Unanticipated diversity, undiscovered ancient paralogs, and Asgard origins of the ARF GTPase protein family

Romana Vargová[1], Jeremy G. Wideman[2], Romain Derelle[3], Richard A. Kahn[4], Vladimír Klimeš[1], Joel B. Dacks[5,6]*, Marek Eliáš[1]*

[1]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, Ostrava, Czech Republic

[2]Biodesign Center for Mechanisms of Evolution, School of Life Sciences, Arizona State University, Tempe, AZ, 85287, USA

[3]Station d'Ecologie Théorique et Expérimentale, UMR CNRS 5321, Moulis, France

[4]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322

[5]Division of Infectious Disease, Department of Medicine, University of Alberta

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

*For correspondence: dacks@ualberta.ca (JBD) and marek.elias@osu.cz (ME)

Abstract

ARF family GTPases act in diverse cellular processes, critical for organellar function today and understanding eukaryotic origins. However, our understanding of ARF family evolution is limited. Our phylogenetically comprehensive *in silico* analyses of ARF family members here doubles the set of ancestral eukaryotic paralogs and challenges existing norms for small GTPases, with examples of non-standard modes of membrane association and novel protein architectures. Evidence for the pan-eukaryotic and ancestral origin of Arf6, Arl13 and Arl16 is presented for the first time, while three newly described ancient sub-families are absent from well-studied model organisms, leaving their functions completely unexplored. Evolutionary analysis of metazoa-specific ARF family GTPases also uncovered several new proteins in animals. Delving back to eukaryogenesis, the relationship within the ARF GTPases sets boundaries for scenarios of vesicle coat origins. Finally, we report the discovery of the archaeal proteins from which the entire eukaryotic ARF family is derived.

Introduction

Understanding how the eukaryotic cell evolved in all its complexity is one of the greatest open questions in evolutionary biology. Eukaryogenesis involved both the origin of new genes and the diversification of key building blocks (Dacks et al., 2016; Eme et al., 2017). Among the different building blocks, particular groups of proteins radiated early in the evolution of eukaryotes and are represented by a large number of pan-eukaryotic orthologs, presumably with conserved functions. One of the largest groups of proteins, acting in an incredibly diverse array of cellular pathways, is the Ras superfamily of GTPases. This superfamily is frequently equated with familiar and extensively studied eukaryotic "small GTPases". However, the more appropriate, i.e. evolutionary, definition conceives it as a major monophyletic subgroup of the vast TRAFAC class of GTPases that also includes prokaryotic representatives, larger proteins combining a Ras-related GTPase domain with other functional domains, and – surprisingly to many in the field – the alpha subunits of heterotrimeric G-proteins (Leipe et al., 2002). Because of its central role in so many fundamental cellular functions, understanding the origin and evolution of this complex superfamily of proteins is necessary for uncovering the processes by which eukaryotes evolved and diversified.

The internal classification of the Ras superfamily is unsettled. In many overviews, especially those concentrating on the eukaryotic small GTPases, the content of the superfamily is pigeonholed into five major families (Ras, Rho, Rab, Ran, Arf/Sar; Rojas et

al., 2012), but this scheme ignores the prokaryotic superfamily members (Wuichet and Søgaard-Andersen, 2014), multi-domain proteins (such as the ROCO family; Bosgraaf and Van Haastert, 2003), and various other lineages clearly distinct from or not easily classified into the well known families, such as the Gtr/Rag family (Klinger et al., 2016) or RJL proteins (Elias and Archibald, 2009). The Ras superfamily genes in extant eukaryotes ultimately derive from a set of genes present in the Last Eukaryote Common Ancestor (LECA). These ancient genes are now represented by tens to hundreds of descendent genes in any given extant eukaryotic genome (Rojas et al., 2012). This complexity of Ras superfamily members within each extant genome can be traced to expansions that occurred in three different evolutionary strata: (1) in prokaryotes, before the emergence of the First Eukaryote Common Ancestor (FECA); (2) during eukaryogenesis (i.e. the FECA-to-LECA transition); (3) and after the emergence of LECA during the diversification of extant eukaryotic lineages.

Some recent studies have specifically investigated the first evolutionary stratum, the prokaryotic origins of this influential protein family. That the eukaryotic Ras superfamily may have two different prokaryotic sources was proposed a decade ago (Dong et al., 2007) and supported by later studies (Yutin et al., 2009; Wuichet and Søgaard-Andersen, 2014), most recently, by the discovery of the Asgard, the archaebacterial superphylum most closely related to eukaryotes (Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017). Analyses of metagenomically reconstructed partial genome sequences from several members of this superphylum unveiled a surprisingly large set of Ras superfamily genes unprecedented among prokaryotes, but similar to numbers in eukaryotes. Detailed analyses of small GTPase genes from one particular Asgard archaebacterium, *Lokiarchaeum* sp. GC14_75, indicated that the eukaryotic genes of the Ras/Rho/Rab/Ran group in fact come from an archaebacterial ancestor (Klinger et al., 2016; Surkont and Pereira-Leal, 2016). In addition, close relatives of the Rag/Gtr family were identified in the *Lokiarchaeum* genome, revealing that this branch of the Ras superfamily evolved prior to the FECA. A group of lokiarchaeal GTPases was denoted ArfL (Arf-like) because of close sequence similarity to eukaryotic Arfs, but phylogenetic analyses remained inconclusive (Klinger et al., 2016). Numerous Ras superfamily GTPases were also found in genomes of other Asgard archaebacteria (including members of Thor-, Odin- and Heimdallarchaeota; Zaremba-Niedzwiedzka et al., 2017), but their relationships to the lokiarchaeal and eukaryotic homologs have not yet been investigated in detail.

