, eukaryogenesis, has seen significant progress recently, but many questions remain unanswered [1]. Eukaryotes arose from a symbiosis between at least two prokaryotic lineages. These lineages include an alphaproteobacterium which gave rise to the mitochondria, and an archaeal cell, which contributed at least some of the other organellar machinery [1–3]. The result of this endosymbiosis then evolved, acquiring hallmark eukaryotic traits and culminating in a population of single-celled eukaryotic organisms, the Last Eukaryotic Common Ancestor (LECA), which gave rise to the extant lineages inhabiting Earth today [1]. The major approach to understanding eukaryogenesis through these ancestors is comparative genomics of extant descendant organisms. For eukaryotic origins, the challenging investigation of the contributing prokaryotic lineages and earliest eukaryogenic events has advanced rapidly, with recent studies describing the Asgardarchaeota from which the eukaryotic host is derived [2], as well as the study of the alphaproteobacterial closest relatives of mitochondria [4]. The LECA marks the end of the eukaryogenic period and is a highly tractable reconstruction point by comparative genomics [5]. From a wide range of analyses, it is inferred to have been a population of microbial eukaryotes, possessing nearly the complete set of organelles observed in many modern cells. This configuration is underpinned by a large genome complement of molecular machinery, characterized by paralogous expansions of protein families [5]. Though challenges remain, this LECA reconstruction is considered to be robust and comprehensive.

One of the major hallmarks of eukaryotic cells that distinguish them from their prokaryotic ancestors is the endomembrane system. This interconnected system of internal membrane-bound organelles, including the endoplasmic reticulum (ER), the Golgi apparatus, endosomes and lysosomes, allowed compartmentalization of biochemical activities and is presumed to have provided a tremendous evolutionary advantage. The secretory pathway assures the delivery of secreted proteins, such as signalling molecules or proteins for predation and defence to the exterior of the cell, and membrane proteins to intracellular locations. The endocytic pathway assures the uptake of extracellular material including nutrients, as well as internalization of plasma membrane (PM) proteins such as signalling receptors for recycling or transport to the lysosome for degradation. In both the secretory and endocytic pathways, communication between organelles is assured by membrane-bound vesicular carriers.

The archaeal ancestor of eukaryotes provided some of the precursors of the protein machinery for vesicular transport found in extant eukaryotes [6–8]. Between this first eukaryotic ancestor and the LECA, extensive elaboration of the endomembrane system occurred, producing the endomembrane organelles. The LECA is inferred as having possessed a highly sophisticated set of membrane trafficking machinery, much of which is comprised of protein families, where the duplicated family members carry out similar biochemical functions, but at different transport pathways or organelles [5,9]. Thus a full understanding of eukaryogenesis requires explanation of how the endomembrane system evolved.

The molecular machinery of membrane trafficking in the endomembrane system includes regulators such as the Arf family GTPases, which are universally present in eukaryotes [10,11]. The Arf family includes Arf, Arf-like (Arl) and Sar GTPases, which are involved in different steps of membrane trafficking including vesicle budding, vesicle tethering, organelle movement, cytoskeleton-membrane interactions and lipid homeostatsis [12–14]. The Arf family GTPases carry out their functions by acting as molecular switches, inactive in their GDP-bound form and active when bound to GTP. Among small GTPases, Arf family proteins are unique in having an N-terminal amphipathic helix, often myristoylated, that mediates membrane binding of the active GTP-bound form [15]. The Arf family is ancient, as the eukaryotic proteins all arose from an ancestral group, the ArfR GTPases, which were present in the archaeal ancestor of eukaryotes [7,16]. The LECA is inferred to have already possessed 15 members of the Arf GTPase family [11]. The originally identified Arf1-6 proteins in mammalian cells, and their orthologues throughout eukaryotes, form a single group within the Arf family. LECA is predicted to have had both Arf1 and Arf6 proteins [11]. Arf1 has expanded in many lineages, including in mammals, to produce Arf1-5 [15]. Arf1 assures essential functions in the secretory pathway, including vesicle formation at the Golgi, while Arf6 proteins are involved in endocytosis pathways and actin cytoskeleton regulation at the PM [12,13,17,18] Arf1 is one of two Arf family proteins found in every eukaryote examined to date, whereas

Arf6 has been lost in a number of eukaryotic lineages, including early in the evolution of the Discoba and of the Archaeplastida (land plants and red, green and glaucophytes algal relatives) [11]. Because Arf1 and Arf6 are each other's closest relatives within the Arf Family [11], their emergence likely represents a relatively late step in eukaryogenesis, close to the LECA.

Comparative genomic analyses based on existing eukaryotic genomes can infer the predicted proteins present in LECA. When Asgard archaeal genomes are also included in comparative genomic analyses, this approach can predict the genes and their protein products that were present in the common ancestor of eukaryotes and the Asgard archaea, which predates LECA and the mitochondrial symbiosis event. However, this approach cannot currently be used to predict the features of our ancestors between these two points in the evolutionary time line, as no extant organisms originating from this evolutionary period have been identified. Given the complexity and sophistication of the LECA, this period is crucial to our understanding of eukaryogenesis, and hence new approaches are required to provide advances in this area. Because the LECA is inferred to have possessed many paralogues of membrane trafficking machinery [9*inter alia*], the challenge is to reconstruct the duplication steps giving rise to that complement. Determining the properties of the pre-duplicated ancestral proteins is precisely what is needed to reconstruct those most recent steps to LECA.

Ancient protein resurrection allows for the prediction of ancestral protein sequences and then the experimental testing of biochemical and cell biological properties of those proteins [19,20]. This approach has been used successfully for metabolic enzymes, transcription factors and structural proteins and back to deep evolutionary time points, including LECA proteins (e.g. [19,21,22]). In this study, we have used APR to move beyond the LECA boundary to reconstruct the ancestor of Arf1 and Arf6 GTPases and to explore whether the cell biological properties of the predicted Arf1/6 ancestral sequences resemble either Arf1, Arf6 or a combination of both.

Results

Ancestral Sequence Reconstruction

In order to perform ancestral sequence reconstruction, it was first necessary to construct a broadly sampled and representative dataset. Starting with the dataset from Vargová et al. 2021 [11], we updated the taxon sampling using genomes or transcriptomes of key eukaryotic lineages that had been sequenced in the intervening period and removed organisms redundant to our overall aim of a broad and even sampling of eukaryotes (Supplementary Table 1). Because the aim of the project was reconstruction of deep nodes in the Arf tree, lineage-specific expansions and highly divergent sequences were identified by preliminary phylogenetic analysis (Supplementary Figure 1A-D) of each of Arf1, Arf6, Arl1, and Arl5. The least divergent representatives were aligned and taxon-specific insertions were masked out such that they were not included in the ancestrally reconstructed sequence. This resulted in a final dataset of 363 sequences and 182 positions.

Under the assumption that the genes in question are vertically inherited [11], a constrained tree was constructed following the species relationships of the taxa sampled [23,24] (Supplementary Figure 2), with an assumed rooting on the Arl1/5 clade (see Methods). The branch lengths in the Arf6 clade are substantially longer than in the Arf1 clade, consistent with past observations of negative selection on Arf1 sequence evolution and divergence in Arf6 (Figure 1, Supplementary Figure 2). Ancestral Sequence Reconstruction was undertaken using this guide tree in both PAML and IQ-TREE (Figure 2A). Initial reconstruction was done with the optimally chosen model (LG+R8) by the embedded ModelFinder in IQ-TREE. In order to test the effect of a mixture model, we implemented a (LG+C60+R10) model within the IQ-TREE which was determined to be the optimal model (Comparative values) and used for final ASR (Supplementary Table 2).

Our ancestral reconstructions focused on three nodes of interest, the ancestor of all Arf1s, the ancestor of all Arf6s, and in particular the Arf1/6 ancestor. The ancestral Arf1/6 sequences predicted by PAML (Arf1/6PAML) and IQ-TREE (Arf1/6IQ) were very similar, sharing 92% identity (Figure 2A). For the Arf1/6 ancestral sequences, both programs predicted a large number of residues with high probability. 165/182 positions (90.66%) in the Arf1/6 IQ-TREE ancestor and 167/182 positions (91.76%) in the Arf1/6 PAML ancestor reconstructed with greater than 0.8 probability (Supplementary Table 2). Of those positions falling below this cut off, 10/17 in the Arf1/6IQ sequence and 7/15 in the Arf1/6PAML sequence conserved the basic chemical properties of the residue (Figure 2A). Notably, in seven cases, the PAML sequence included, as the residue reconstructed, the residue that was the second most likely from the IQ-TREE reconstruction. Reciprocally, there were four instances where the IQ-TREE sequence included the second most likely PAML residue at the same position. Thus we are confident that both the IQ-TREE and PAML reconstructions are robust and that together they cover the sequence space for possible ancestral reconstruction of the Arf1/6 ancestral protein (Figure 2A, Supplementary Table 2).

The ancestral Arf1/6 sequences predicted by PAML and IQ-TREE share a high level of identity to existing eukaryotic Arf1 and Arf6 sequences, especially in the nucleotide binding regions (G1 to G5, Figure 2B), which are highly conserved among Arf1 and Arf6 proteins themselves. The ancestral Arf1/6 sequences were more similar to Arf1 than to Arf6 proteins (Figure 2B). For example, Arf1/6PAML and Arf1/6IQ were 84% and 89% identical to human Arf1, and 73% and 72% identical to human Arf6, respectively. There was less conservation compared to individual eukaryotic Arf1 and Arf6 sequences in the N-terminal amphipathic helical (AH) region, which among eukaryotic sequences is also quite variable. However, the feature of this region that is important for Arf protein function is its amphipathic nature, which mediates tight membrane binding of the active GTP-bound form of the protein. The AHs of the Arf1/6 predicted ancestral sequences clearly have this amphipathic property (Figure 2C). In addition, both Arf1 and Arf6 are myristoylated on the glycine residue following the first methionine, and both predicted ancestral sequences have this feature (Figure 2B, C).

The C-terminal regions of Arf1 and Arf6 proteins across eukaryotes are quite variable, both in length and amino acid composition, likely explaining the lower confidence levels of some C-terminal residues in the predicted ancestral sequences in this region (Figure 2B, Supplementary Table 2). Because of the length variation observed at the C-terminus of Arf1 and Arf6 sequences across eukaryotes, the positions corresponding to position 182 in the initial predictions could be included or excluded as a discretionary decision. We chose to test predicted sequences that were 181 and 182 amino acids in length, with the latter having an additional lysine residue at the C-terminus.

Localization of Predicted Ancestral Proteins

In order to test the properties of the predicted ancestral proteins in cells, we chose two model eukaryotic systems, mammalian (MDCK) cells and budding yeast. Starting with MDCK cells, we expressed Arf1/Arf6 ancestral proteins predicted by both the PAML and IQ-TREE programs, testing two predicted versions by each program, as described in the previous section. All four predicted proteins were fused to GFP at their C-termini, and compared to human Arf1 and Arf6 expressed from the same vector. Western blot and in-gel GFP fluorescence analyses indicated similar levels of expression in mammalian cells (Supplementary Figure 3). The predicted ancestral proteins localized to the Golgi, as did Arf1, colocalizing with the Golgi marker GM130 (Figure 3). In contrast to Arf1, both Arf1/6PAML and Arf1/6IQ also localized to the PM, in a manner similar to Arf6, colocalizing with the 8+ PM probe (Figures 4 and 5). The major difference between the predicted ancestors and Arf6 that we observed is that the cytosolic pool of the Arf1/6 predicted ancestral proteins was less than that of Arf6, due to the fact that a large fraction of Arf1/6PAML and Arf1/6IQ also localized to the Golgi, unlike the case for Arf6. We found that the two versions of each predicted ancestral protein, differing in the number of C-terminal residues, had identical localization properties (Figure 5, Supplementary Figure 4).

We also tested localization of Arf1/6PAML and Arf1/6IQ in yeast cells. Western blot analysis indicated similar levels of expression in yeast (data not shown). The yeast Golgi is not a single, centrosomally localized organelle as in mammalian cells, but is fragmented into individual elements dispersed throughout the cytosol. We used yeast Arf1 fused to mCherry as a Golgi marker, as the majority of Arf1 at steady state is present at the Golgi, colocalizing with the early Golgi marker α-COP (Cop1), part of the COPI coat, and with the late Golgi marker clathrin heavy chain (Chc1) [25]. We found a high level of colocalization of both Arf1/6 predicted ancestral proteins with yeast Arf1 (Figure 6). In addition, unlike yeast Arf1, both Arf1/6PAML and Arf1/6IQ were also found at the PM (Figure 7). As was the case in mammalian cells, we found that the two versions of each predicted ancestral protein, differing in the number of C-terminal residues, had identical localization properties (Supplementary Figure 5). Hence in two different eukaryotic cellular systems, the Arf1/6 predicted ancestral proteins have localization properties of both

Arf1 and Arf6, localizing to both the Golgi and the PM, suggesting that the common ancestor was able to localize to both of these compartments.

Mechanism of Localization

Previous work has shown that one localization determinant for the Golgi in Arf1 is present in the alpha3 helix and downstream residues (aa 101 to 116 in *Homo sapiens* and *Arabidopsis thaliana* Arf1), and contains the MXXE motif [26–28]. This motif is lost in Arf6 proteins. However, when these amino acids are re-introduced into human Arf6, they are sufficient to target Arf6 to the Golgi [26]. We found that the predicted ancestral proteins have the MXXE motif, consistent with their Golgi localization. We also noted that the predicted ancestral proteins have multiple lysines at their C-termini, a known PM localization determinant [29–31]. We mutated these lysine residues to glutamines, and expressed the mutants (Arf1/6PAML-3KQ and Arf1/6IQ-4KQ) in mammalian cells. For both Arf1/6PAML and Arf1/6IQ predicted ancestral sequences, the mutant versions failed to localize to the PM (Figures 4 and 5). However, we observed Golgi localization for these C-terminal KQ mutants. Similarly, when the C-terminal KQ mutants of both Arf1/6PAML and Arf1/6IQ sequences were expressed in yeast cells, localization to the PM was lost, but localization to the Golgi was unaffected (Figure 7). We conclude that a property of the Arf1/6 predicted ancestral protein is that it had the capacity to localize to both the Golgi and the PM, and that PM localization is mediated by C-terminal basic residues.

Diverse Modern Arf1 Homologues Possess C-terminal Lysines

Because Arf6 localizes and functions at the PM, we expected that the PM localization of the Arf1/6 ancestor has its basis in Arf6 sequences. Indeed, the Arf6 predicted ancestral sequences do have multiple C-terminal basic residues. However, surprisingly, the predicted Arf1 ancestor (using either PAML or IQ-TREE) also has multiple C-terminal positively charged residues (Supplementary Table 2), Examination of the Arf1 sequences used in the ASR indicates that 31 of 120 Arf1 sequences in extant eukaryotes have multiple C-terminal basic residues. To further explore the prevalence of the polybasic C-terminal region in modern Arf1 proteins, we examined the broad diversity of eukaryotic Arf1 sequences in our broadest sampled dataset, a subset of which was used for the ASR. From a highly conserved

W (aa 172 in human Arf1) onwards, we examined sequences to determine the number of basic residues. We found that, with the exception of the plants and green algae, Arf1 with positively charged residues in the C-terminus are found all across the eukaryotic tree of life. A selection of Arf1 sequences with multiple basic residues, from a range of eukaryotes, and human Arf1 for comparison, is shown in Figure 8, and includes organisms of medical, and ecological impact. Notably, in addition to those Arf1 sequences with C-terminal basic residues, there were many cases in which a given taxon possessed multiple Arf1 homologues, and that used for the ASR was not the one with a highest number of C-terminal basic residues. This means that if anything the ASR analysis represents an under-estimate of the basic residue count in the Arf1/6 ancestral sequences. It also means that the C-terminal basic tail, and by inference the PM-localization of Arf1, is a prevalent feature of modern eukaryotic Arf1 proteins.

Discussion

In this work we reconstructed the ancestor of Arf1 and Arf6, a protein which predated LECA, and examined its characteristics. In the process of addressing this question of ancient evolution, we learned new properties of existing Arf1 and Arf6 proteins in the wide range of eukaryotes. Arf family proteins are key regulators of membrane trafficking, and are ancient, having arisen in the archaeal ancestor of eukaryotes [7,16]. These results place the Arf family GTPases at the very beginning of the process of eukaryogenesis. Between the archaeal ancestor and LECA, expansion of the Arf family proteins occurred, leading to the LECA complement of 15 members of this family, including Arf1 and Arf6 [11]. In well-studied eukaryotic systems, the major localization of Arf1 is at the Golgi, where a highly conserved function is in vesicular trafficking from the Golgi to either the ER or the endosomal system [18,32,33], whereas Arf6 localizes predominantly to the PM and endosomes, where it regulates the actin cytoskeleton and endocytic pathways [12,13,17]. Although having primary functions at different organelles, Arf1 and Arf6 are similar at the sequence level, even sharing many amino acids in the switch regions, which interact with effectors in the GTP-bound state [34–39]. Indeed, localization of mammalian Arf1 and Arf6 to different membrane sites is thought to be a major factor in their distinct functions. For all of these reasons, the Arf1/6 ancestor is a theoretical reconstruction point particularly well suited to the APR approach of addressing pre-LECA evolution.