Several evolutionary analyses have addressed the second and third evolutionary strata, reconstructing LECA's complement of particular Ras superfamily subgroups and detailing

the downstream innovation within eukaryotes. Prominent examples include analyses of the Rab (Diekmann et al., 2011; Elias et al., 2012; Klöpper et al., 2012) and Ras families (van Dam et al., 2011), and some isolated lineages like RJL (Elias and Archibald, 2009) or Miro (Vlahou et al., 2011). These investigations demonstrated that a large number of functionally investigated paralogs were present in the LECA, emphasizing the role of loss or streamlining of genomic complement in many eukaryotic lineages. They also identified ancient LECA paralogs of unknown function that have been lost in lineages leading to conventional model systems but which are present in diverse eukaryotic lineages of ecological and medical importance. Paralogs with such an evolutionary distribution were recently coined jotnarlogs (More et al., 2020). Finally, these studies also inevitably shed light on the diversification of GTPases in the post-LECA expansion phase. For example, divergent paralogs of unclear evolutionary relationships are found in various taxa (e.g., Pereira-Leal, 2008), most likely resulting from rapid sequence evolution of lineage-specific paralogs linked to their neofunctionalization. Additionally, the inherently small nature of the GTPases makes them particularly susceptible to molecular tinkering, such as accretion of additional domains or gain/loss of motifs mediating specific post-translational modifications (e.g., Záhonová et al., 2018).

Not yet addressed in a comparable evolutionary framework is the ARF protein family. This large protein family is comprised of the "true" ADP Ribosylation Factors (i.e., Arfs), as well as Arf-like proteins (Arls), Arf-related protein 1 (Arfrp1), and Sar1. Clearly related are the beta subunits of the signal recognition particle receptor (SRβ; Schwartz and Blobel, 2003). Sequence analyses have also revealed that an Arf-like ancestor, modified by insertion of a novel α-helical region into its GTPase domain and high sequence divergence, gave rise to the alpha subunits of heterotrimeric G-proteins (abbreviated Gα; Neuwald, 2007; Anantharaman et al., 2011). The distinction between Arf and Arf-like (Arl) proteins was originally made based upon activity in the cholera toxin-catalyzed ADP-ribosylation of the stimulator of adenylyl cyclase, $G\alpha_s$, as all tested Arfs retain this functionality while the Arls did not (Tamkun et al., 1991; Clark et al., 1993). However, this activity has proven of very limited utility in studies of cellular functions for ARF family members as greater appreciation of both the size of the family in model organisms as well as the diversity of functions became clear. Thus, little if any weight should be given to whether a gene is named as an Arf, an Arl, an Arfrp1, or a Sar. The ARF family is functionally heterogeneous and comprises proteins involved in membrane vesicle formation (Arfs, Sar1), other aspects of vesicle traffic and maintenance of membranous organelle morphology (e.g., Arl1, Arl5, or Arfrp1), microtubule

dynamics and mitochondrial fusion (Arl2), and cilium biogenesis and function (Arl3, Arl6, Arl13) (Gillingham and Munro, 2007; Donaldson and Jackson, 2011; Francis et al., 2016). Members of this family are critical to these diverse cellular activities and dysfunction results in numerous human diseases. Family members are generally considered to be single-domain small GTPases. Post-translational modifications (N-terminal myristoylation or acetylation) are also often critical to the protein's localization and function.

An early phylogenetic study on the ARF family, limited by a lack of taxonomic breadth in available genomic sequences, provided an early estimate of the ancient complexity of the family in LECA and identified putative lineage-specific expansions in metazoans (Li et al., 2004). The analyses showed that LECA contained at least eight ancient groups of orthologs inferred from representatives being present in metazoans and at least one non-opisthokont (protist or plant) eukaryote. This analysis also demonstrated that some of the metazoan family members lacked close relatives in other eukaryotes, suggesting that lineage-specific expansions related to metazoan multicellularity occurred. However, since that study, several family members from the target species (including humans) have been identified (Kahn et al., 2006) and a much broader taxon sampling, including members of the Asgard archaea, is available. In addition, progress in methods of phylogenetic analyses of protein sequences, including the development of the ScrollSaw approach facilitating analyses of complex paralog-rich families (Elias et al., 2012), now invite attempts to obtain a much better picture of the evolution of ARF family.

To this end, we assembled, extensively curated and phylogenetically analysed a dataset of ARF family sequences from a taxonomically broad selection of eukaryotic species. This enabled us to revise the set of ancestral eukaryotic ARF family paralogs, which has now expanded to between 14 and 16 genes. Two sub-families, described here for the first time, are not represented in well studied models and point to hitherto unstudied molecular functions mediated by the ARF family. We observed an unexpected diversity of domain architectures challenging the dogma that ARF family proteins are only small and single-domain proteins. Our analyses also unveiled a range of predicted post-translational modifications (PTMs), including but not limited to well-established N-terminal myristoylation, and other molecular adaptations that facilitate membrane association as a central feature of ARF family biology. In addition, we identified well supported relationships between the sub-families, which have implications for the inferred function of the primordial family members during eukaryogenesis. Finally, our (re)analyses of archaeal GTPases unveiled the actual closest prokaryotic relatives of the eukaryotic ARF family.

Results and Discussion

A comprehensive dataset and phylogeny of the ARF family

We first gathered all ARF family sequences (including SRβ but excluding the highly divergent Gα proteins) from a broad diversity of eukaryotes, exploiting both publicly available and privately curated genomes and transcriptomes. We did not rely solely on predicted protein sequence sets but also checked the genome and transcriptome assemblies to ensure maximal accuracy when it comes to statements about the absence of particular genes in different taxa. All sequences were carefully validated, as described under Materials and Methods, and when needed, edited (by modifications of the originally predicted gene models or by changes in the assembled nucleotide sequences based on inspection of raw sequencing data) to ensure maximal quality and completeness of the data. Our final dataset, provided as supplementary Dataset 1, included >2,000 manually curated sequences from 114 species (Supplementary table 1). The number of ARF family genes in individual species ranged from 5 in the yeast *Schizosaccharomyces pombe* to 70 in the rotifer *Adineta vaga* (this high number apparently reflecting the tetraploid origin of its genome; Flot et al., 2013).

The genes were initially annotated based on their similarity to previously characterized or named ARF family genes in model organisms scored by BLAST. While this procedure enabled us to recognize candidate groups of orthologs and to assign most of the genes into these groups, the assignment of many sequences was uncertain or unclear and a more rigorous method for establishing orthologous relationships – phylogenetic analysis of a multiple sequence alignment – was required to corroborate the proposed groups of orthologs and to possibly identify additional ones not readily apparent from sequence-similarity comparisons. Such an analysis of the whole dataset was impractical, if not impossible, for its size and the existence of divergent sequences that tend to disrupt the results of phylogenetic inference. We therefore utilized of the ScrollSaw protocol previously developed to deal with a similarly complex family of Rab GTPases (Elias et al., 2012). This protocol enables one to infer a “backbone” phylogeny of a protein family by concentrating on preselected sequences likely representing slowly-evolving members of the main clades and thus conserved across distantly related organismal lineages. Briefly (see Materials and Methods for details), we divided the sampled species into 13 groups corresponding to major eukaryotic lineages, and for each pair of groups we identified all pairs of sequences (the two sequences representing the two different groups) that had mutually minimal genetic distances calculated by the maximum likelihood method from a multiple sequence alignment. We then gathered all the sequence pairs of all the comparisons, removed redundancies, and inferred trees from the full

resulting dataset (Supplementary figure 1) or after pruning sequences from selected species to further decrease the complexity of the analysis (Figure 1; Supplementary figure 2). This resulted in a final phylogeny, which enabled us to infer various aspects about the evolutionary and diversity history of the ARF family in eukaryotes.

LECA possessed an extensive array of ARF family paralogs

Dissection of the "ScrollSaw" trees next enabled us to define 13 clades that seem to correspond to paralogs differentiated before the divergence of major eukaryotic lineages and perhaps present in the LECA, as each Arf clade is represented in both major eukaryotic clades (Opimoda and Diphoda) defined by the most recent hypothesis on the position of the root of the eukaryote phylogeny (Derelle et al., 2015). Most of these clades have received high statistical support (posterior probability, SH-aLRT support, and ultrafast bootstrap values greater than or equal to 0.98, 98, and 98, respectively) (Figure 1). An exception is the clade denoted Arf1 and comprising prototypical Arf sequences, but there is little doubt that it constitutes a coherent group of orthologs. The weak signal for its monophyly may stem from a very slow evolution of Arf1 sequences (apparent also from very short branches in the tree) having precluded accumulation of paralog-specific sequence features that would enable strong phylogenetic separation from the related, more rapidly evolving (and much more strongly supported) paralogs. Nevertheless, a focused analysis restricted to Arf1, Arf6, Arl1 and Arl5 allowed us to use protein alignment with more positions and recovered Arf1 as a supported monophyletic clade (Supplementary figure 3). We previously observed a similar behaviour with the highly conserved Rab1 paralog, whose undoubted monophyly was also difficult to recover (Elias et al., 2012).

The existence of two separate clades of Arfs originated before the divergence of metazoans, fungi, and plants was hypothesized previously but not convincingly demonstrated (Li et al., 2004). Our study is the first to show that the opisthokont class III Arfs, typified by the sole mammalian member, Arf6, have robustly supported orthologs in various protists spanning the phylogenetic breadth of eukaryotes. The existence of a separate eukaryotic Arf6 clade is further supported by comparison of intron positions in Arf genes (Supplementary figure 4). In contrast, the mammalian Arf1-Arf5 proteins (class I and II Arfs) all cluster into the Arf1 clade. Our analyses also newly demonstrate that the metazoan Arl16 has orthologs present in diverse protists and thus represents an ancient ARF family paralog. Another previously unrecognized ancient paralog was missed because it is not represented in metazoans and has no characterized or named member. As it is most closely related to Arl8, we propose to call it

Arl18. The separation of Arl8 and Arl18 is apparent not only from the phylogenetic analysis (Figure 1, Supplementary figure 2) but also from their exon-intron structure (Supplementary figure 5).