We found that the predicted Arf1-6 ancestors are far more similar to Arf1 than to Arf6 proteins, and that these ancestors had the capacity to target the PM using a C-terminal polybasic motif. This predicted PM targeting motif, however, is a feature of the Arf1 and not the Arf6 sequences used in our reconstructions. The presence of C-terminal positively charged residues in many Arf1 proteins in eukaryotes today is surprising, and challenges the predominant view that Arf1 functions predominantly on internal organelles such as the Golgi in eukaryotic cells.

From the onset it is worth considering the limitations of our work and clarifying the steps that were taken to mitigate these factors. While ASR is of course a theoretical reconstruction, we took several steps to provide the most reasonable reconstruction and avoid biases. We sampled organisms from across the broad diversity of eukaryotes, providing a balanced sampling. We performed upstream analyses to select the least divergent representatives and we used two different ASR methods, both with the most sophisticated models of sequence evolution. The results of these analyses showed overall conservation with well-characterized Arf proteins (Figure 2). In fact these were nearly identical in some cases to Arf proteins found in existing organisms (e.g. *Planomonas*), which lies near the most recent predicted position of the root of the eukaryotic tree [40]. From an evolutionary perspective, Arf1 and Arf6 are one another's closest relatives in an unrooted tree of Arf family proteins, as compared to other possible reconstructions where there are other protein subfamilies intervening. We did assume that the root of the Arf tree does not lie on Arf1, Arf6 or basally to Arf1/6, making the Arl1/5 clade an outgroup. Although this is a declared assumption, we believe that it a strong one. Past attempts to root the Arf family tree with the recently described ArfR archael homologues were unresolved but gave no indication of a root on the Arf1, Arf6, or Arf1/6 groups [7]. Furthermore, a rooting on Arf1, Arf6 or Arf1/6 would then imply that all other Arf family sub-families are derived from Arfs and evolved at a later evolutionary point. This would include the COPII GTPase Sar1. We believe this is unlikely, particularly since other components of the COPII machinery (e.g. Sec23/24) homologues have been identified in Asgardarchaea [6]. Finally, recognizing the theoretical nature of our reconstructions, we have limited our conclusions to broad properties of the sequences, rather than specific amino acids, and corroborated with experimental data or sequences from existing organisms. Even with the steps taken for robustness in our predictions, we chose to test sequences derived from both methods to cover the range of sequence that we had reconstructed and balance against methodological bias. We also chose to perform our analyses in two model organisms, those that are the best studied and from which a large proportion of the biological knowledge about Arf family proteins have been determined. Importantly, we observed experimental consistency in the two cell biological model systems and in the predicted ancestral sequences reconstructed using both methods. With all of these efforts for robustness, we believe that there are important conclusions that can be drawn.

Arf1 and Arf6 are localized to membranes through multiple mechanisms, including recruitment of the GDP-bound form through a membrane-bound receptor, interaction with specifically localized Arf guanine nucleotide exchange factors (GEFs) and by activation and interaction with effectors. The complete mechanism of localization for these GTPases has not been fully resolved [18]. Although in vitro, Arf1 and Arf6 are soluble in their GDP-bound forms, in cells the GDP-bound forms of both GTPases have been shown to bind to membranes via protein-protein interactions. Arf6 in its GDP-bound form localizes to the PM [41] and interacts specifically with the TBC domain proteins TRE17 and TBC1D24/TBC-K, mediating PM association [42,43]. Arf1 has been shown to bind to at least three Golgi proteins in its GDP-bound form, the Golgi-localized SNARE protein membrin [26,28] and members of the p24 family [44,45] in mammalian cells, and to members of the ArfGAP2/ArfGAP3/Glo3 family of ArfGAPs in plant cells[46].

The interaction of Arf1-GDP with membrin is the best studied, and has been shown to mediate specific localization of the Arf1 GTPase to the Golgi [26–28]. The region of Arf1-GDP that interacts with membrin, the alpha3 helix and four residues downstream, is sufficient for Golgi localization, and has a markedly different sequence in Arf6 [26]. The importance for Golgi localization of the MXXE motif within this region was first demonstrated in mammalian cells [26] and was also shown for a plant Arf1 orthologue [27,28]. Indeed, mutagenesis studies showed that mutations in the M110 and E113 residues within the MXXE motif of both *Homo sapiens* and

Arabidopsis thaliana Arf1 reduce Golgi localization [26,27]. Our results show that this feature is widespread throughout eukaryotic Arf1 orthologues, but is not found in Arf6 sequences. Our studies predict that it was present in the ancestor of Arf1 and Arf6 proteins.

Even though Arf6 proteins do not have a C-terminal polybasic motif, there is one known mechanism of Arf6 localization to the PM that does depend on one. This mechanism is via an Arf cascade in metazoan cells, whereby PM-localized Arl4 recruits the Arf6 activator (GEF) cytohesin, which in turn activates and recruits Arf6 [47–49]. There are three Arl4 proteins in mammalian cells (Arl4A, C, D), all of which have nine lysine and/or arginine residues at their C-termini that are required for PM localization through binding to anionic PM lipids such as PI(4,5)P_2 [31,47]. This Arl4-dependent mechanism of Arf6 localization was established late in evolution, as Arl4 proteins arose in early metazoans, likely from Arl19. Arl19 arose in the stem branch of Holozoa, possibly from Arf1 [15]. Arl19 proteins in a number of eukaryotes also have multiple basic residues at their C-termini [11].

This work has important implications for understanding both the cell biology of Arf proteins in modern eukaryotes and the process of eukaryogenesis. Our results predict that the eukaryotic ancestor of Arf1 and Arf6 proteins localized to the Golgi and the PM via a mechanism that was maintained in a number of modern lineages. Arf1 paralogues with three or more C-terminal lysines are present in diverse eukaryotes including important parasites of global health importance (e.g. *Giardia intestinalis*), economically relevant agricultural or veterinary pathogens (e.g. *Tritrichomonas foetus*, *Ustilago maydis*, *Plasmodiophora brassicae*) or protists with key ecological roles (e.g. the diatom *Thallasiosira pseudonana,* the dinoflagellate *Symbiodinium minutum,* or the heterotrophic flagellate *Paradiplonema papillatum*). For these organisms, we predict that some of their Arf1 proteins will not follow the current paradigm for Arf1 and will localize to the PM as well as the Golgi. Even in mammalian cells, the idea that Arf1 proteins have broader functions than at the Golgi has been gaining support, where Arf1 has been shown to have functions at the PM, for example in phagocytosis, cell migration, neutrophil chemotaxis, and the CLIC/GEEC endocytic pathway [50–56]. In addition, Arf6 itself can recruit cytohesin which then recruits Arf1 to the PM in another Arf activation cascade [36,37]. This mechanism enables a strong and sustained activation signal at the PM, as Arf1 is more abundant than either Arf6 or Arl4 [37]. Our results support the conclusion that the multifunctionality of Arf1 is an evolutionarily conserved feature and that the paradigm of Arf1 functionality should include PM activity as part of its broader role.

With respect to the events and processes of eukaryogenesis, our results strongly suggest that the Arf1/6 ancestor was a generalist protein acting at multiple cellular membranes in organisms within the LECA population. While ancient protein resurrection has been used to investigate deep nodes within the eukaryotic tree [22] and even aspects of LECA proteins [21], to our knowledge, this is the first attempt to reconstruct a pre-LECA protein. In this case it appears that a multifunctional protein duplicated, with one paralogue retaining its generalist nature and the other paralogue undergoing divergence and loss in many cases. Loss after duplication is a common result but long-term retention, divergence, and sporadic loss over a 1.5 billion year span suggests complex cell biological involvement. This requires a much deeper understanding of Arf6 function to explain, particularly in non-opisthokont organisms where cell biological information is lacking. Nonetheless, our work does demonstrate that the ASR approach to understanding events in the intervening eukaryogenic period between first eukaryotic common ancestor (FECA) and LECA is tractable and can yield concrete results. We contend that it opens up a highly promising new avenue of investigation for an important and currently intransigent problem in evolutionary biology.

Methods

Homology Searching

Our sampling was build upon that from Vargová et al. 2021 [11]. Nonetheless, to incorporate sampling from newly sequenced or described taxonomic lineages (e.g. ancyromonads, CRuMs, Telonemia), we searched 24 additional ‘omics dataset (Supplementary Table 1). Human Arf1, Arf6, Arl1, and Arl5 were used as the query sequences to search into the dataset for Arf family homologs with the human genome being used as the reference genome.

Homology searching for the Arf family paralogs were carried out with Analysis of MOlecular Evolution with BAtch Entry (AMOEBAE) [57]. Forward searches into the assembled database were carried out with BLASTp to identify potential Arf homologs that have sufficient E-value support(E<0.05). Potential homologues were used as queries to search into the reference genome to check if confirmed Arf family subunits are also detected with proper E value support (E<0.05 and must be a factor of 10^2 less than next nonredundant hit). Sequences that were clearly fragmentary were discarded. The remaining data was combined with a selection from Vargová et al. 2021 [11].

Phylogenetic Analyses

For computational tractability, the dataset was then subdivided into subdatasets for Arf1, Arf6, Arl1, and Arl5. Alignment was performed with MUSCLE (V5.1) and masked/trimmed manually using Mesquite (V3.70). IQ-TREE (V2.2) was run for 1,000 ultrafast bootstraps for each of the Arf subfamilies with default model finding criteria using the embedded Model finder (analysis details in legends of Supplementary Figure 1A-D). The least divergent paralogue (shortest branch length) for each of the Arf families analyzed was collected for further analysis for a total of 120 Arf1, 56 Arf6, 108 Arl1, and 79 Arl5 homologues across a total of 120 species and an alignment of 182 position.

Ancestral Sequence Reconstruction

The datasets were recombined and aligned using MAFFT and manually trimmed to remove unique indels. An unresolved guide tree was imposed, following the most up to date topology of the eukaryotic tree of life [23,24] and respecting the sister clade relationship of Arf1/Arf6 and Arl1/Arl5 [11]. The Arf1 and Arf6 tree was calculated with a rooted constraint of the Arl1/Arf5 clades that are the closest sister clades in an unrooted phylogeny from Vargová et al. 2021[11] as an assumed outgroup.

Ancestral sequence reconstruction using the constrained guidetree and an alignment of eukaryotic Arfs was performed using an empirical (model=2) non mixture LG model on Phylogenetic Analysis by Maximum Likelihood (PAML V4.8) based on the results of IQ-TREE ModelFinder. Ancestral sequence reconstruction was also performed using a LG+C60+R8 mixture model in IQ-TREE, derived from a ModelFinder search for the most appropriate mixture model. The ASR was also calculated for the Arf1/6 ancestors [20]. Because of the length variation observed at the C-terminus across eukaryotes, the positions corresponding to positions 182-183 could be included or excluded as a discretionary decision. To be conservative, we eliminated the positions, resulting in final sequences of the Arf1 ancestor (181 positions in both IQ-TREE and PAML) ending with a chain of three lysine residues and the Arf1/6 ancestor (181 positions in both IQ-TREE and PAML) with four lysines (Figure 2).

The N-terminal AHs of the PAML and IQ-TREE predicted ancestral sequences were determined from alignment with human Arf1 and Arf6, whose AH regions have been determined by crystal structure analysis [58,59]. Helical wheel plots were generated by Heliquest [60].

Cell Culture and Plasmids

MDCK cells (ATCC CCL-34) were grown in Dulbecco's modified Eagle's medium (DMEM) (Gibco #31885023) supplemented with 10% FBS (Gibco #A5256701) and 1% penicillin-streptomycin (Gibco #15140122). For transfections, cells were plated at 50% confluency in 6- or 24-well plates, and containing 1.5 mm coverslips for immunofluorescence microscopy. After 24 hours cells were transfected with 1 µL/cm^2 of Lipofectamine 2000 (Thermofisher #11668027) and a total of 0.5 µg/cm^2 of DNA, including plasmids of interest and carrier DNA where appropriate. After 48 hours, cells were fixed and prepared for immunofluorescence microscopy, or cell lysates were prepared either for western blotting or for in gel fluorescence (IGF) using a ChemiDoc MP imager (Bio-Rad) [61]. As a loading control, total proteins were visualized in commercial pre-cast SDS–PAGE gels (stain-free TGX gels, 4%–20%; Bio-Rad) after photoactivation induced by a 45 second UV treatment using a ChemiDoc MP imager (Bio-Rad).

Yeast transformations were carried out using the lithium acetate method. Briefly, the BY4742 *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* strain [62] was grown in YPD medium at 30°C to exponential phase, and approximately 4 OD_{600} equivalents of cells per transformation reaction were harvested, washed and resuspended in 60 µL of transformation solution 1 (10 mM Tris pH7.5, 100 mM LiAc, 1 mM EDTA). 2 µg of plasmid DNA and carrier DNA from salmon testes (Sigma #D9156-1ml, incubated at 95°C for 5 min), were added, then 300 µL of 40% PEG in transformation solution 1. Cells were incubated 30 min at 22°C or 30°C, then at 42°C for 15 min. Cells were resuspend in YPD, incubated 1 hour, then plated on selective medium. For live cell imaging, cells were grown in selective liquid medium, to a maximum OD_{600} of 0.3, concentrated, then observed by spinning disk microscopy. For western blotting analysis, cells were grown to a maximum OD_{600} of 0.5, and cell lysates prepared. After transfer of gels, nitrocellulose membranes were incubated with anti-GFP primary antibodies (Roche #11814460001) and anti-mouse IgG-HRP secondary antibodies (GE Healthcare, #NA931V). As a loading control, total proteins were visualized in commercial pre-cast SDS–PAGE stain-free gels after photoactivation induced by a 45 second UV treatment using a ChemiDoc MP imager (Bio-Rad).

Plasmids used in this study are listed in Supplementary Table 3. The 8+mCherry PM probe was derived from the 8+GFP PM probe described previously [29].

Immunofluorescence

The transfected MDCK cells previously seeded on 1.5 mm coverslips were fixed with 4% paraformaldehyde (PFA) for 15 min. PFA was inactivated with 50 mM NH_4Cl for 20 min. After 2 1× PBS washes, cells were permeabilized for 30 min with 1× PBS, 0.025% saponin (Sigma-Merck #84510), then blocked with 1× PBS, 0.025% saponin, 1% BSA for 30 min. The primary mouse antibody targeting GM130 (BD Biosciences #610822) was hybridized in 1× PBS, 0.025% saponin, 1% BSA at 1:500 overnight at 4°C. After multiple washes, goat anti-mouse IgG Alexa fluor 647 (ThermoFisher #A-21235) was added at 1:500 for 1 hour. After multiple washes, nuclei were stained using Hoechst33342 (ThermoFisher #H3570) in 1× PBS at 1:10,000 for 15 min. Coverslips were washed again, rinced in water, then mounted using Fluoromount (ThermoFisher #00-4958-02). After 24 hours at room temperature, the cells were observed by confocal microscopy.

Microscopy

For MDCK cells, acquisitions were performed either on a Zeiss confocal 980 Airyscan microscope driven by ZEN software (version 3.9.101.03000), using a 63× objective, or on a Leica Spinning Disk CSU-X1 microscope driven by MetaMorph software (version 7.10.4.407), using a 100× objective. For the images obtained using the Zeiss confocal 980 microscope, 40 z-sections separated by 0.13 µm were acquired for each field of cells using excitation/emission wavelengths 353/465 (Hoechst), 488/509 (GFP), and 587/610 (mCherry), and/or 653/668 (Alexa 647). For the Leica Spinning Disk microscope, 20 z-sections separated by 0.2 µm were acquired for each field of cells using excitation wavelengths of 405 (Hoechst), 488 (GFP), 561 (mCherry), 642 (Alexa 647), and emission filters 450/60 (Hoechst), 525/50 (GFP), 590/35 (mCherry), 670/40 (Alexa 647). Localization of GFP-tagged proteins and compartment markers was analysed in ImageJ2 V2.14.0/1.54f. Median and gaussian blur filters were applied to image stacks, and a projection of four z-sections was used for quantifications. A six-pixel wide line was drawn across a portion of the compartment of interest (PM or Golgi), identified by specific markers (8+mCherry or GM130, respectively). Examples are shown in Figures 3 and 4, yellow lines. Fluorescence intensity in each channel, expressed as a gray value, was obtained for each pixel along the line. Plots of fluorescence intensities across the yellow lines in the upper right panels of each part of Figures 3 and 4 are shown in the bottom right panels. To determine the membrane/cytosol ratio (Figure 5), the value of the peak of fluorescence at the PM was determined, and the average cytoplasolic fluorescence intensity was obtained by selection of a portion of the line passing over cytosol free of organelles, and averaging the fluorescence values obtained. The membrane/cytosol ratio for PM staining was calculated by dividing the membrane fluorescence value by the average cytosolic fluorescence value. For this quantification of PM localization, 20-53 cells from separate fields were analysed (numbers for each condition are given in Supplementary Table 4). The number “n” refers to the number of individual cells quantified for each plasmid transfected, chosen from different microscopy fields and from at least three different transfection experiments. Mean and standard deviation of the calculated membrane/cytosol ratios were obtained, and statistical analysis (unpaired t test with Welch's correction, two-tailed p value) was performed, using Graphpad Prism software (V10.5.0). ****

indicates $p < 0.0001$; **, $p < 0.01$; ns, not significant. The values of p for each pairwise comparison are given in Supplementary Table 5. Numerical values of the raw data obtained, as well as the results of statistical analyses are in Supplementary Table 4.