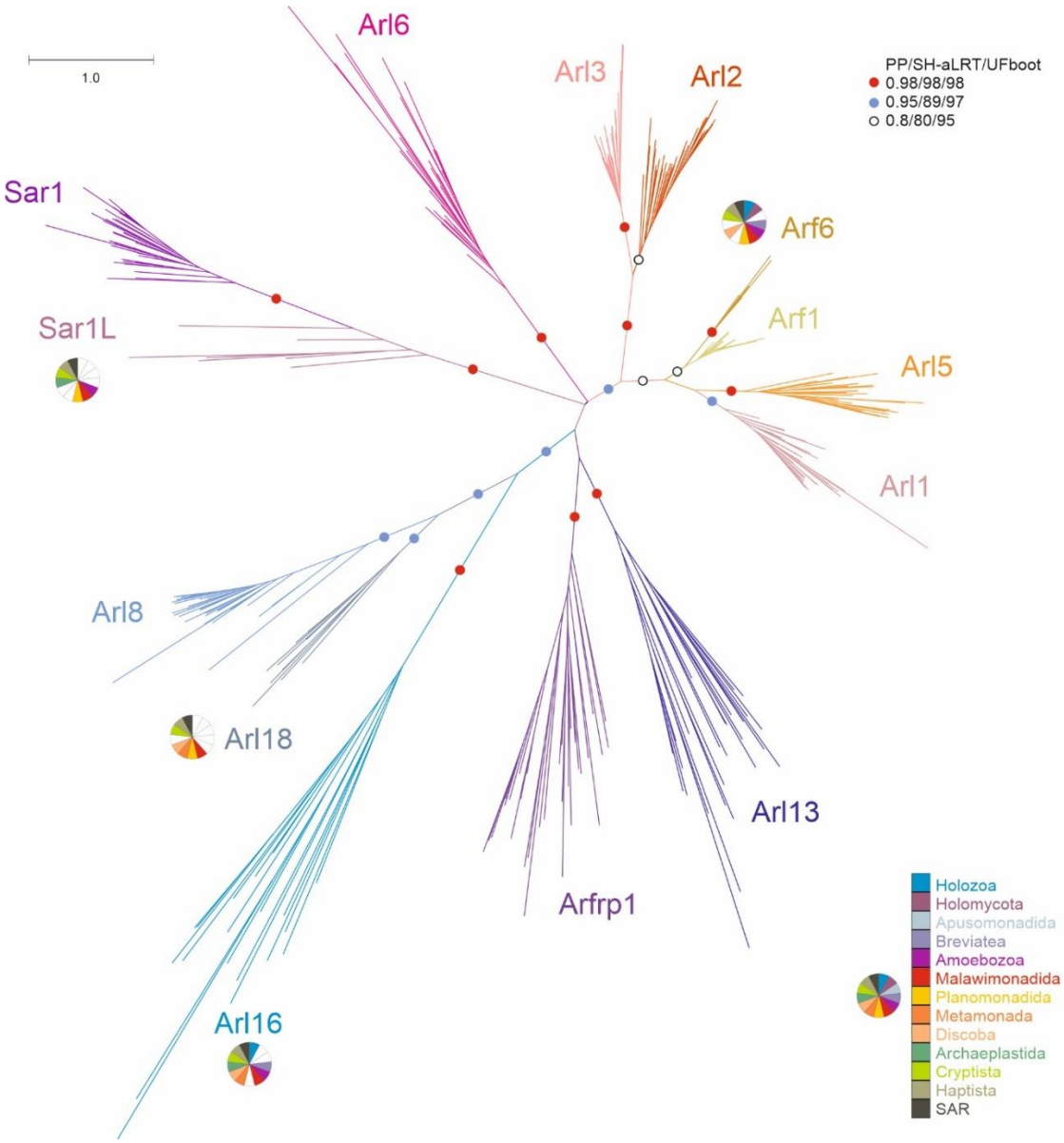

Figure 1. Maximum likelihood phylogenetic tree of the ARF family based on a reduced ScrollSaw dataset. The tree was inferred using IQ-TREE with LG+I+G4 model (the model selected by the program itself) with the ultrafast bootstrap algorithm and the SH-aLRT test (both 10000 replicates), as described under Materials and Methods. Dots at branches represent bootstrap values as indicated in the graphical legend (top right). The bar on the top corresponds to the estimated number of substitutions per site. The pie charts indicate the occurrence of Arf6, Arl16, Arl18 and Sar1L in main eukaryotic lineages (indicated by different colours explained in the graphical legend in the lower right). The remaining paralogs have ubiquitous distribution (i.e., are present in all main lineages analysed). A full version of the tree is provided in Supplementary figure 2.

Two additional ancient eukaryotic ARF family paralogs may exist, although not unambiguously supported by present analyses. One is suggested by a group of Sar1-like (Sar1L) genes from phylogenetically diverse protist lineages that branch off basally to the strongly supported Sar1 clade (Figure 1; Supplementary figure 2). These are not simply divergent Sar1 orthologs, as they always co-occur with a *bona fide* Sar1 in each species. We call these genes Sar1L and tentatively consider them as representatives of one ancient eukaryotic paralog despite their non-monophyly in the phylogenetic tree. It is of note that some intron positions in Sar1L genes are exclusive for this group and not shared with Sar1 (Supplementary figure 6). Sar1L sequences also differ from Sar1 in the functionally important Walker B motif, also known as G3 motif/box (Leipe et al., 2002): whereas Sar1L exhibits a conserved tryptophan residue shared by other ARF family members and Gα proteins, Sar1 typically has a phenylalanine residue at this position (Supplementary figure 7; Vetter, 2014).

The second additional potential ancient paralog, here proposed to be called Arl17, is represented by uncharacterized proteins from various protists, certain fungi, and a single metazoan lineage. These proteins are comprised of a GTPase domain most similar to Arfs and one, two, or three non-identical copies of a novel conserved domain (Figure 2; Supplementary figure 8). Despite their unique domain architecture, no Arl17 sequences passed the ScrollSaw filter and their monophyly is suggested, but not supported, by phylogenetic analysis (Figure 2), which may reflect the fact that the GTPase domain in Arl17 proteins tends to be very divergent (Supplementary figure 9). The novel ~100 residue, C-terminal domain displays no discernible homology to previously described domains (even when tested by the highly sensitive HHpred searches) despite being present also in other proteins from some opisthokonts and bacteria, either as a stand-alone protein (e.g., EGF92317.1) or in combination with various non-GTPase domains (e.g., XP_004347279.1). Having established the main lineages of the ARF family, we attempted to assign all genes from our full dataset (including those that were excluded by the ScrollSaw protocol) into them by considering sequence similarity scored by BLAST, comparison to lineage-specific profile HMMs by HMMER, and by targeted phylogenetic analyses. The majority of genes in our dataset could be allocated with confidence to a specific, ancient ARF family paralog, enabling us to evaluate the pattern of retention of the ancient paralogs in modern eukaryotes and to map the presumed gene losses to the eukaryote phylogeny (Figure 3; Supplementary table 2). Nevertheless, a relatively small number of genes (160 out of > 2,000 sequences) remained unclassified. A majority of these likely correspond to taxon-specific duplications of

the standard ARF family members that have diverged substantially obscuring their actual evolutionary origin. Some cases, however, may represent excessively divergent, unrecognized direct orthologs of the widespread genes. For example, several unclassified genes showed potential affinity to Arf6, yet without significant support. These sequences all share one or more intron positions specific to Arf6 (Supplementary figure 4), supporting their annotation as highly derived Arf6 genes. Future studies with a more comprehensive sampling may help resolve cases such as this.

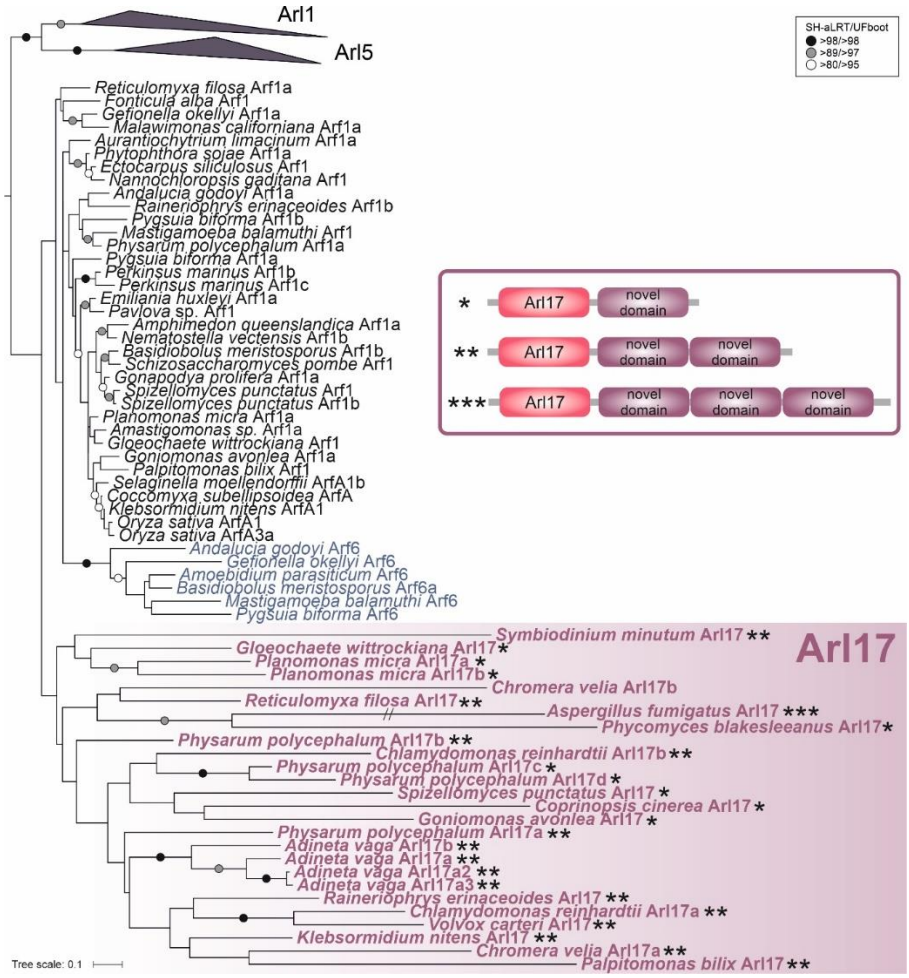

Figure 2. Phylogenetic analysis and domain architecture of Arl17. The tree shown is a result of a ML analysis of all Arl17 sequences and a subset of the reduced "scrollsawed" dataset restricted to Arf1, Arf6, Arl1, and Arl5 sequences (the latter two collapsed as triangles). The alignment was trimmed manually. The tree was inferred using IQ-TREE with LG+I+G4 model (the model selected by the program itself) with the ultrafast bootstrap algorithm and the SH-aLRT test (both 10000 replicates). Dots at branches represent bootstrap values as indicated in the graphical legend (top right). Inset: schematic representation of three different variants of the Arl17 domain architecture (correspondence to specific proteins in the tree is indicated by the asterisks). The exact architecture of the *Ch. velia* Arl17b protein could not be determine due to incompleteness of the genome assembly.