For yeast cells, acquisitions were performed on a Leica Spinning Disk CSU-X1 microscope, driven by Metamorph software, using a 100× objective. 21 z-sections separated by 0.2 µm were acquired for each field of cells, using excitation wavelengths of 488 (GFP) and 561 (mCherry), and emission filters 525/50 (GFP) and 590/35 (mCherry). Localization of GFP-tagged proteins was analysed using ImageJ2 by selecting an appropriate z-section, choosing cells with well-defined organelle localization, and drawing a line across the cell (examples are shown in Figures 6 and 7, yellow lines). Fluorescence intensity was obtained for each pixel along the line. Plots of pixel intensities across the yellow lines in Figures 6 and 7 are shown in the right panels. At least three independent transformations were performed for each plasmid, including appropriate controls each time.

Acknowledgements

We acknowledge the ImagoSeine core facility of Institut Jacques Monod, member of France-BioImaging (ANR-10-INBS-04) and IBiSA, with the support of Labex "Who Am I", Inserm Plan Cancer, Region Ile-de-France and Fondation Bettencourt Schueller. Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic. This study was supported by the CNRS, France and grant ANR-20-CE13-0007 from the ANR, France to CLJ. Work in the Dacks lab is supported by grants from the Natural Sciences and Engineering Research Council of Canada (RES0043758 and RES0046091).

Figure Legends

Figure 1: Illustrated constraint tree calculated for Arf1, Arf6, Arl1, and Arl5. This phylogeny shows the more divergent nature of Arf6 sequences as compared with Arf1. Clades are collapsed for clarity (for full tree see Supplementary Figure 2).

Figure 2: Properties of the Arf1/6 ancestral proteins predicted by PAML and IQ-TREE. (A) Alignment of the two predicted ancestral sequences, Arf1/6PAML and Arf1/6IQ. Key G-Boxes and and sequence motifs are annotated above. Residue-specific probabilities are colour-coded as inset. (B) Alignment of Arf1/6PAML and Arf1/6IQ predicted ancestral sequences and selected eukaryotic Arf1 and Arf6 proteins. Identical residues in 80% or more (all but 2) of the sequences are indicated by dark purple boxes; in 70% (all but 3) by light purple boxes. Same motifs as in part A are annotated. (C) Helical wheel plots of N-terminal AHs of some of the sequences shown in part B. Helical wheel representations of the N-terminal AHs, generated by Heliquest [60], are shown. Color code for the amino acids is yellow: Val, Leu, Ile, Met, Phe, and Trp; dark purple: Ser and Thr; red: Asp and Glu; pink: Asn and Gln; blue: Lys and Arg; grey: other residues. Arrow represents the helical hydrophobic moment, with length proportional to the calculated value and direction indicating that the helix is amphipathic in the perpendicular direction.

Figure 3: Localization of Arf1/6 predicted ancestral proteins to the Golgi in MDCK cells. MDCK cells were transfected with GFP (A); Arf1-GFP (B); Arf6-GFP (C); ARF1/6PAML-GFP (D); ARF1/6IQ-GFP (E); ARF1/6PAML-3KQ-GFP (F); or ARF1/6IQ-4KQ-GFP (G), and prepared for immunofluorescence analysis using antibodies against GM130, an early Golgi marker (magenta) and Hoechst dye to visualize nuclei (blue). Representative merged images are shown on the left for each transformant, with GFP-tagged protein (green), Golgi (magenta) and nuclei (blue). Insets (yellow square) are shown on the right, with the GFP and Golgi channels shown separately (left inset panels). The yellow lines shown in the merged inset panels (upper right) are those used to generate plots of pixel intensities for the GFP-tagged proteins (green) and the GM130 Golgi marker (magenta) (lower right) of each part (A-G). At least three independent transformations were performed for each mutant plasmid, always including the appropriate control plasmids. Scale bar, 5 μm.

Figure 4: Localization of Arf1/6 predicted ancestral proteins to the PM in MDCK cells. MDCK cells were co-transfected with GFP (A); Arf1-GFP (B); Arf6-GFP (C); ARF1/6PAML-GFP (D); ARF1/6IQ-GFP (E); ARF1/6PAML-3KQ-GFP (F); or ARF1/6IQ-4KQ-GFP (G), and the 8+mCherry PM probe, then prepared for immunofluorescence analysis using Hoechst dye to visualize nuclei (blue). Representative merged images are shown on the left for each transformant, with GFP-tagged protein (green), PM (magenta) and nuclei (blue). Insets (yellow square) are shown on the right, with the GFP and PM channels shown separately (left inset panels). The yellow lines shown in the merged inset panels (upper right) are those used to generate plots of pixel intensities for the GFP-tagged proteins (green) and the PM marker (magenta) (lower right) of each part (A-G). At least three independent transformations were performed for each mutant plasmid, always including the appropriate control plasmids. Scale bar, 5 µm.

Figure 5: Quantification of PM localization. Levels of GFP-tagged proteins at the PM in individual cells (n) from the experiments shown in Figure 4 and Supplementary Figure 4 were quantified. For GFP, n=30; Arf1-GFP, n=30; Arf6-GFP, n=15; ARF1/6PAML-GFP, n=33; ARF1/6IQ-GFP, n=19; ARF1/6PAML+1-GFP, n=18; ARF1/6IQ+1-GFP, n=22; ARF1/6PAML-3KQ-GFP, n=18; ARF1/6IQ-4KQ-GFP, n=18. The membrane/cytosol ratio was calculated by dividing the membrane fluorescence value by the average cytosolic fluorescence value, as described in Methods. Data are presented as mean values +/- standard deviation, with each individual point shown. Statistical analysis (unpaired t test with Welch's correction, two-tailed p value) was performed. ** indicates $p < 0.01$; ****, $p < 0.0001$; ns, not significant. p = for the comparison of GFP alone and Arf1-GFP. Numerical values obtained for all calculated ratios are in Supplementary Table 4, and p values for all pairwise comparisons are in Supplementary Table 5.

Figure 6: Localization of Arf1/6 predicted ancestral proteins in yeast cells. Plasmids encoding Arf1-GFP (A), Arf1/6PAML-GFP (B) and Arf1/6IQ-GFP (C) were co-transformed into *Saccharomyces cerevisiae* cells with a plasmid encoding Arf1-mCherry. Representative images for each GFP-tagged protein expressed (left panels), with Arf1-mCherry (middle panels) are shown. The yellow lines are those

used to generate plots of pixel intensities (right panels) for the GFP-tagged proteins (green) and Arf1-mCherry (magenta). At least three independent transformations were performed for each mutant plasmid, always including the WT control plasmid. Scale bar, 5 µm.

Figure 7: Localization of Arf1/6 predicted ancestral proteins to the PM in yeast cells. Plasmids encoding GFP (A), Arf1-GFP (B), Arf1/6PAML-GFP (C), Arf1/6IQ-GFP (D), Arf1/6PAML-3KQ-GFP (E), and Arf1/6IQ-4KQ-GFP (F) were transformed into *Saccharomyces cerevisiae* cells. Localizations of each GFP-tagged protein were quantified as described in Methods. Representative images for each protein expressed in yeast are shown (left panels). The yellow lines are those used to generate plots of pixel intensities (right panels). “0” (left panels) indicates the beginning of the line shown, and corresponds to “0” on the x-axis of the plots (left panels). At least three independent transformations were performed for each mutant plasmid, always including the WT control plasmid. Scale bar, 5 µm.

Figure 8: C-terminal region of selected Arf1 homologues from diverse eukaryotes. This alignment illustrates the presence of multiple positively charged residues in the C-terminus of Arf1 proteins from phylogenetically diverse eukaryotes, including animals, fungi, and organisms with agricultural (*P. brassicae*, *T. foetus*), ecological (*T. pseudonana*, *P. papillatum*), and medical (*G. intestinalis*, *C. neoformans*, *T. brucei*) impact. The sequence abbreviation used in Supplementary Table 1 is given before the organism name. Positively charged residues are coloured lime.

Supplementary Figure Legends

Supplementary Figure 1. Phylogenetic analyses of Arf sub-families to identify least divergent paralogs. (A) Arf1. Model=Q.Insect+R8 Matrix=451 taxa, 171 positions. (B) Arf6 Model=Q.Insect+I+G4 Matrix=68, 170 positions. (C) Arl1. Model=Q.Yeast.R5 Matrix=126 taxa, 163 positions (D) Arl5. Model=Q.Yeast.R6. Matrix=92 taxa, 161 positions.

Supplementary Figure 2. Constrained tree topology for ASR analyses. This tree was manually constrained based on the topology from Vargová et al. 2021 [11], assuming that Arl1/5 represents an outgroup root. Internal eukaryotic relationships within each clade are based on consensus eukaryotic relationships [23,24] and with a polytomy at the root, reflecting the knowledge at the time of analysis. Notably, the Carpediemonas ‘Arf6’ was included in the calculation yielding the sequences tested. However, recalculation without the sequence resulted in only a single AA residue change at a low probability site, and thus did not substantially change the inferred Arf1/6 ancestral sequence (Supplementary Table 2).

Supplementary Figure 3. Expression levels of Arf1/6 predicted ancestral proteins and controls. MDCK cells were transfected with the plasmids encoding the indicated proteins fused to GFP (or GFP alone) (A), and samples prepared for GFP in gel fluorescence (see Methods) (A). As a loading control, total proteins were visualized in commercial pre-cast SDS–PAGE stain-free gels. (B). Uncropped version of blot in Part A is shown (C).

Supplementary Figure 4. Localization of alternative Arf1/6 predicted ancestral proteins to the Golgi and PM in MDCK cells. MDCK cells were transfected with ARF1/6PAML+1-GFP (A,B) or ARF1/6IQ+1-GFP (C,D) and prepared for immunofluorescence analysis using Hoechst dye to visualize nuclei (blue). Cells were either co-transfected with the 8+mCherry PM probe (magenta) (A,C) or prepared for immunofluorescence analysis using antibodies against GM130, an early Golgi marker (magenta) (B,D). Representative merged images are shown on the left for each transformant, with GFP-tagged protein (green), nuclei (blue), and either PM (A,C) or Golgi (B,D) in magenta. Insets (yellow square) are shown on the right, with each channel shown separately (left inset panels). The yellow lines shown in the merged inset panels (upper right) are those used to generate plots of pixel intensities for the GFP-tagged proteins (green) and the organelle marker (magenta) (lower right) of each part (A-D). At least three independent transformations were performed for each mutant plasmid, always including the appropriate control plasmids. Scale bar, 5 µm.

Supplementary Figure 5. Localization of alternative Arf1/6 predicted ancestral proteins in yeast cells. Plasmids encoding Arf1/6PAML+1-GFP (A) and Arf1/6IQ+1-GFP (B) were co-transformed into *Saccharomyces cerevisiae* cells with a plasmid encoding Arf1-mCherry. Representative images for each GFP-tagged protein expressed (left panels), with Arf1-mCherry (middle panels) are shown. The merged images are shown in the right panels. At least three independent transformations were performed for each mutant plasmid, always including the WT control plasmid. Scale bar, 5 µm.

References

1. Donoghue, P. C. J. *et al.* Defining eukaryotes to dissect eukaryogenesis. *Curr. Biol.* **33**, R919–R929 (2023).
2. Vosseberg, J. *et al.* The emerging view on the origin and early evolution of eukaryotic cells. *Nature* **633**, 295–305 (2024).
3. López-García, P. & Moreira, D. The symbiotic origin of the eukaryotic cell. *C. R. Biol.* **346**, 55–73 (2023).
4. Muñoz-Gómez, S. A. *et al.* Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* **6**, 253–262 (2022).
5. Richards, T. A. *et al.* Reconstructing the last common ancestor of all eukaryotes. *PLoS Biol.* **22**, e3002917 (2024).
6. Eme, L. *et al.* Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 992–999 (2023).
7. Vargová, R. *et al.* The Asgard archaeal origins of Arf family GTPases involved in eukaryotic organelle dynamics. *Nat. Microbiol.* **10**, 495–508 (2025).
8. Tran, L. T., Akıl, C., Senju, Y. & Robinson, R. C. The eukaryotic-like characteristics of small GTPase, roadblock and TRAPPC3 proteins from Asgard archaea. *Commun. Biol.* **7**, 273 (2024).
9. Prokopchuk, G. *et al.* Lessons from the deep: mechanisms behind diversification of eukaryotic protein complexes. *Biol. Rev. Camb. Philos. Soc.* **98**, 1910–1927 (2023).
10. Li, Y. *et al.* Functional genomic analysis of the ADP-ribosylation factor family of GTPases: phylogeny among diverse eukaryotes and function in C. elegans. *FASEB J.* **18**, 1834–1850 (2004).
11. Vargová, R. *et al.* A Eukaryote-Wide Perspective on the Diversity and Evolution of the ARF GTPase Protein Family. *Genome Biol. Evol.* **13**, evab157 (2021).
12. Donaldson, J. G. & Jackson, C. L. ARF family G proteins and their regulators: roles in membrane transport, development and disease. *Nat. Rev. Mol. Cell Biol.* **12**, 362–375 (2011).
13. Sztul, E. *et al.* Arf GTPases and their GEFs and GAPS: Concepts and challenges. *Mol. Biol. Cell* **30(11):124**, 1249–1271 (2019).

14. Li, F.-L. & Guan, K.-L. The Arf family GTPases: Regulation of vesicle biogenesis and beyond. *Bioessays* **45**, e2200214 (2023).

15. Jackson, C. L., Ménétrey, J., Sivia, M., Dacks, J. B. & Eliáš, M. An evolutionary perspective on Arf family GTPases. *Curr. Opin. Cell Biol.* **85**, 102268 (2023).

16. Zhu, J. *et al.* Asgard Arf GTPases can act as membrane-associating molecular switches with the potential to function in organelle biogenesis. *Nat. Commun.* **16**, 2622 (2025).

17. Mayor, S., Parton, R. G. & Donaldson, J. G. Clathrin-independent pathways of endocytosis. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).

18. Dejgaard, S. Y. & Presley, J. F. Arfs on the Golgi: four conductors, one orchestra. *Front. Mol. Biosci.* **Volume 12**, (2025).

19. Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* **46**, 247–269 (2017).

20. Mascotti, M. L. Resurrecting Enzymes by Ancestral Sequence Reconstruction. *Methods Mol. Biol.* **2397**, 111–136 (2022).

21. Mendes, L. F. S. *et al.* Resurrecting Golgi proteins to grasp Golgi ribbon formation and self-association under stress. *Int. J. Biol. Macromol.* **194**, 264–275 (2022).

22. Boverio, A. *et al.* Structure, mechanism, and evolution of the last step in vitamin C biosynthesis. *Nat. Commun.* **15**, 4158 (2024).

23. Adl, S. M. *et al.* Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119 (2019).

24. Tikhonenkov, D. V *et al.* Microbial predators form a new supergroup of eukaryotes. *Nature* **612**, 714–719 (2022).

25. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).

26. Honda, A., Al-Awar, O. S., Hay, J. C. & Donaldson, J. G. Targeting of Arf-1 to the early Golgi by membrin, an ER-Golgi SNARE. *J. Cell Biol.* **168**, 1039–1051 (2005).

27. Matheson, L. A., Suri, S. S., Hanton, S. L., Chatre, L. & Brandizzi, F. Correct targeting of plant ARF GTPases relies on distinct protein domains. *Traffic* **9**, 103–120 (2008).

28. Marais, C. *et al.* The Qb-SNARE Memb11 interacts specifically with Arf1 in the Golgi apparatus of Arabidopsis thaliana. *J. Exp. Bot.* **66**, 6665–6678 (2015).

29. Yeung, T. *et al.* Membrane Phosphatidylserine Regulates Surface Charge and Protein Localization. *Science* **319**, 210–213 (2008).

30. Hammond, G. R. V *et al.* PI4P and PI(4,5)P2 Are Essential But Independent Lipid Determinants of Membrane Identity. *Science* **337**, 727–730 (2012).

31. Heo, W. Do *et al.* PI(3,4,5)P_3 and PI(4,5)P_2 Lipids Target Proteins with Polybasic Clusters to the Plasma Membrane. *Science* **314**, 1458–1461 (2006).

32. Wong-Dilworth, L. *et al.* STED imaging of endogenously tagged ARF GTPases reveals their distinct nanoscale localizations. *J. Cell Biol.* **222**, (2023).

33. Gillingham, A. K. & Munro, S. The small G proteins of the Arf family and their regulators. *Annu. Rev. Cell Dev. Biol.* **23**, 579–611 (2007).

34. Takatsu, H., Yoshino, K., Toda, K. & Nakayama, K. GGA proteins associate with Golgi membranes through interaction between their GGAH domains and ADP-ribosylation factors. *Biochem. J.* **365**, 369–378 (2002).

35. Dubois, T. *et al.* Golgi-localized GAP for Cdc42 functions downstream of ARF1 to control Arp2/3 complex and F-actin dynamics. *Nat. Cell Biol.* **7**, 353–364 (2005).

36. Cohen, L. A. *et al.* Active Arf6 recruits ARNO/cytohesin GEFs to the PM by binding their PH domains. *Mol. Biol. Cell* **18**, 2244–2253 (2007).

37. Stalder, D. *et al.* Kinetic studies of the Arf activator Arno on model membranes in the presence of Arf effectors suggest control by a positive feedback loop. *J. Biol. Chem.* **286**, 3873–3883 (2011).

38. Malaby, A. W., van den Berg, B. & Lambright, D. G. Structural basis for membrane recruitment and allosteric activation of cytohesin family Arf GTPase exchange factors. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14213–14218 (2013).