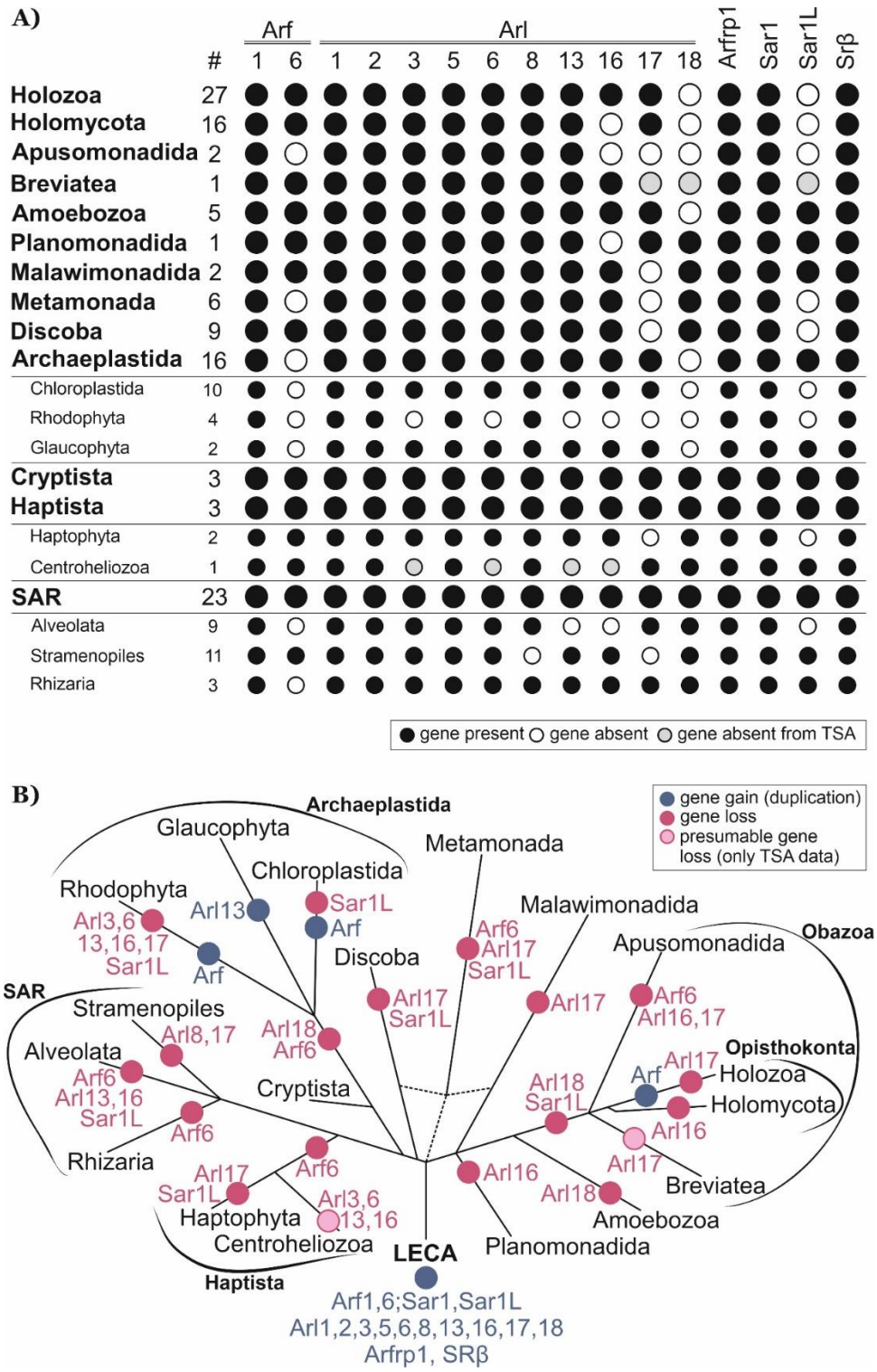

Figure 3. Retention of ancient paralogs of the ARF family in main lineages of eukaryotes. (**A**) Black circle: the paralog is present in at least one member of the lineage. White circle: the gene is absent from the lineage (evidenced by genome sequence data). Grey circle: the gene was not found in the transcriptome data available (lineages with transcriptome assemblies only). The hashtag (#) indicates the number of species included in the analysis. (**B**) Gene gains (blue circles) and losses (pink circles) mapped onto the eukaryote phylogeny. Only duplications specific to whole lineages listed in the picture are considered. The acquisition of Arl17 via HGT in rotifers (here represented by *A. vaga*) is indicated with a blue circle with an asterisk within.

Complex cellular repertoire inferred from the LECA complement

The analyses presented above indicate that the LECA possessed at least 15 ARF family genes; Arf1 and 6, Arl1, 2, 3, 5, 6, 8, 13, 16, 17, and 18, Arfrp1, Sar1, and Sar1L. In addition, it certainly encoded SRβ, excluded from our ScrollSaw analysis (hence absent from the trees in Figure 1 and Supplementary figure 1) due to its marked divergence from the (core) ARF family and because SRβ orthologs can be unambiguously recognized by sequence similarity. Eight of these clades (Arf1, Arl1, 2, 3, 5, and 8, Arfrp1, and Sar1) were previously recognized as likely ancient (Li et al., 2004) and the existence of orthologs of the metazoan Arl13 in protists was also noted (e.g., Miertzschke et al., 2014), although perhaps never documented by a phylogenetic analysis. Our analysis thus doubles the reconstructed complement in LECA and further strengthens the idea that the LECA was a fully-fledged eukaryotic cell making broad use of a complex molecular machinery.

The cellular function of many of the 16 ARF family GTPases in the LECA in principle can be inferred from what has been learned about their descendants in modern eukaryotes, although our present knowledge about the function of various GTPases comes from a limited number of phylogenetically biased model eukaryotes (primarily metazoans and the yeast *Saccharomyces cerevisiae*, i.e. the opisthokonts) and it is not always certain to what extent we can generalize from them to eukaryotes as a whole. In addition, each ARF family member studied in any depth in mammalian cells has been found to act in more than one pathway and typically with multiple downstream effectors (Kahn et al., 2009; Sztul et al., 2019), often making it difficult to assess which of these are ancient and which were acquired at later times. With this caveat in mind, we summarize the key findings about the different paralogs to paint a hypothetical picture of the cellular engagement of the ARF family members in the LECA.

Most of the ARF family paralogs clearly play a role in the endomembrane dynamics. As a subunit of the receptor of the signal recognition particle, SRβ mediates co-translational import of proteins into the ER (Schwartz and Blobel, 2003). Sar1 also associates with the ER and recruits subunits of the COPII coat complex to promote budding of transport vesicles from the ER (Miller and Barlowe, 2010). Four paralogs – Arf1, Arfrp1, Arl1 and Arl5 – are physically and functionally associated with the Golgi/*trans*-Golgi network (TGN). One key function of Arf1 (including the metazoan Arf1 to Arf5) is to recruit different types of vesicle coats (COPI, AP-1/clathrin, AP-3) to different parts of the Golgi (Jackson and Bouvet, 2014). Arl1 and Arfrp1 (confusingly called Arl3p in the yeast *S. cerevisiae*) are functionally linked,

the latter shown to be critical for Arl1 recruitment to the *trans*-Golgi in both yeast and mammalian cells (Panic et al., 2003; Setty et al., 2003; Zahn et al., 2006). Arl1 recruits several effectors (e.g. golgins, arfaptins, and Arf-GEFs) to the *trans*-Golgi network (TGN) and is important for endosome-to-TGN traffic (Yu and Lee, 2017). The function of Arl5 is less-well understood, but it may partly overlap with that of Arl1, as it also localizes to the *trans*-Golgi (Houghton et al., 2012), and both the fly Arl5 and the yeast Arl1 each interact with the GARP tethering complex (Panic et al., 2003; Rosa-Ferreira et al., 2015). In contrast to the Golgi localizing and acting members of the ARF family, Arf6 acts predominantly at the cell surface and endosomes to mediate endosome recycling, cell motility, and membrane extensions, which together influence cell division and lipid/cholesterol metabolism (Funakoshi et al., 2011; Schweitzer et al., 2011). Arl8 has been implicated in controlling lysosomal motility and traffic in metazoan cells (Khater et al., 2015). Its localization to the vacuolar membranes in *A. thaliana* (Heazlewood et al., 2007) suggests that functional association of Arl8 with the lysosomal/vacuolar compartment is ancestral and conserved.