39. Chavrier, P. & Ménétrey, J. Toward a structural understanding of arf family:effector specificity. *Structure* **18**, 1552–1558 (2010).

40. Williamson, K. *et al.* A robustly rooted tree of eukaryotes reveals their excavate ancestry. *Nature* **640**, 974–981 (2025).

41. Macia, E. *et al.* The GDP-bound form of Arf6 is located at the plasma membrane. *J. Cell Sci.* **117**, 2389–2398 (2004).

42. Martinu, L. *et al.* The TBC (Tre-2/Bub2/Cdc16) domain protein TRE17 regulates plasma membrane-endosomal trafficking through activation of Arf6. *Mol. Cell. Biol.* **24**, 9752–9762 (2004).

43. Falace, A. *et al.* TBC1D24, an ARF6-interacting protein, is mutated in familial infantile myoclonic epilepsy. *Am. J. Hum. Genet.* **87**, 365–370 (2010).

44. Gommel, D. U. *et al.* Recruitment to Golgi membranes of ADP-ribosylation factor 1 is mediated by the cytoplasmic domain of p23. *EMBO J.* **20**, 6751–6760 (2001).

45. Sabido-Bozo, S. *et al.* The p24 Complex Contributes to Specify Arf1 for COPI Coat Selection. *Int. J. Mol. Sci.* **22**, (2021).

46. Min, M. K. *et al.* Recruitment of Arf1-GDP to Golgi by Glo3p-type ArfGAPs is crucial for golgi maintenance and plant growth. *Plant Physiol.* **161**, 676–691 (2013).

47. Hofmann, I., Thompson, A., Sanderson, C. M. & Munro, S. The Arl4 family of small G proteins can recruit the cytohesin Arf6 exchange factors to the plasma membrane. *Curr. Biol.* **17**, 711–716 (2007).

48. Li, C.-C. *et al.* ARL4D recruits cytohesin-2/ARNO to modulate actin remodeling. *Mol. Biol. Cell* **18**, 4420–4437 (2007).

49. Lee, D. M., Rodrigues, F. F., Yu, C. G., Swan, M. & Harris, T. J. C. PH Domain-Arf G Protein Interactions Localize the Arf-GEF Steppke for Cleavage Furrow Regulation in Drosophila. *PLoS One* **10**, e0142562 (2015).

50. Mazaki, Y. *et al.* ARF1 recruits RAC1 to leading edge in neutrophil chemotaxis. *Cell Commun. Signal.* **15**, 36 (2017).

51. Lewis-Saravalli, S., Campbell, S. & Claing, A. ARF1 controls Rac1 signaling to regulate migration of MDA-MB-231 invasive breast cancer cells. *Cell. Signal.* **25**, 1813–1819 (2013).

52. Braun, V. *et al.* AP-1 and ARF1 control endosomal dynamics at sites of FcR mediated phagocytosis. *Mol. Biol. Cell* **18**, 4921–4931 (2007).

53. Boulay, P.-L., Cotton, M., Melançon, P. & Claing, A. ADP-ribosylation factor 1 controls the activation of the phosphatidylinositol 3-kinase pathway to regulate epidermal growth factor-dependent growth and migration of breast cancer cells. *J. Biol. Chem.* **283**, 36425–36434 (2008).

54. Haines, E., Schlienger, S. & Claing, A. The small GTPase ADP-Ribosylation

Factor 1 mediates the sensitivity of triple negative breast cancer cells to EGFR tyrosine kinase inhibitors. *Cancer Biol. Ther.* **16**, 1535–1547 (2015).

55. Saila, S., Gyanwali, G. C., Hussain, M., Gianfelice, A. & Ireton, K. The Host GTPase Arf1 and Its Effectors AP1 and PICK1 Stimulate Actin Polymerization and Exocytosis To Promote Entry of Listeria monocytogenes. *Infect. Immun.* **88**, (2020).
56. Kumari, S., Mg, S. & Mayor, S. Endocytosis unplugged: multiple ways to enter the cell. *Cell Res.* **20**, 256–275 (2010).
57. Barlow, L. D. *et al.* Comparative Genomics for Evolutionary Cell Biology Using AMOEBAE: Understanding the Golgi and Beyond. *Methods Mol. Biol.* **2557**, 431–452 (2023).
58. Amor, J. C., Harrison, D. H., Kahn, R. A. & Ringe, D. Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* **372**, 704–708 (1994).
59. Ménétrey, J., Macia, E., Pasqualato, S., Franco, M. & Cherfils, J. Structure of Arf6-GDP suggests a basis for guanine nucleotide exchange factors specificity. *Nat. Struct. Biol.* **7**, 466–469 (2000).
60. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* **24**, 2101–2102 (2008).
61. Sanial, M. *et al.* Direct observation of fluorescent proteins in gels: A rapid, cost-efficient, and quantitative alternative to immunoblotting. *Biol. cell* **117**, e2400161 (2025).
62. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).

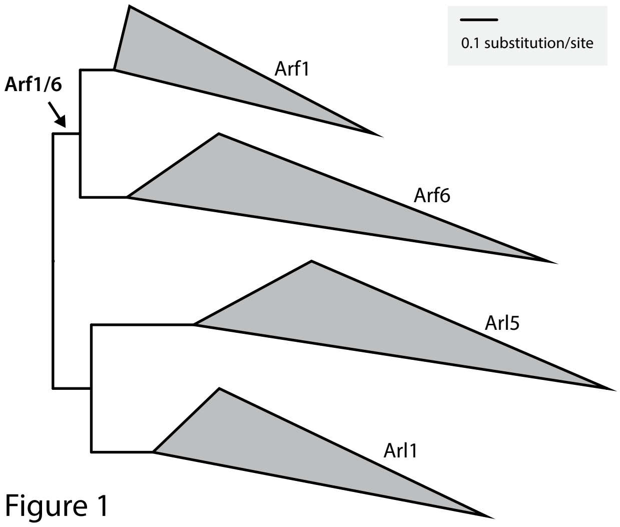
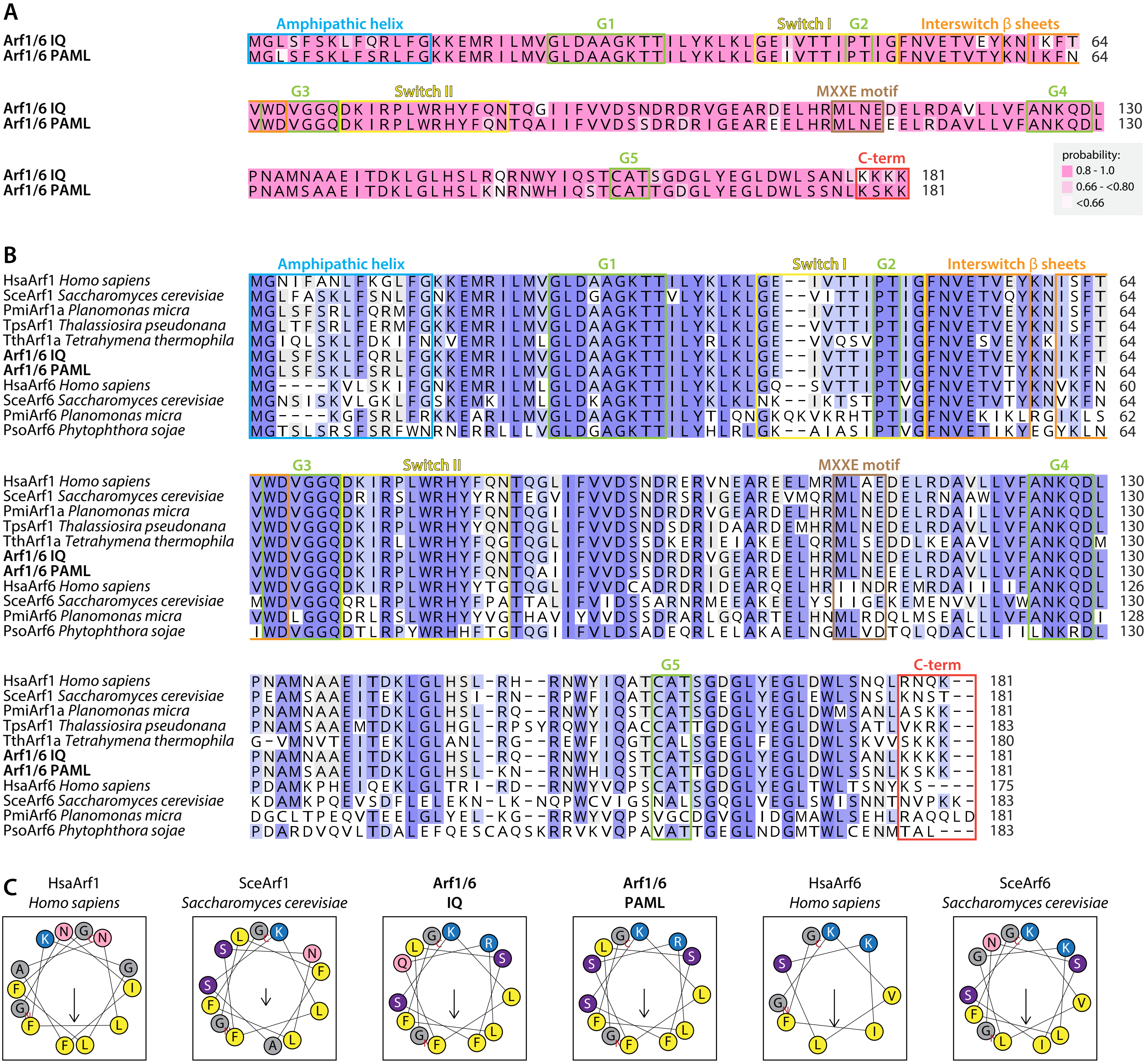

0.1 substitution/site
Arf1/6
Arf1
Arf6
Arl5
Arl1

Figure 1



A

```
Arf1/6 IQ    MGLSFSKLFQRLFGKKEMRILMVGLDAAGKTTILYKLKLGEIVTTIPTIGFNVETVEYKNIKFT  64
Arf1/6 PAML  MGLSFSKLFSRLFGKKEMRILMVGLDAAGKTTILYKLKLGEIVTTIPTIGFNVETVTYKNIKFN  64

Arf1/6 IQ    VWDVGGQDKIRPLWRHYFQNTQGIIFVVDSNDRDRVGEARDELHRMLNEDELRDAVLLVFANKQDL  130
Arf1/6 PAML  VWDVGGQDKIRPLWRHYFQNTQAIIFVVDSSDRDRIGEAREELHRMLNEEELRDAVLLVFANKQDL  130

Arf1/6 IQ    PNAMNAAEITDKLGLHSLRQRNWYIQSTCATSGDGLYEGLDWLSANLKKKK  181
Arf1/6 PAML  PNAMSAAEITDKLGLHSLKNRNWHIQSTCATTGDGLYEGLDWLSSNLKSKK  181
```

B

```
HsaArf1 Homo sapiens                  MGNIFANLFKGLFGKKEMRILMVGLDAAGKTTILYKLKLGE--IVTTIPTIGFNVETVEYKNISFT  64
SceArf1 Saccharomyces cerevisiae     MGLFASKLFSNLFGNKEMRILMVGLDGAGKTTVLYKLKLGE--VITTIPTIGFNVETVQYKNISFT  64
PmiArf1a Planomonas micra            MGLSFSRLFQRMFGKKEMRILMVGLDAAGKTTILYKLKLGE--IVTTIPTIGFNVETVEYKNISFT  64
TpsArf1 Thalassiosira pseudonana     MGLTFSRLFERMFGKKEMRILMVGLDAAGKTTILYKLKLGE--VVTTIPTIGFNVETVEYKNISFT  64
TthArf1a Tetrahymena thermophila     MGIQLSKLFDKIFNKVEMRILMLGLDAAGKTTILYRLKLGE--VVQSVPTIGFNVESVEYKKIKFT  64
Arf1/6 IQ                           MGLSFSKLFQRLFGKKEMRILMVGLDAAGKTTILYKLKLGE--IVTTIPTIGFNVETVEYKNIKFT  64
Arf1/6 PAML                         MGLSFSKLFSRLFGKKEMRILMVGLDAAGKTTILYKLKLGE--IVTTIPTIGFNVETVTYKNIKFN  64
HsaArf6 Homo sapiens                  MG----KVLSKIFGNKEMRILMLGLDAAGKTTILYKLKLGQ--SVTTIPTVGFNVETVTYKNVKFN  60
SceArf6 Saccharomyces cerevisiae     MGNSISKVLGKLFGSKEMKILMLGLDKAGKTTILYKLKLNK--IKTSTPTVGFNVETVTYKNVKFN  64
PmiArf6 Planomonas micra             MG----KGFSRLFRKKEARILMVGLDAAGKTTILYTLQNGKQKVKRHTPTIGFNVEKIKLRGIKLS  62
PsoArf6 Phytophthora sojae           MGTSLSRSFSRFWNRNERRLLLVGLDGAGKTTILYHLRLGK--AIASIPTVGFNVETIKYEGYKLN  64

HsaArf1 Homo sapiens                  VWDVGGQDKIRPLWRHYFQNTQGLIFVVDSNDRERVNEAREELMRMLAEDELRDAVLLVFANKQDL  130
SceArf1 Saccharomyces cerevisiae     VWDVGGQDRIRSLWRHYYRNTEGVIFVVDSNDRSRIGEAREVMQRMLNEDELRNAAWLVFANKQDL  130
PmiArf1a Planomonas micra            VWDVGGQDKIRPLWRHYFQNTQGIIFVVDSNDRDRVGEARDELHRMLNEDELRDAILLVFANKQDL  130
TpsArf1 Thalassiosira pseudonana     VWDVGGQDKIRPLWRHYYQNTQGLIFVVDSNDSDRIDAARDEMHRMLNEDELRDAVLLVFANKQDL  130
TthArf1a Tetrahymena thermophila     VWDVGGQDKIRLLWRHYFQGTQGLIFVVDSSDKERIEIAKEELQRMLSEDDLKEAAVLVFANKQDM  130
Arf1/6 IQ                           VWDVGGQDKIRPLWRHYFQNTQGIIFVVDSNDRDRVGEARDELHRMLNEDELRDAVLLVFANKQDL  130
Arf1/6 PAML                         VWDVGGQDKIRPLWRHYFQNTQAIIFVVDSSDRDRIGEAREELHRMLNEEELRDAVLLVFANKQDL  130
HsaArf6 Homo sapiens                  VWDVGGQDKIRPLWRHYYTGTQGLIFVVDCADRDRIDEARQELHRIINDREMRDAIILIFANKQDL  126
SceArf6 Saccharomyces cerevisiae     MWDVGGQQRLRPLWRHYFPATTALIFVIDSSARNRMEEAKEELYSIIGEKEMENVVLLVWANKQDL  130
PmiArf6 Planomonas micra             VWDLGGQDRLRSLWRHYYVGTHAVIYVVDSSDRARLGQARTELHNMLRDQLMSEALLLVFANKQDI  128
PsoArf6 Phytophthora sojae           IWDVGGQDTLRPYWRHHFTGTQGIIFVLDSADEQRLELAKAELNGMLVDTQLQDACLLIILNKRDL  130

HsaArf1 Homo sapiens                  PNAMNAAEITDKLGLHSL-RH--RNWYIQATCATSGDGLYEGLDWLSNQLRNQK--  181
SceArf1 Saccharomyces cerevisiae     PEAMSAAEITEKLGLHSI-RN--RPWFIQATCATSGEGLYEGLEWLSNSLKNST--  181
PmiArf1a Planomonas micra            PNAMNAAEITDKLGLHSL-RQ--RNWYIQSTCATSGDGLYEGLDWMSANLASKK--  181
TpsArf1 Thalassiosira pseudonana     PNAMSAAEMTDKLGLHGL-RPSYRQWYIQACCATTGDGLYEGLDWLSATLVKRK--  183
TthArf1a Tetrahymena thermophila     G-VMNVTEITEKLGLANL-RG--REWFIQGTCALSGEGLFEGLDWLSKVVSKKK--  180
Arf1/6 IQ                           PNAMNAAEITDKLGLHSL-RQ--RNWYIQSTCATSGDGLYEGLDWLSANLKKKK--  181
Arf1/6 PAML                         PNAMSAAEITDKLGLHSL-KN--RNWHIQSTCATTGDGLYEGLDWLSSNLKSKK--  181
HsaArf6 Homo sapiens                  PDAMKPHEIQEKLGLTRI-RD--RNWYVQPSCATSGDGLYEGLTWLTSNYKS----  175
SceArf6 Saccharomyces cerevisiae     KDAMKPQEVSDFLELEKN-LK-NQPWCVIGSNALSGQGLVEGLSWISNNTNVPKK-  183
PmiArf6 Planomonas micra             DGCLTPEQVTEELGLYEL-KG--RRWYVQPSVGCDGVGLIDGMAWLSEHLRAQQLD  181
PsoArf6 Phytophthora sojae           PDARDVQVLTDALEFQESCAQSKRRVKVQPAVATTGEGLNDGMTWLCENMTAL---  183
```

Figure 2

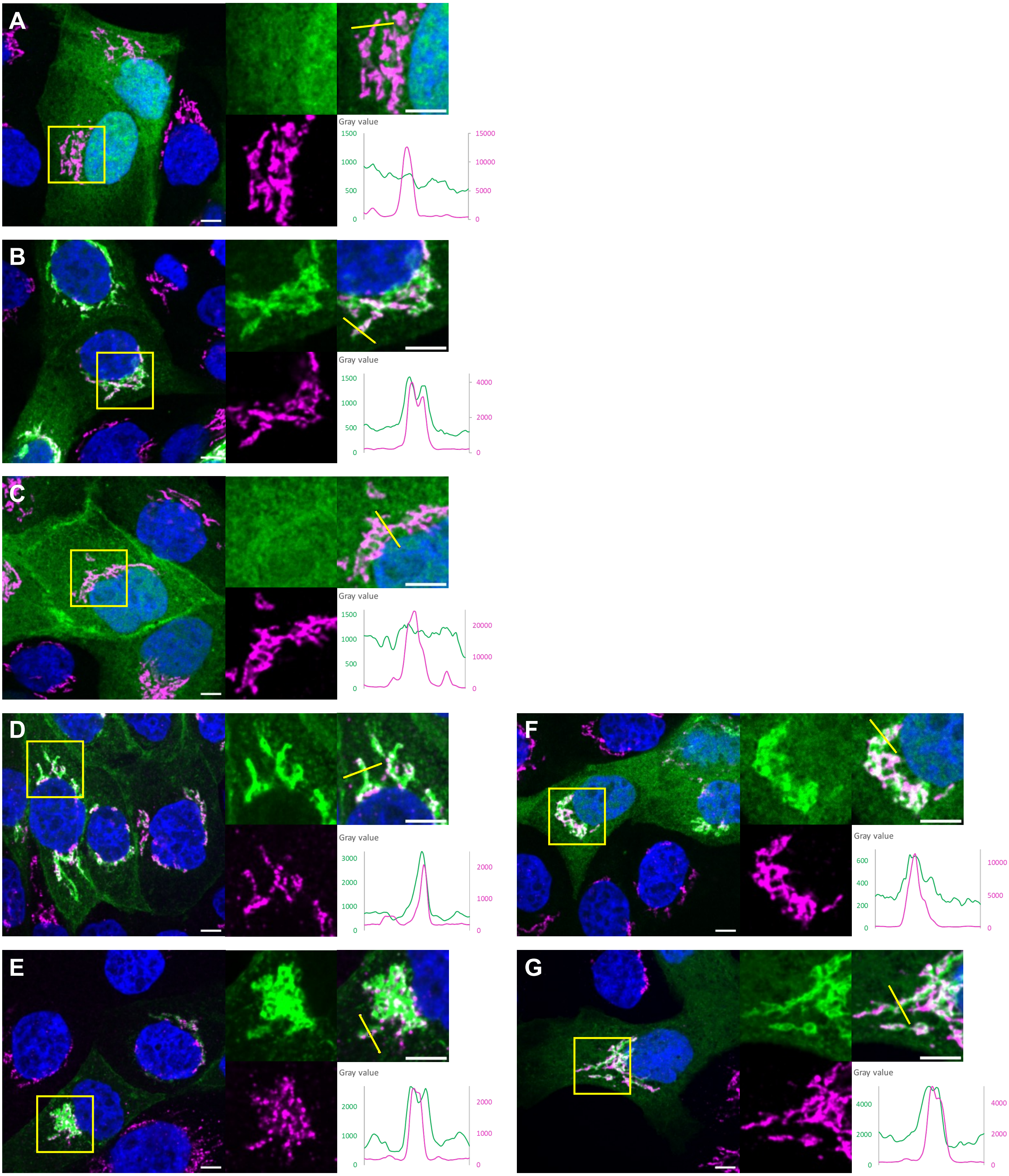

A
Gray value
1500
1000
500
0
15000
10000
5000
0
B
Gray value
1500
1000
500
0
4000
2000
0
C
Gray value
1500
1000
500
0
20000
10000
0
D
Gray value
3000
2000
1000
0
2000
1000
0
E
Gray value
2000
1000
0
2000
1000
0
F
Gray value
600
400
200
0
10000
5000
0
G
Gray value
4000
2000
0
4000
2000
0

Figure 3

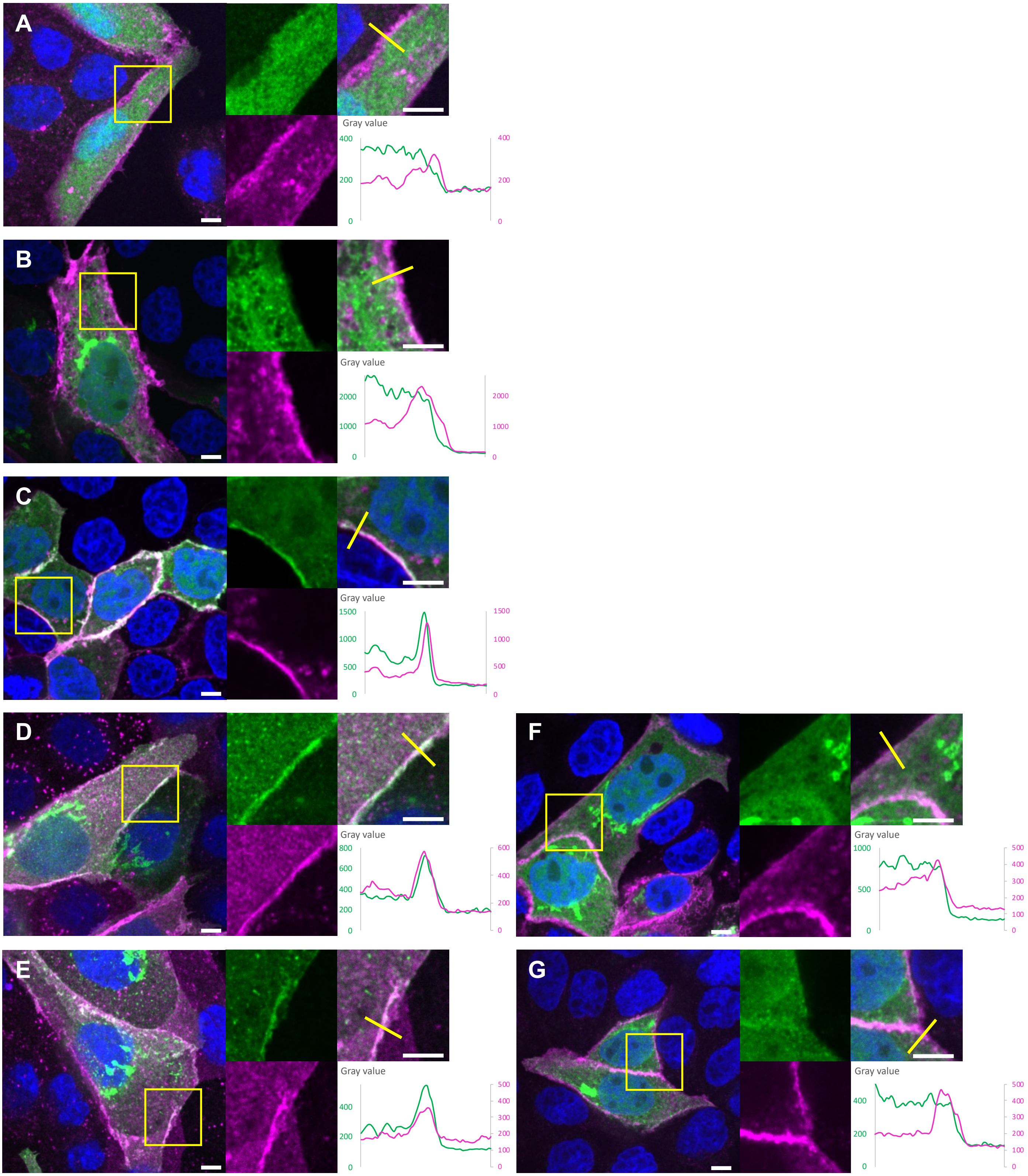
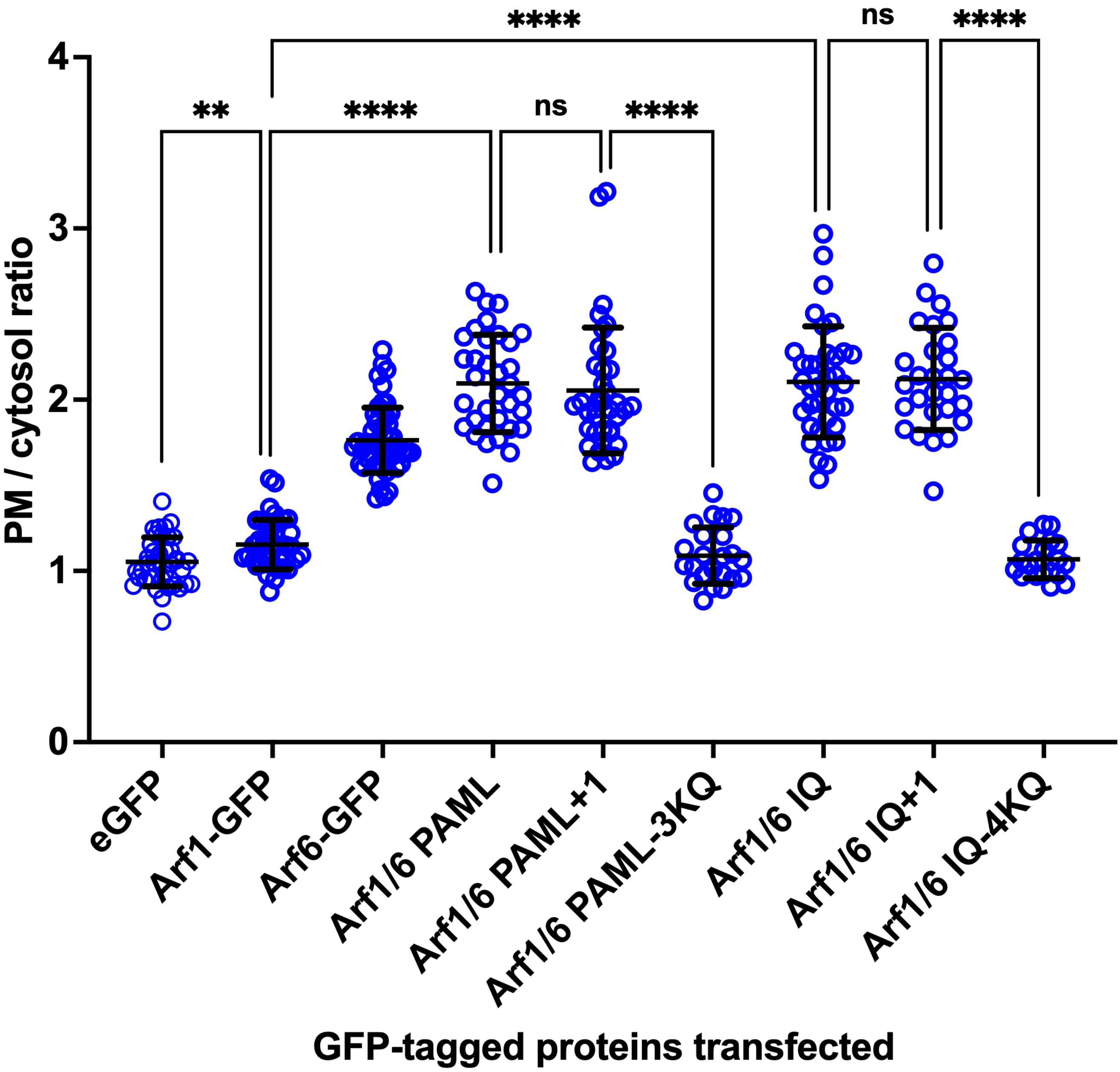

A
Gray value
400
200
0
400
200
0
B
Gray value
2000
1000
0
2000
1000
0
C
Gray value
1500
1000
500
0
1500
1000
500
0
D
Gray value
800
600
400
200
0
600
400
200
0
E
Gray value
400
200
0
500
400
300
200
100
0
F
Gray value
1000
500
0
500
400
300
200
100
0
G
Gray value
400
200
0
500
400
300
200
100
0

Figure 4



4
3
2
1
0
PM / cytosol ratio
**
****
ns
****
****
ns
****
eGFP
Arf1-GFP
Arf6-GFP
Arf1/6 PAML
Arf1/6 PAML+1
Arf1/6 PAML-3KQ
Arf1/6 IQ
Arf1/6 IQ+1
Arf1/6 IQ-4KQ
GFP-tagged proteins transfected

Figure 5

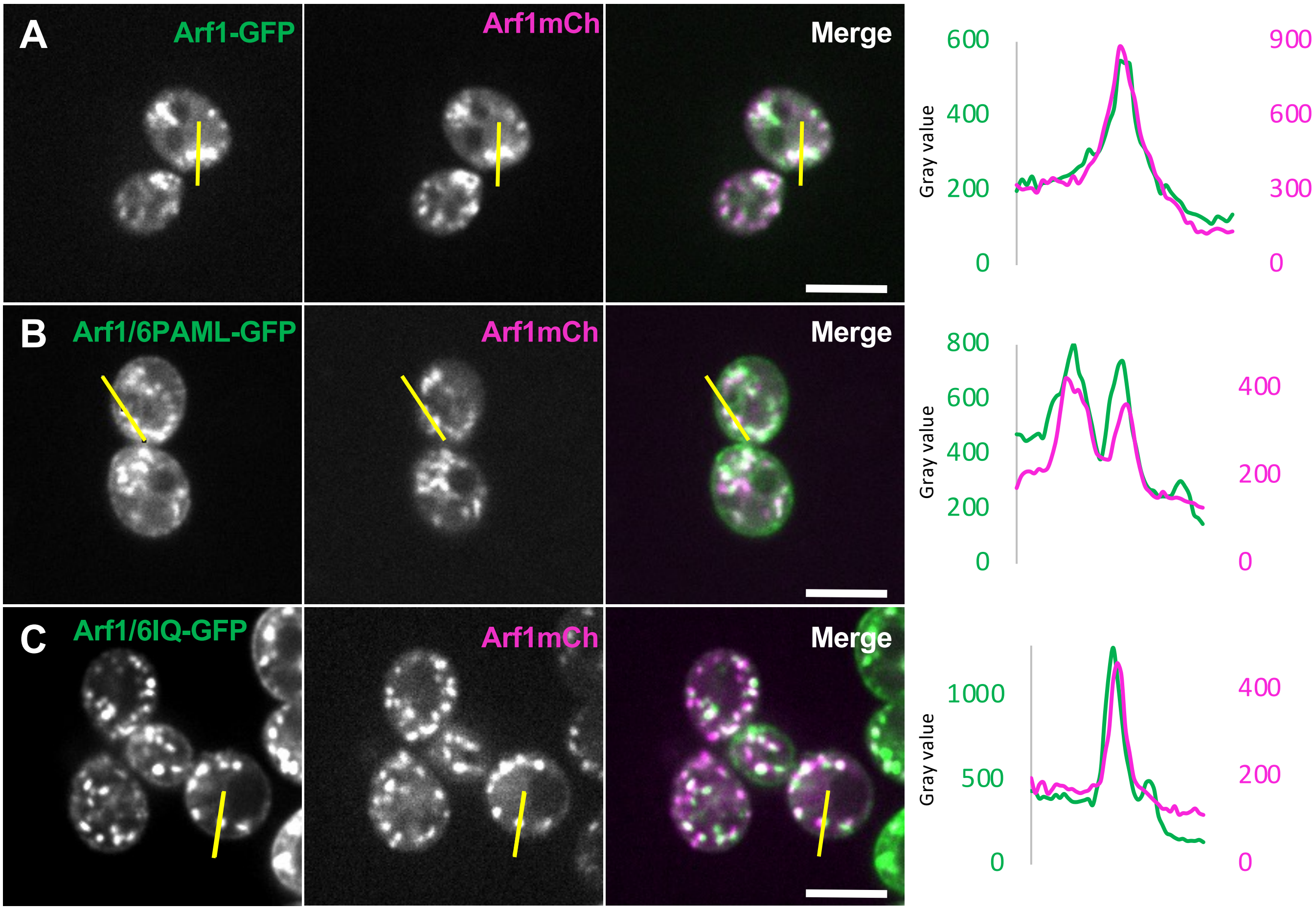

A
Arf1-GFP
Arf1mCh
Merge
600
400
200
0
900
600
300
0
Gray value
B
Arf1/6PAML-GFP
Arf1mCh
Merge
800
600
400
200
0
400
200
0
Gray value
C
Arf1/6IQ-GFP
Arf1mCh
Merge
1000
500
0
400
200
0
Gray value