Three paralogs, Arl3, Arl6, and Arl13 have been implicated in flagellar function (Fisher et al., 2020). Arl3 has been proposed to regulate the delivery of N-myristoylated and prenylated proteins to the cilium (Fansa and Wittinghofer, 2016; Stephen and Ismail, 2016). Arl6 (also called BBS3) regulates the function of the BBSome (a protein complex involved in intraflagellar transport; Mourão et al., 2014). Arl13 is involved in ciliary protein import and export, at least partly mediated by its activity as a positive regulator (guanine nucleotide exchange factor, GEF) for Arl3 (Gotthardt et al., 2015; Ivanova et al., 2017). Arl2 shares some effectors with Arl3 and is probably involved in traffic of lipidated proteins (Van Valkenburgh et al., 2001; Fansa and Wittinghofer, 2016), but it has its own specific agenda, as it regulates the assembly of αβ-tubulin dimers (Al-Bassam, 2017; Francis et al., 2017a; Francis et al., 2017b) and mitochondrial fusion (Newman et al., 2017).

Only a single study addressing the function of Arl16 has been published, reporting that the mammalian Arl16 inhibits the function of the RIG-I protein, involved in the defence against RNA viruses (Yang et al., 2011), but more specific functional insights are lacking. Functions for of the newly discovered paralogs Sar1L, Arl17, and Arl18 are completely unknown, as these paralogs are missing from all common model eukaryotes and thus represent examples of “jotnarlogs”, proteins that are present across eukaryotes, but missing in well-studied cell biological models (More et al., 2020). This adds further credence to the proposal that this is a substantial evolutionary cell biological phenomenon and highlights the gap in our understanding of the cell biology of the ARF family in eukaryotes. Nevertheless,

some clues as to the function of these proteins are provided by the phylogenetic relationship to other paralogs, as relatedness within the ARF family appears to signify functional similarity. Indeed, the aforementioned functional aspects shared by the pairs Arl2-Arl3 and Arl1-Arl5 are reflected by close relationships of the paralogs in the pairs (Figure 1). Likewise, the related Arf1 and Arf6 paralogs, although different in terms of the intracellular localization and effectors they deploy (Jackson and Bouvet, 2014), share the same class of GEFs and GTPase activating proteins, though to a very incompletely characterized extent (GAPs; Casanova, 2007; Kahn et al., 2008; Sztul et al., 2019). Hence, by analogy we propose that Arl18 may have similar functional attributes as its closest paralog Arl8 (e.g. it may likewise function in the lysosomal/vacuolar sector of the endomembrane system), and that Sar1L functions similarly to the canonical Sar1 protein in the secretory pathway (Sato and Nakano, 2007; Melville et al., 2020). The specific relationship of Arl17 and true Arfs may be less informative concerning the function of the latter, given the unique domain architecture of Arl17 proteins and the generally divergent nature of their GTPase domains (compare the branch lengths of Arl17 sequences in the tree in Figure 2).

Phylogenetic profiles of some ancestral eukaryotic ARF family paralogs illuminate differential simplification of endomembrane system functions in eukaryote evolution

A detailed scrutiny of the taxonomic distribution of some of the ancestral ARF family paralogs in extant eukaryotes provides interesting insights into the variation of their roles in cell functions across eukaryotes. While a hallmark of the ARF family perhaps is that members are commonly found to be active in multiple, distinct pathways in the same cells (Francis et al., 2016; Sztul et al., 2019), here we discuss their known or predicted functionalities with respect to their best known activities, recognizing the limitations that result.

Arfs (specifically the Arf1 paralog), Sar1, and SRβ are all found in every eukaryote sampled (with one exception in case of SRβ, most likely due to incompleteness of the data; Supplementary table 2), indicating that they belong to the functional core of the eukaryotic protein toolkit. Nearly ubiquitous is Arl2, being absent only from *Entamoeba histolytica*. Inspection of genomes of other *Entamoeba* species suggest that Arl2 loss is not an artefact and predates the radiation of the genus. Given the role of Arl2 in the assembly of tubulin dimers and in mitochondrial fusion (Francis et al., 2016), its absence in *Entamoeba* may be related to a unique combination of traits of this taxon including divergent tubulin sequences

and a highly reduced microtubular cytoskeleton (Roy and Lohia, 2004; Meza et al., 2006), and a simplified mitochondrion (i.e., a mitosome; Makiuchi and Nozaki, 2014).

Five of the ancestral paralogs functionally linked to the endomembrane system (based on data from model eukaryotes) show various degrees of patchiness in their occurrence (Figure 3A; Supplementary table 2). Arl1, Arl5, and Arfrp1, all associated with the Golgi apparatus, have been preserved in all main eukaryote lineages sampled, but have been lost from some more terminal branches. Arl1 is missing from the fission yeast (*Schizosaccharomyces pombe*), diplomonads, and some apicomplexans. Arfrp1 is absent from the same set of species plus two more (the highly reduced endosymbiotic kinetoplastid *Perkinsela* sp. CCAP 1560/4 and the tiny green alga *Micromonas commoda*). The similar patterns of loss of these two GTPases may reflect the fact that they were shown to work in the same functional cascade (see above). How Arl1 functions in the absence of Arfrp1 in *Perkinsela* or *Micromonas* remains an open question.

Arl5 is missing from many more eukaryotes, including even some metazoans (e.g., the flatworm *Schmidtea mediterranea*). A minimum of 20 independent losses of Arl5 is required to explain its distribution in our dataset (Supplementary table 2), suggesting that this GTPase is a less critical component of the basic infrastructure of the eukaryotic cell. In accord, disruption of the Arl5 gene in *Drosophila melanogaster* does not alter the fly's viability or fertility (Rosa-Ferreira et al., 2015). Arl5 is closely related to Arl1 and the two GTPases may share some effectors (see above). It is thus possible that Arl5 loss is facilitated by partial functional redundancy with Arl1. The presence of neither Arl1 nor Arl5 correlates strongly with stacked versus unstacked Golgi structure (Supplementary table 2). However, we did note a correlation between the presence of Arl5 and GRIP domain-containing proteins that act in the tethering of incoming vesicles at the Golgi (Cheung and Pfeffer, 2016) (Supplementary table 2). Further investigation of the functional overlap and interactions between Arl1, Arl5 and GRIP-proteins may be fruitful to understand their cellular roles. This in turn may help explain why both proteins are retained in so many lineages, and why the two ancient paralogs may have evolved in the first place.