Figure 6

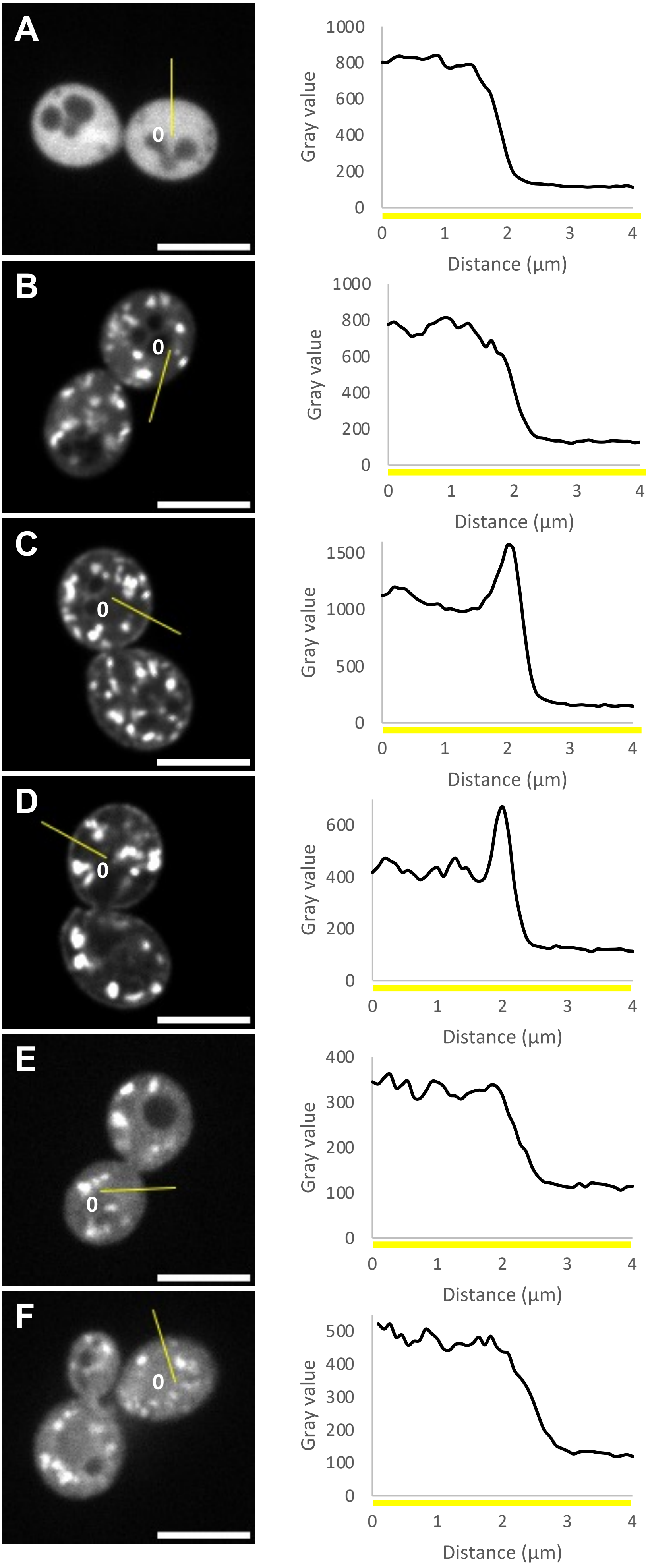
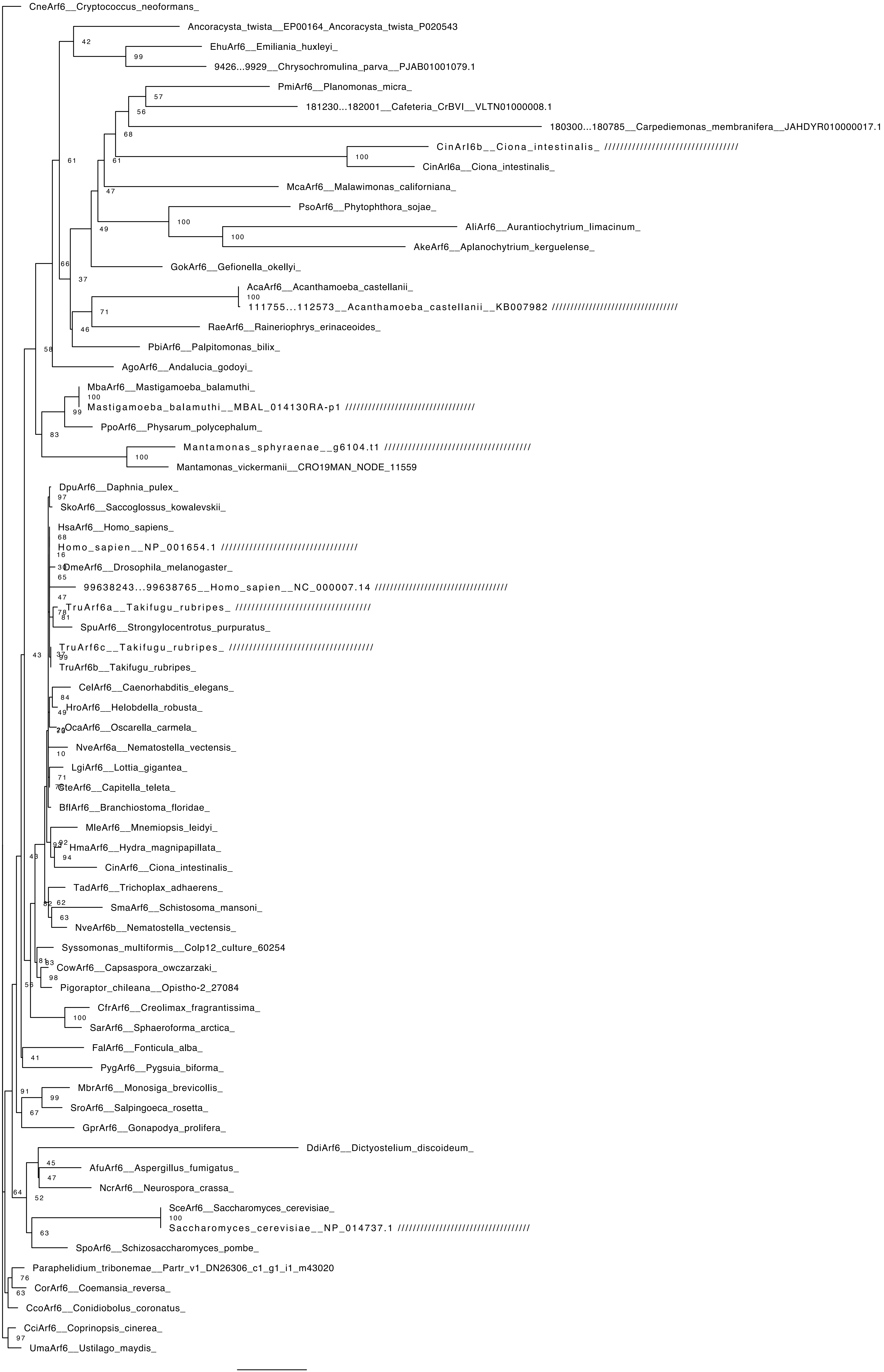

A
0
Gray value
1000
800
600
400
200
0
Distance (µm)
0
1
2
3
4
B
0
Gray value
1000
800
600
400
200
0
Distance (µm)
0
1
2
3
4
C
0
Gray value
1500
1000
500
0
Distance (µm)
0
1
2
3
4
D
0
Gray value
600
400
200
0
Distance (µm)
0
1
2
3
4
E
0
Gray value
400
300
200
100
0
Distance (µm)
0
1
2
3
4
F
0
Gray value
500
400
300
200
100
0
Distance (µm)
0
1
2
3
4

Figure 7

Supplementary Figure 1A


CneArf6__Cryptococcus_neoformans_
Ancoracysta_twista__EP00164_Ancoracysta_twista_P020543
42
EhuArf6__Emiliania_huxleyi_
99
9426...9929__Chrysochromulina_parva__PJAB01001079.1
PmiArf6__Planomonas_micra_
57
181230...182001__Cafeteria_CrBVI__VLTN01000008.1
56
180300...180785__Carpediemonas_membranifera__JAHDYR010000017.1
68
CinArl6b__Ciona_intestinalis_ //////////////////////////////////
100
61
61
CinArl6a__Ciona_intestinalis_
McaArf6__Malawimonas_californiana_
47
PsoArf6__Phytophthora_sojae_
100
AliArf6__Aurantiochytrium_limacinum_
49
100
AkeArf6__Aplanochytrium_kerguelense_
66
GokArf6__Gefionella_okellyi_
37
AcaArf6__Acanthamoeba_castellanii_
100
111755...112573__Acanthamoeba_castellanii__KB007982 //////////////////////////////////
71
46
RaeArf6__Raineriophrys_erinaceoides_
PbiArf6__Palpitomonas_bilix_
58
AgoArf6__Andalucia_godoyi_
MbaArf6__Mastigamoeba_balamuthi_
100
99
Mastigamoeba_balamuthi__MBAL_014130RA-p1 //////////////////////////////////
PpoArf6__Physarum_polycephalum_
83
Mantamonas_sphyraenae__g6104.t1 //////////////////////////////////////
100
Mantamonas_vickermanii__CRO19MAN_NODE_11559
DpuArf6__Daphnia_pulex_
97
SkoArf6__Saccoglossus_kowalevskii_
HsaArf6__Homo_sapiens_
68
Homo_sapien__NP_001654.1 //////////////////////////////////
16
30
DmeArf6__Drosophila_melanogaster_
65
99638243...99638765__Homo_sapien__NC_000007.14 //////////////////////////////////
47
TruArf6a__Takifugu_rubripes_ //////////////////////////////////
78
81
SpuArf6__Strongylocentrotus_purpuratus_
TruArf6c__Takifugu_rubripes_ //////////////////////////////////////
37
99
43
TruArf6b__Takifugu_rubripes_
CelArf6__Caenorhabditis_elegans_
84
HroArf6__Helobdella_robusta_
49
OcaArf6__Oscarella_carmela_
NveArf6a__Nematostella_vectensis_
10
LgiArf6__Lottia_gigantea_
71
CteArf6__Capitella_teleta_
BflArf6__Branchiostoma_floridae_
MleArf6__Mnemiopsis_leidyi_
92
93
43
HmaArf6__Hydra_magnipapillata_
94
CinArf6__Ciona_intestinalis_
TadArf6__Trichoplax_adhaerens_
62
82
SmaArf6__Schistosoma_mansoni_
63
NveArf6b__Nematostella_vectensis_
Syssomonas_multiformis__Colp12_culture_60254
81
83
CowArf6__Capsaspora_owczarzaki_
98
56
Pigoraptor_chileana__Opistho-2_27084
CfrArf6__Creolimax_fragrantissima_
100
SarArf6__Sphaeroforma_arctica_
FalArf6__Fonticula_alba_
41
PygArf6__Pygsuia_biforma_
MbrArf6__Monosiga_brevicollis_
91
99
SroArf6__Salpingoeca_rosetta_
67
GprArf6__Gonapodya_prolifera_
DdiArf6__Dictyostelium_discoideum_
45
AfuArf6__Aspergillus_fumigatus_
47
NcrArf6__Neurospora_crassa_
64
52
SceArf6__Saccharomyces_cerevisiae_
100
Saccharomyces_cerevisiae__NP_014737.1 //////////////////////////////////
63
SpoArf6__Schizosaccharomyces_pombe_
Paraphelidium_tribonemae__Partr_v1_DN26306_c1_g1_i1_m43020
76
CorArf6__Coemansia_reversa_
63
CcoArf6__Conidiobolus_coronatus_
CciArf6__Coprinopsis_cinerea_
97
UmaArf6__Ustilago_maydis_
0.4

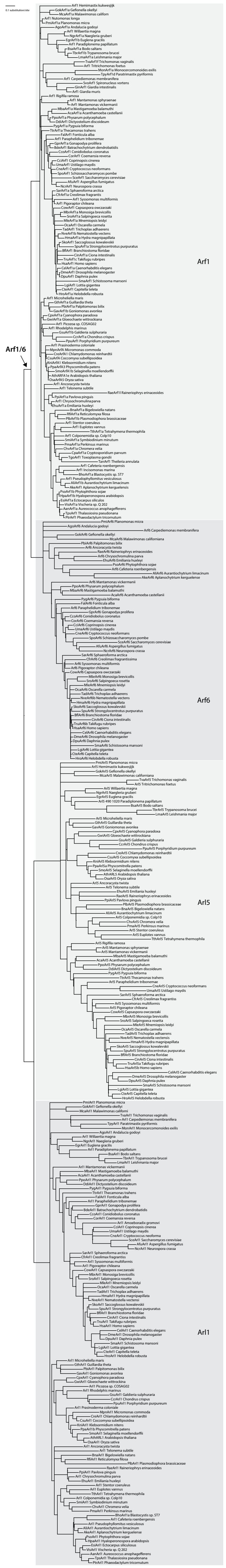
Supplementary Figure 1B

MbrArl1__Monosiga_brevicollis_
82
CcoArl1__Conidiobolus_coronatus_
87
TtrArl1__Thecamonas_trahens_
CorArl1__Coemansia_reversa_
PpoArl1__Physarum_polycephalum_
98
DdiArl1__Dictyostelium_discoideum_
100
Dictyostelium_discoideum__XP_636876.2 //////////////////////////////////
85
MbaArl1__Mastigamoeba_balamuthi_
100
Mastigamoeba_balamuthi__MBAL_008919RA-p1 //////////////////////////////////
95
PygArl1__Pygsuia_biforma_
77
1207...1719__Mantamonas_vickermanii__CRO19MAN_NODE_5561
100
Mantamonas_sphyraenae__g636.t1 //////////////////////////////////
99
Acanthamoeba_castellanii__ACA1_062370-t26_1-p1 //////////////////////////////////
100
AcaArl1__Acanthamoeba_castellanii_
GsuArl1__Galdieria_sulphuraria_
97
CcrArl1__Chondrus_crispus_
81
100
PpuArl1__Porphyridium_purpureum_
89
Bigelowiella_natans__EP00466_Bigelowiella_natans_P010972 //////////////////////////////////
100
BnaArl1__Bigelowiella_natans_
88
75
Rhodelphis_marinus__Rmarinus_m.13250
41
EgrArl1__Euglena_gracilis_
NgrArl1__Naegleria_gruberi_
100
14267...14809__Willaertia_magna__CADWAF010001842.1 //////////////////////////////////
40
Telonema_subtile__EP00656_Telonema_subtile_P010313
50
BsaArl1__Bodo_saltans_
100
TbrArl1__Trypanosoma_brucei_
99
45
LmaArl1__Leishmania_major_
155...691__Paradiplonema_papillatum__GJNJ01035967.1
88
AgoArl1__Andalucia_godoyi_
98
TvaArl1__Trichomonas_vaginalis_
98
MonArl1__Monocercomonoides_exilis_
99
TpyArl1__Paratrimastix_pyriformis_
BhoArl1b__Blastocystis_sp__ST7_ //////////////////////////////////
93
BhoArl1a__Blastocystis_sp__ST7_
75
77
Cafeteria_CrBVI__KAA0153185.1
PtrArl1__Phaeodactylum_tricornutum_
75
74
1160...1687__Incisomonas_marina__INCI3091 //////////////////////////////////
TpsArl1__Thalassiosira_pseudonana_
65
HpaArl1__Hyaloperonospora_arabidopsidis_
89
100
PsoArl1__Phytophthora_sojae_
62
VisArl1__Vischeria_sp__
69
99
Syssomonas_multiformis__Colp12_culture_34527 //////////////////////////////////
99
EsiArl1__Ectocarpus_siliculosus_
59
AanArl1__Aureococcus_anophagefferens_
83
AkeArl1__Aplanochytrium_kerguelense_
74
AliArl1__Aurantiochytrium_limacinum_
44...586__Pseudophyllomitus_vesiculosus__PSEU34746
29
45
Colponemidia_spColp10__EP00742_Colponemidia_sp_Colp-10_P004806
85
SmiArl1__Symbiodinium_minutum_
88
ChvArl1__Chromera_velia_
86
PmaArl1__Perkinsus_marinus_
Amoeboradix_gromovi__NODE_300_length_11422_cov_0.297978.p6
88
PbiArl1__Palpitomonas_bilix_
86
261...794__Euplotes_vannus__Contig34262
85
75
98
Tetrahymena_thermophila__XP_001009425.2 //////////////////////////////////
100
TthArl1__Tetrahymena_thermophila_
97
16370...16903__Stentor_coeruleus__SteCoe_contig_470
100
Stentor_coeruleus__SteCoe_35360 //////////////////////////////////
GokArl1__Gefionella_okellyi_
100
67
McaArl1__Malawimonas_californiana_
53
38
Ancoracysta_twista__EP00164_Ancoracysta_twista_P003818
RfiArl1__Reticulomyxa_filosa_
91
PmiArl1__Planomonas_micra_
PpiArl1__Pavlova_pinguis_
96
100
EhuArl1__Emiliania_huxleyi_
28
34
3112...3736__Chrysochromulina_parva__PJAB01003944.1
GthArl1__Guillardia_theta_
85
Microheliella_maris__TRINITY_DN4591_c0_g1_i1
82
GavArl1__Goniomonas_avonlea_
RaeArl1__Raineriophrys_erinaceoides_
92
PlbArl1__Plasmodiophora_brassicae_
49
CpxArl1__Cyanophora_paradoxa_
100
GwiArl1__Gloeochaete_wittrockiana_
30
Carpediemonas_membranifera__KAG9393728.1
78
CciArl1__Coprinopsis_cinerea_
78
93
CneArl1__Cryptococcus_neoformans_
76
UmaArl1__Ustilago_maydis_
CorArf1__Coemansia_reversa_
100
CneArf1__Cryptococcus_neoformans_
52
55
COSAG01__COSAG01_NODE_7467_length_3198_cov_37.299774_2_plen_181_part_00 //////////////////////////////////
86
100
COSAG04__COSAG04_NODE_747_length_10615_cov_4.568372_10_plen_170_part_00 //////////////////////////////////
COSAG02__COSAG02_NODE_2551_length_8553_cov_102.765673_2_plen_201_part_00
67
MpnArl1__Micromonas_commoda_
SmoArl1__Selaginella_moellendorffii_
83
91
PpaArl1a__Physcomitrella_patens_ //////////////////////////////////
77
78
PpaArl1b__Physcomitrella_patens_
1023957...1024830__Prasinoderma_coloniale__Chr2
AthARL1__Arabidopsis_thaliana_
70
100
OsaArl1__Oryza_sativa_
87
KniArl1__Klebsormidium_nitens_
45
CreArl1__Chlamydomonas_reinhardtii_
99
CsuArl1__Coccomyxa_subellipsoidea_
71
BdeArl1__Batrachochytrium_dendrobatidis_
89
Paraphelidium_tribonemae__Partr_v1_DN23552_c0_g1_i1_m14259
100
Paraphelidium_tribonemae__Partr_v1_DN23552_c0_g1_i2_m14260 //////////////////////////////////
81
AfuArl1__Aspergillus_fumigatus_
100
NcrArl1__Neurospora_crassa_
67
89
100
SceArl1__Saccharomyces_cerevisiae_
100
Saccharomyces_cerevisiae__NP_009723.3 //////////////////////////////////
GprArl1__Gonapodya_prolifera_
FalArl1__Fonticula_alba_
CfrArl1__Creolimax_fragrantissima_
100
SarArl1__Sphaeroforma_arctica_
CowArl1__Capsaspora_owczarzaki_
SmaArl1__Schistosoma_mansoni_
49
DmeArl1__Drosophila_melanogaster_
100
CelArl1__Caenorhabditis_elegans_
46
100
DpuArl1__Daphnia_pulex_
74
LgiArl1__Lottia_gigantea_
98
84
CteArl1__Capitella_teleta_
58
HroArl1__Helobdella_robusta_
CinArl1__Ciona_intestinalis_
84
Homo_sapien__NP_001168.1 //////////////////////////////////
100
96
HsaArl1__Homo_sapiens_
94
TruArl1__Takifugu_rubripes_
96
93
BflArl1__Branchiostoma_floridae_
SkoArl1__Saccoglossus_kowalevskii_
99
SpuArl1__Strongylocentrotus_purpuratus_
97
OcaArl1__Oscarella_carmela_
98
TadArl1__Trichoplax_adhaerens_
92
NveArl1__Nematostella_vectensis_
95
HmaArl1__Hydra_magnipapillata_
MleArl1__Mnemiopsis_leidyi_
62
72
Syssomonas_multiformis__Colp12_culture_32588
Pigoraptor_chileana__Opistho-2_15903
SroArl1__Salpingoeca_rosetta_
0.2

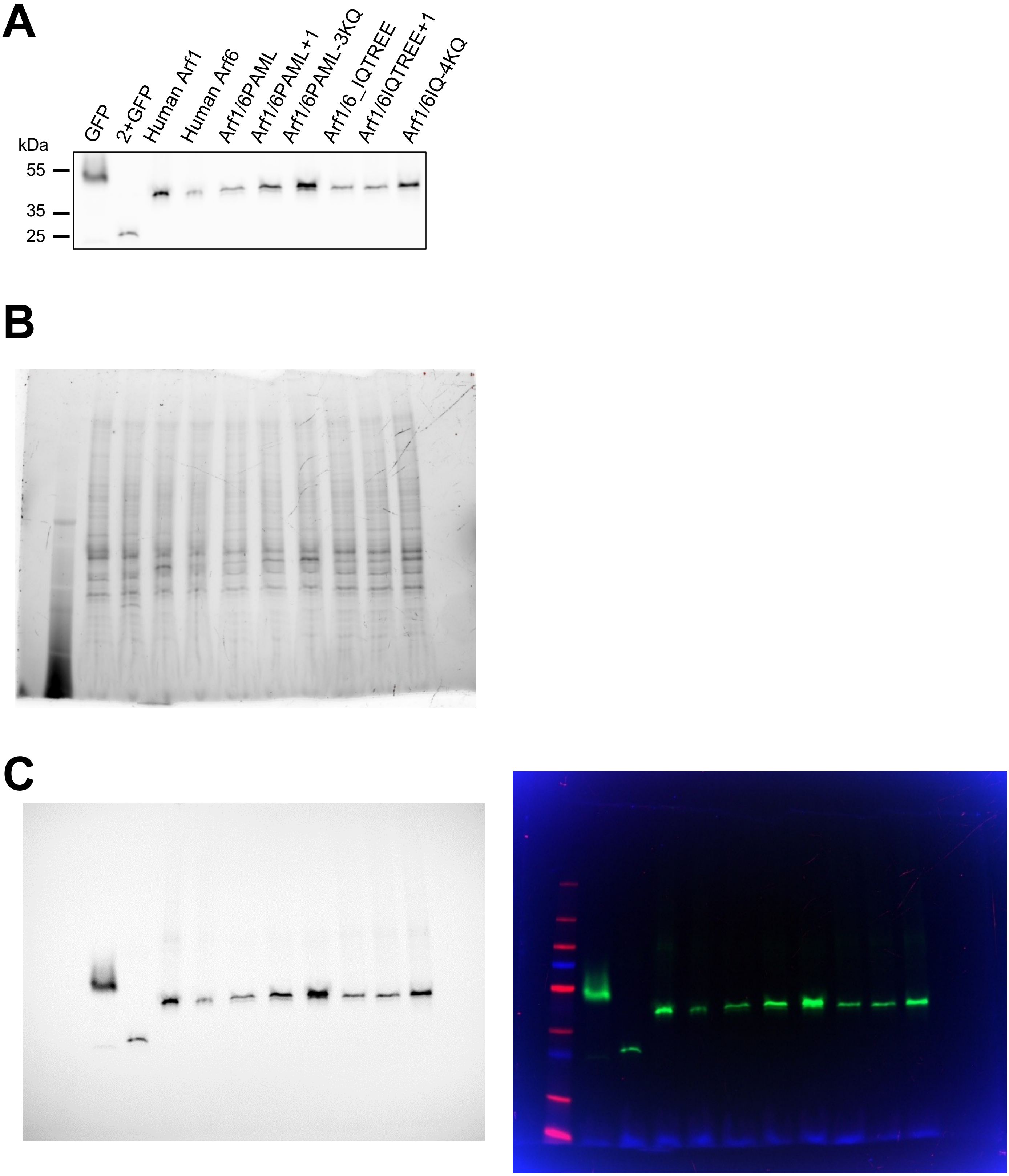
Supplementary Figure 1C

SmoArl5__Selaginella_moellendorffii_
LmaArl5__Leishmania_major_
100
TbrArl5__Trypanosoma_brucei_
100
BsaArl5__Bodo_saltans_
91
490...1020__Paradiplonema_papillatum__GJNJ01046161
90
EgrArl5__Euglena_gracilis_
68
GokArl5__Gefionella_okellyi_
90
McaArl5__Malawimonas_californiana_
PygArl5__Pygsuia_biforma_
SkoArl5__Saccoglossus_kowalevskii_
78
SpuArl5__Strongylocentrotus_purpuratus_
82
47
HmaArl5__Hydra_magnipapillata_
DpuArl5__Daphnia_pulex_
93
63
DmeArl5__Drosophila_melanogaster_
NveArl5__Nematostella_vectensis_
61
70
BflArl5__Branchiostoma_floridae_
86
LgiArl5__Lottia_gigantea_
HsaArl5a__Homo_sapiens_
100
Homo_sapien__NP_036229.1
100
TruArl5a__Takifugu_rubripes_
70
36
Homo_sapien__NP_848930.1
98
HsaArl5b__Homo_sapiens_
100
69
TruArl5b__Takifugu_rubripes_
CinArl5__Ciona_intestinalis_
71
OcaArl5__Oscarella_carmela_
56
80
64
TruArl5c__Takifugu_rubripes_
MleArl5__Mnemiopsis_leidyi_
76
SmaArl5__Schistosoma_mansoni_
62
83
TadArl5__Trichoplax_adhaerens_
CteArl5__Capitella_teleta_
71
82
HroArl5__Helobdella_robusta_
CelArl5__Caenorhabditis_elegans_
54
SroArl5__Salpingoeca_rosetta_
52
71
58
MbrArl5__Monosiga_brevicollis_
69
CowArl5__Capsaspora_owczarzaki_
Syssomonas_multiformis__Colp12_culture_630
83
Pigoraptor_chileana__Opistho-2_40154
Paraphelidium_tribonemae__Partr_v1_DN22402_c0_g1_i
84
CfrArl5__Creolimax_fragrantissima_
69
100
SarArl5__Sphaeroforma_arctica_
CneArl5__Cryptococcus_neoformans_
73
100
UmaArl5__Ustilago_maydis_
68
23
Mantamonas_sphyraenae__g8303.t1
100
Mantamonas_vickermanii__CRO19MAN_NODE_12311
Rigifila_ramosa__EP00004_Rigifila_ramosa_P027130
19
100
Rigifila_ramosa__EP00004_Rigifila_ramosa_P027131
53
TtrArl5__Thecamonas_trahens_
Acanthamoeba_castellanii__ACA1_295340-t26_1-p1
22
100
AcaArl5__Acanthamoeba_castellanii_
PmiArl5__Planomonas_micra_
82
MbaArl5__Mastigamoeba_balamuthi_
100
Mastigamoeba_balamuthi__MBAL_008219RA-p1
76
Dictyostelium_discoideum__XP_645417.1
100
DdiArl5__Dictyostelium_discoideum_
97
PpoArl5__Physarum_polycephalum_
ChvArl5__Chromera_velia_
59
PmaArl5__Perkinsus_marinus_
Stentor_coeruleus__SteCoe_30962
96
100
Stentor_coeruleus__SteCoe_22763
96
42
183...678__Euplotes_vannus__Contig35052
53
34
Colponemidia_spColp10__EP00742_Colponemidia_sp_Col
CpxArl5__Cyanophora_paradoxa_
GavArl5__Goniomonas_avonlea_
NgrArl5__Naegleria_gruberi_
100
2693...3369__Willaertia_magna__CADWAF010001940.1
45
92
79
RaeArl5__Raineriophrys_erinaceoides_
GsuArl5__Galdieria_sulphuraria_
92
CcrArl5__Chondrus_crispus_
38
98
PpuArl5__Porphyridium_purpureum_
Tetrahymena_thermophila__XP_001032077.2
44
100
TthArl5__Tetrahymena_thermophila_
37
BnaArl5__Bigelowiella_natans_
38
100
Bigelowiella_natans__EP00466_Bigelowiella_natans_P
83
96
Tritrichomonas_foetus__OHS97321.1
33
100
TvaArl5__Trichomonas_vaginalis_
PlbArl5__Plasmodiophora_brassicae_
Microheliella_maris__TRINITY_DN10836_c0_g1_i2
GwiArl5__Gloeochaete_wittrockiana_
92
95
Hemimastix_kukwesjijk__EP00696_Hemimastix_kukwesji
80
PpiArl5__Pavlova_pinguis_
99
59
EhuArl5__Emiliania_huxleyi_
Ancoracysta_twista__EP00164_Ancoracysta_twista_P00
100
AliArl5__Aurantiochytrium_limacinum_
60
Telonema_subtile__EP00656_Telonema_subtile_P056830
GthArl5__Guillardia_theta_
98
CsuArl5__Coccomyxa_subellipsoidea_
99
CreArl5__Chlamydomonas_reinhardtii_
99
KniArl5__Klebsormidium_nitens_
PpaArl5a__Physcomitrella_patens_
100
PpaArl5b__Physcomitrella_patens_
AthARL5__Arabidopsis_thaliana_
99
OsaArl5__Oryza_sativa_
0.2

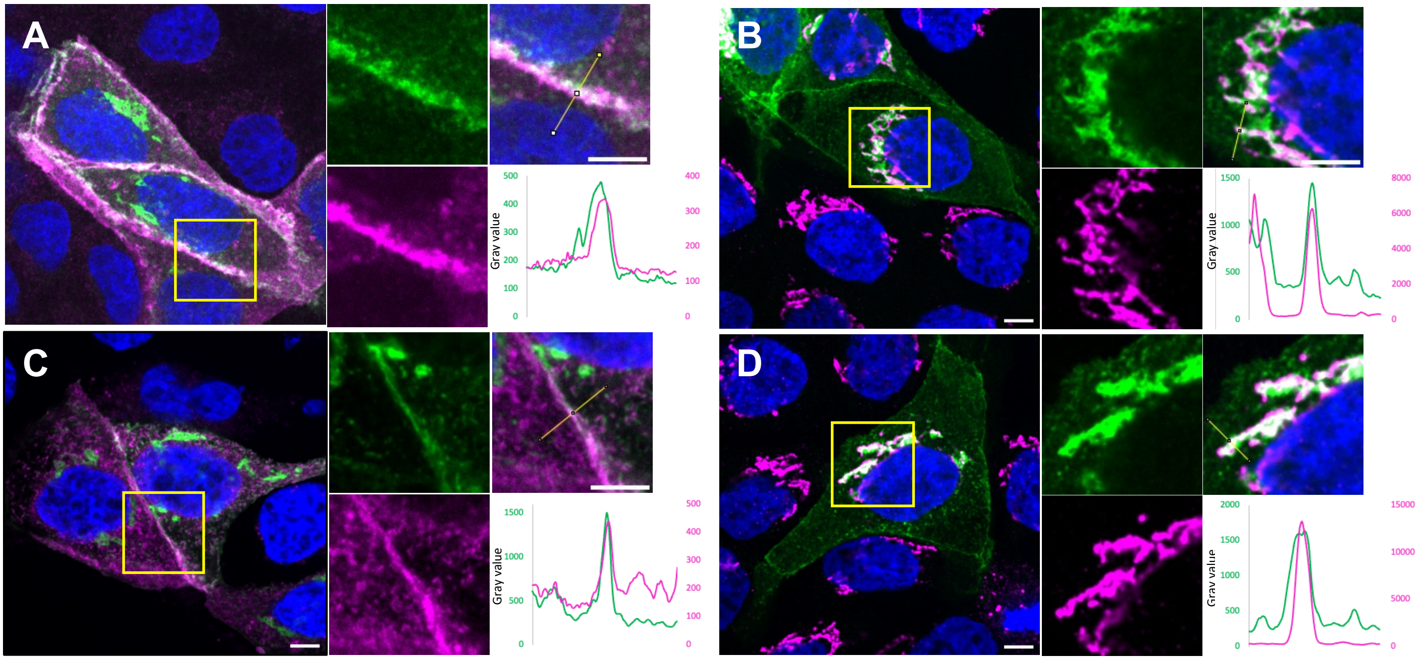
Supplementary Figure 1D


0.1 substitution/site
Arf1/6
Arf1 Hemimastix kukwesjijk
GokArf1a Gefionella okellyi
McaArf1a Malawimonas californ
Arf1 Nutomonas longa
PmiArf1a Planomonas micra
AgoArf1a Andalucia godoyi
Arf1 Willaertia magna
NgrArf1a Naegleria gruberi
EgrArf1b Euglena gracilis
Arf1 Paradiplonema papillatum
BsaArf1a Bodo saltans
TbrArf1b Trypanosoma brucei
LmaArf1a Leishmania major
TvaArf1f Trichomonas vaginalis
Arf1 Tritrichomonas foetus
MonArf1a Monocercomonoides exilis
TpyArf1d Paratrimastix pyriformis
Arf1 Carpediemonas membranifera
SvoArf1 Spironucleus vortens
GinArf1 Giardia intestinalis
Arf1 Giardia muris
Arf1 Rigifila ramosa
Arf1 Mantamonas sphyraenae
Arf1 Mantamonas vickermanii
MbaArf1a Mastigamoeba balamuthi
AcaArf1a Acanthamoeba castellanii
PpoArf1a Physarum polycephalum
DdiArf1 Dictyostelium discoideum
PygArf1a Pygsuia biforma
TtrArf1a Thecamonas trahens
FalArf1 Fonticula alba
Arf1 Paraphelidium tribonemae
GprArf1a Gonapodya prolifera
BdeArf1 Batrachochytrium dendrobatidis
CcoArf1 Conidiobolus coronatus
CorArf1 Coemansia reversa
CciArf1 Coprinopsis cinerea
UmaArf1 Ustilago maydis
CneArf1 Cryptococcus neoformans
SpoArf1 Schizosaccharomyces pombe
SceArf1 Saccharomyces cerevisiae
AfuArf1 Aspergillus fumigatus
NcrArf1 Neurospora crassa
SarArf1a Sphaeroforma arctica
CfrArf1a Creolimax fragrantis
Arf1 Syssomonas multiformis
Arf1 Pigoraptor chileana
CowArf1 Capsaspora owczarzaki
MbrArf1a Monosiga brevicollis
SroArf1a Salpingoeca rosetta
MleArf1a Mnemiopsis leidyi
OcaArf1 Oscarella carmela
TadArf1 Trichoplax adhaerens
NveArf1b Nematostella vectens
HmaArf1a Hydra magnipapillata
SkoArf1 Saccoglossus kowalevskii
SpuArf1a Strongylocentrotus purpuratus
BflArf1 Branchiostoma floridae
CinArf1a Ciona intestinalis
TruArf1c Takifugu rubripes
HsaArf1 Homo sapiens
CelArf1a Caenorhabditis elegans
DmeArf1 Drosophila melanogaster
DpuArf1 Daphnia pulex
SmaArf1 Schistosoma mansoni
LgiArf1 Lottia gigantea
CteArf1 Capitella teleta
HroArf1a Helobdella robusta
Arf1 Microheliella maris
GthArf1a Guillardia theta
PbiArf1a Palpitomonas bilix
GavArf1b Goniomonas avonlea
CpxArf1a Cyanophora paradoxa
GwiArf1a Gloeochaete wittrockiana
Arf1 Picozoa sp. COSAG02
Arf1 Rhodelphis marinus
GsuArf1b Galdieria sulphuraria
CcrArf1a Chondrus crispus
PpuArf1 Porphyridium purpureum
Arf1 Prasinoderma coloniale
MpnArfA Micromonas commoda
CreArfA1 Chlamydomonas reinhardtii
CsuArfA Coccomyxa subellipsoidea
KniArfA1 Klebsormidium nitens
PpaArfA3 Physcomitrella patens
SmoArfA1b Selaginella moellendorffii
AthARFA1e Arabidopsis thaliana
OsaArfA5 Oryza sativa
Arf1 Ancoracysta twista
Arf1 Telonema subtile
RaeArf1l Raineriophrys erinaceoides
PpiArf1a Pavlova pinguis
Arf1 Chrysochromulina parva
EhuArf1a Emiliania huxleyi
BnaArf1a Bigelowiella natans
RfiArf1a Reticulomyxa filosa
PlbArf1b Plasmodiophora brassicaceae
Arf1 Stentor coeruleus
Arf1 Euplotes vannus
TthArf1a Tetrahymena thermophila
Arf1 Colponemidia sp. Colp10
SmiArf1a Symbiodinium minutum
PmaArf1a Perkinsus marinus
ChvArf1a Chromera velia
CpaArf1a Cryptosporidium parvum
TgoArf1 Toxoplasma gondii
TanArf1 Theileria annulata
Arf1 Cafeteria roenbergensis
Arf1 Incisomonas marina
BhoArf1a Blastocystis sp. ST7
Arf1 Pseudophyllomitus vesiculosus
AliArf1b Aurantiochytrium limacinum
AkeArf1 Aplanochytrium kerguelensis
PsoArf1b Phytophthora sojae
HpaArf1b Hyaloperonospora arabidopsis
EsiArf1a Ectocarpus siliculos
VisArf1a Vischeria sp. Q 202
AanArf1a Aureococcus anophagefferens
TpsArf1 Thalassiosira pseudonana
PtrArf1 Phaeodactylum tricornutum
Arf1
PmiArf6 Planomonas micra
AgoArf6 Andalucia godoyi
Arf6 Carpediemonas membranifera
GokArf6 Gefionella okellyi
McaArf6 Malawimonas californiana
PbiArf6 Palpitomonas bilix
Arf6 Ancoracysta twista
RaeArf6 Raineriophrys erinaceoides
Arf6 Chrysochromulina parva
EhuArf6 Emiliania huxleyi
PsoArf6 Phytophthora sojae
Arf6 Cafeteria roenbergensis
AliArf6 Aurantiochytrium limacinum
AkeArf6 Aplanochytrium kerguelense
Arf6 Mantamonas vickermanii
PpoArf6 Physarum polycephalum
MbaArf6 Mastigamoeba balamuthi
AcaArf6 Acanthamoeba castellanii
PygArf6 Pygsuia biforma
FalArf6 Fonticula alba
Arf6 Paraphelidium tribonemae
GprArf6 Gonapodya prolifera
CcoArf6 Conidiobolus coronatus
CorArf6 Coemansia reversa
CciArf6 Coprinopsis cinerea
UmaArf6 Ustilago maydis
CneArf6 Cryptococcus neoformans
SpoArf6 Schizosaccharomyces pombe
SceArf6 Saccharomyces cerevisiae
AfuArf6 Aspergillus fumigatus
NcrArf6 Neurospora crassa
SarArf6 Sphaeroforma arctica
CfrArf6 Creolimax fragrantissima
Arf6 Syssomonas multiformis
Arf6 Pigoraptor chileana
CowArf6 Capsaspora owczarzaki
MbrArf6 Monosiga brevicollis
SroArf6 Salpingoeca rosetta
MleArf6 Mnemiopsis leidyi
OcaArf6 Oscarella carmela
TadArf6 Trichoplax adhaerens
NveArf6b Nematostella vectens
HmaArf6 Hydra magnipapillata
SkoArf6 Saccoglossus kowalevskii
SpuArf6 Strongylocentrotus purpuratus
BflArf6 Branchiostoma floridae
CinArf6 Ciona intestinalis
TruArf6b Takifugu rubripes
HsaArf6 Homo sapiens
CelArf6 Caenorhabditis elegans
DmeArf6 Drosophila melanogaster
DpuArf6 Daphnia pulex
SmaArf6 Schistosoma mansoni
LgiArf6 Lottia gigantea
CteArf6 Capitella teleta
HroArf6 Helobdella robusta
Arf6
PmiArl5 Planomonas micra
Arl5 Hemimastix kukwesjijk
GokArl5 Gefionella okellyi
McaArl5 Malawimonas californiana
TvaArl5 Trichomonas vaginalis
Arl5 Tritrichomonas foetus
Arl5 Willaertia magna
NgrArl5 Naegleria gruberi
EgrArl5 Euglena gracilis
Arl5 490 1020 Paradiplonema papillatum
BsaArl5 Bodo saltans
TbrArl5 Trypanosoma brucei
LmaArl5 Leishmania major
Arl5 Microheliella maris
GthArl5 Guillardia theta
GavArl5 Goniomonas avonlea
CpxArl5 Cyanophora paradoxa
GwiArl5 Gloeochaete wittrockiana
GsuArl5 Galdieria sulphuraria
CcrArl5 Chondrus crispus
PpuArl5 Porphyridium purpureum
CreArl5 Chlamydomonas reinhardtii
CsuArl5 Coccomyxa subellipsoidea
KniArl5 Klebsormidium nitens
PpaArl5a Physcomitrella patens
SmoArl5 Selaginella moellendorffii
AthARL5 Arabidopsis thaliana
OsaArl5 Oryza sativa
Arl5 Ancoracysta twista
Arl5 Telonema subtile
EhuArl5 Emiliania huxleyi
RaeArl5 Raineriophrys erinaceoides
PpiArl5 Pavlova pinguis
PlbArl5 Plasmodiophora brassicaceae
BnaArl5 Bigelowiella natans
AliArl5 Aurantiochytrium limacinum
Arl5 Colponemidia sp. Colp10
ChvArl5 Chromera velia
PmaArl5 Perkinsus marinus
Arl5 Stentor coeruleus
Arl5 Euplotes vannus
TthArl5 Tetrahymena thermophila
Arl5 Rigifila ramosa
Arl5 Mantamonas sphyraenae
Arl5 Mantamonas vickermanii
MbaArl5 Mastigamoeba balamuthi
AcaArl5 Acanthamoeba castellanii
PpoArl5 Physarum polycephalum
DdiArl5 Dictyostelium discoideum
PygArl5 Pygsuia biforma
TtrArl5 Thecamonas trahens
Arl5 Paraphelidium tribonemae
CneArl5 Cryptococcus neoformans
UmaArl5 Ustilago maydis
SarArl5 Sphaeroforma arctica
CfrArl5 Creolimax fragrantiss
Arl5 Syssomonas multiformis
Arl5 Pigoraptor chileana
CowArl5 Capsaspora owczarzaki
MbrArl5 Monosiga brevicollis
SroArl5 Salpingoeca rosetta
MleArl5 Mnemiopsis leidyi
OcaArl5 Oscarella carmela
TadArl5 Trichoplax adhaerens
NveArl5 Nematostella vectensis
HmaArl5 Hydra magnipapillata
SkoArl5 Saccoglossus kowalevskii
SpuArl5 Strongylocentrotus purpuratus
BflArl5 Branchiostoma floridae
CinArl5 Ciona intestinalis
TruArl5a Takifugu rubripes
HsaArl5b Homo sapiens
CelArl5 Caenorhabditis elegans
DmeArl5 Drosophila melanogaster
DpuArl5 Daphnia pulex
SmaArl5 Schistosoma mansoni
LgiArl5 Lottia gigantea
CteArl5 Capitella teleta
HroArl5 Helobdella robusta
Arl5
PmiArl1 Planomonas micra
GokArl1 Gefionella okellyi
McaArl1 Malawimonas californi
TvaArl1 Trichomonas vaginalis
Arl1 Carpediemonas membranifera
TpyArl1 Paratrimastix pyriformis
MonArl1 Monocercomonoides exilis
AgoArl1 Andalucia godoyi
Arl1 Willaertia magna
NgrArl1 Naegleria gruberi
EgrArl1 Euglena gracilis
Arl1 Paradiplonema papillatum
BsaArl1 Bodo saltans
TbrArl1 Trypanosoma brucei
LmaArl1 Leishmania major
Arl1 Mantamonas vickermanii
MbaArl1 Mastigamoeba balamuthi
AcaArl1 Acanthamoeba castellanii
PpoArl1 Physarum polycephalum
DdiArl1 Dictyostelium discoideum
PygArl1 Pygsuia biforma
TtrArl1 Thecamonas trahens
FalArl1 Fonticula alba
Arl1 Paraphelidium tribonemae
GprArl1 Gonapodya prolifera
BdeArl1 Batrachochytrium dendrobatidis
CcoArl1 Conidiobolus coronatus
CorArl1 Coemansia reversa
Arl1 Amoeboradix gromovi
CciArl1 Coprinopsis cinerea
UmaArl1 Ustilago maydis
CneArl1 Cryptococcus neoforma
SceArl1 Saccharomyces cerevisiae
AfuArl1 Aspergillus fumigatus
NcrArl1 Neurospora crassa
SarArl1 Sphaeroforma arctica
CfrArl1 Creolimax fragrantiss
Arl1 Syssomonas multiformis
Arl1 Pigoraptor chileana
CowArl1 Capsaspora owczarzaki
MbrArl1 Monosiga brevicollis
SroArl1 Salpingoeca rosetta
MleArl1 Mnemiopsis leidyi
OcaArl1 Oscarella carmela
TadArl1 Trichoplax adhaerens
HmaArl1 Hydra magnipapillata
NveArl1 Nematostella vectensi
SkoArl1 Saccoglossus kowalevskii
SpuArl1 Strongylocentrotus purpuratus
BflArl1 Branchiostoma floridae
CinArl1 Ciona intestinalis
TruArl1 Takifugu rubripes
HsaArl1 Homo sapiens
CelArl1 Caenorhabditis elegans
DmeArl1 Drosophila melanogaster
DpuArl1 Daphnia pulex
SmaArl1 Schistosoma mansoni
LgiArl1 Lottia gigantea
CteArl1 Capitella teleta
HroArl1 Helobdella robusta
Arl1 Microheliella maris
GthArl1 Guillardia theta
PbiArl1 Palpitomonas bilix
GavArl1 Goniomonas avonlea
CpxArl1 Cyanophora paradoxa
GwiArl1 Gloeochaete wittrockina
Arl1 Picozoa sp. COSAG02
Arl1 Rhodelphis marinus
GsuArl1 Galdieria sulphuraria
CcrArl1 Chondrus crispus
PpuArl1 Porphyridium purpureum
Arl1 Prasinoderma coloniale
MpnArl1 Micromonas commoda
CreArl1 Chlamydomonas reinhardtii
CsuArl1 Coccomyxa subellipsoidea
KniArl1 Klebsormidium nitens
PpaArl1b Physcomitrella patens
SmoArl1 Selaginella moellendorffii
AthARL1 Arabidopsis thaliana
OsaArl1 Oryza sativa
Arl1 Ancoracysta twista
Arl1 Telonema subtile
BnaArl1 Bigelowiella natans
RfiArl1 Reticulomyxa filosa
PlbArl1 Plasmodiophora brassicaceae
RaeArl1 Raineriophrys erinaceoides
PpiArl1 Pavlova pinguis
Arl1 Chrysochromulina parva
EhuArl1 Emiliania huxleyi
Arl1 Stentor coeruleus
Arl1 Euplotes vannus
TthArl1 Tetrahymena thermophila
Arl1 Colponemidia sp. Colp10
SmiArl1 Symbiodinium minutum
ChvArl1 Chromera velia
PmaArl1 Perkinsus marinus
BhoArl1a Blastocystis sp. ST7
Arl1 Cafeteria roenbergensis
Arl1 Pseudophyllomitus vesiculosus
AliArl1 Aurantiochytrium limacinum
AkeArl1 Aplanochytrium kerguelense
PsoArl1 Phytophthora sojae
HpaArl1 Hyaloperonospora arabidopsis
EsiArl1 Ectocarpus siliculosus
VisArl1 Vischeria sp. Q 202
AanArl1 Aureococcus anophagefferens
TpsArl1 Thalassiosira pseudonana
PtrArl1 Phaeodactylum tricornutum
Arl1

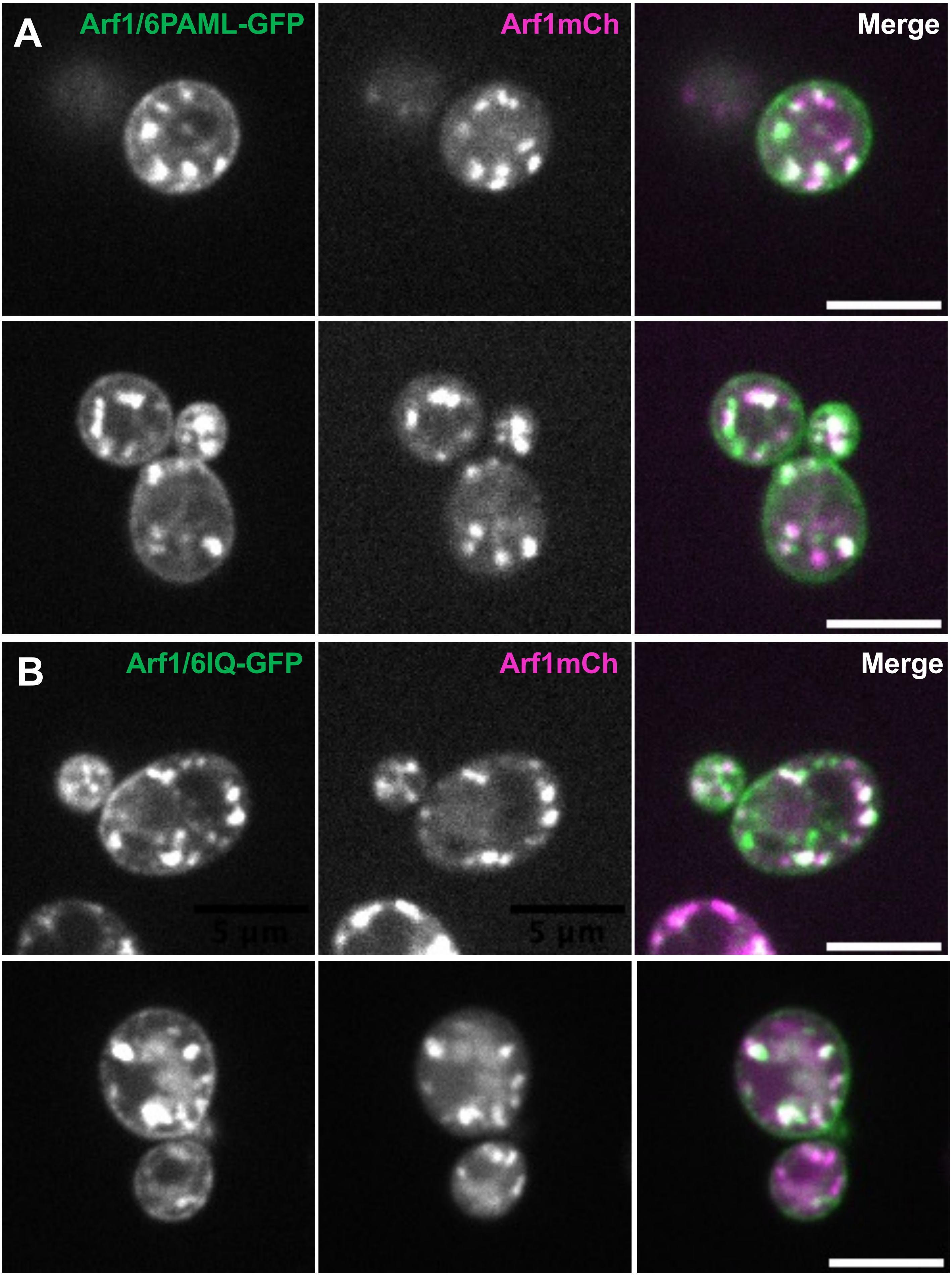
Supplementary Figure 2


A
GFP
2+GFP
Human Arf1
Human Arf6
Arf1/6PAML
Arf1/6PAML+1
Arf1/6PAML-3KQ
Arf1/6_IQTREE
Arf1/6IQTREE+1
Arf1/6IQ-4KQ
kDa
55
35
25
B
C

Supplementary Figure 3


A
Gray value
500
400
300
200
100
0
400
300
200
100
0
B
Gray value
1500
1000
500
0
8000
6000
4000
2000
0
C
Gray value
1500
1000
500
0
500
400
300
200
100
0
D
Grav value
2000
1500
1000
500
0
15000
10000
5000
0

Supplementary Figure 4


A
Arf1/6PAML-GFP
Arf1mCh
Merge
B
Arf1/6IQ-GFP
Arf1mCh
Merge
5 µm
5 µm

Supplementary Figure 5