Similar to Arl5, the distribution of Arl8 in extant eukaryotes has been shaped by multiple (at least 14) independent losses, including one in the lineage leading to the main eukaryotic taxon Stramenopiles (Figure 3B; Supplementary table 2). Comparison of phylogenetic profiles of Arl8 and the related uncharacterized paralog Arl18 reveals that the former paralog has been retained more frequently than the latter, but in a few taxa (e.g., stramenopiles) Arl18 occurs in the absence of Arl8 (Figure 3A; Supplementary table 2). It

would be interesting to investigate whether a level of functional redundancy might allow Arl18 to have taken over some of the Arl8 functions in these organisms. The presence of both Arl8 and Arl18 in model systems like *Tetrahymena thermophila* and *Trypanosoma cruzi* (Supplementary table 2) provides a promise that functional dissection of these closely related paralogues is possible.

The patchy distribution of Arf6 is somewhat surprising, at least in part because it contrasts with the near universal distribution of Arf1 paralogs. While Arf6 is perhaps most commonly associated with endocytosis and plasma membrane dynamics (see above) we speculate that perhaps it is its role in pericentriolar localization of specific subsets of recycling endosomes that are required for midbody formation and abscission (Fielding et al., 2005; Wilson et al., 2005; Turn et al., 2020) that vary amongst species. The nature and composition of centrioles, as well as associated components are known to vary, including losses or differences in Archaeplastida and SAR (Nabais et al., 2020).

The unexpected discovery of the sporadically distributed, yet potentially ancestral Sar1L paralog (Figure 1; Figure 3A; Supplementary table 2) raises an interesting possibility of a specific elaboration of the ER function in the LECA lost for some reason(s) by most major eukaryotic groups. Direct functional characterization of Sar1L in suitable model organisms is necessary before the causes behind the retention/loss pattern of the gene may be understood. However, it is interesting to compare Sar1L with the recently uncovered complexity of the ancestral set of paralogs of the COPII coat complex, including the Sec24III paralog as patchily distributed as Sar1L (Schlacht and Dacks, 2015). The phylogenetic profiles of Sar1L and Sec24III do not overlap well (e.g., Sar1L is missing from Chloroplastida and Sec24III is absent from diatoms), so we are not suggesting a specific functional link between these two proteins. Nevertheless, the existence of both proteins implies the existence of an interesting degree of variation in the COPII vesicle formation at the ER in different eukaryotes.

A rare, robust example of horizontal transfer of a Ras superfamily gene

The newly recognized functionally uncharacterized Arl17 group of ARF family protein is unusual not only because of its unique domain architecture (Figure 2), but also due to its very patchy taxonomic distributions (Figure 3A; Supplementary table 2). Based on our current sampling, Arl17 is completely missing from several major eukaryotic clades (Malawimonadida, Metamonada, Discoba, Stramenopiles, Haptophyta, and Rhodophyta), whereas its occurrence in the other groups is typically sporadic. Particularly interesting is

identification of a group of four closely related Arl17 homologs in the rotifer *A. vaga* (Figure 2; Supplementary table 1), which is the sole representative of the densely sampled Holozoa clade possessing Arl17 (Supplementary table 2). Transcriptome data from *A. vaga* relatives indicate that Arl17 is not restricted to a single rotifer species (data not shown), ruling out contamination in the *A. vaga* genome data. Hence, the isolated occurrence of Arl17 in a rotifer lineage strongly indicates gain via horizontal gene transfer (HGT) from a protist or fungal lineage, with subsequent gene duplications (at least partly accounted for by tetraploidy of the *A. vaga* genome, see above). Indeed, analyses of rotifer genomes revealed propensity of these peculiar microscopic animals for gene gain from various sources, and three of the four *A. vaga* Arl17 paralogs were included in the list of HGT candidates in the *A. vaga* genome (Flot et al., 2013). To our knowledge, this is the first convincing case of a eukaryote-to-eukaryote HGT in the whole Ras superfamily. Even though phylogenetic analysis of the Arl17 GTPase domain did not shed light on the origin of rotifer's Arl17 (Figure 2), a specific relationship to Arl17 proteins from *Physarum polycephalum* is suggested by a phylogeny inferred for the different copies of the C-terminal novel domains (Supplementary figure 10), suggesting that rotifers acquired Arl17 from an amoebozoan.

Expansion of the ARF family in Holozoa

Given the prominent position of metazoan model systems (humans, *M. musculus, D. melanogaster*, *Caenorhabditis elegans*) in research on the ARF family, we carried out a separate analysis concentrating on the family members in widely sampled representatives of Metazoa and their closest protist relatives, together constituting the taxon called Holozoa. Analogously to our eukaryote-scale ScrollSaw analysis described above, we compared 18 groups of sequences corresponding to the main holozoan lineages (phyla). This approach narrowed our original holozoan dataset of nearly 550 sequences to ~320 sequences and phylogenetic analysis of this reduced dataset revealed a set of strongly supported clades that provided a basis for defining ARF family paralogs conserved across the main holozoan or metazoan lineages (Figure 4). All ancient eukaryotic paralogs represented in this taxon, except Arf1, form supported clades (note that Arl17 failed to pass the ScrollSaw step as it is present only in rotifers). Furthermore, six additional groups could be identified based on this analysis, namely Arf4 (class II Arf), Arl4, 10, 15, 19 and TRIM23. Most of them are named according to previously annotated vertebrate genes (Gillingham and Munro, 2007). An exception is a novel group, here named Arl19, which is not a resolved clade, but seems to represent a coherent evolutionary lineage based on additional evidence (see below). Analysis

of intron positions in a subgroup of ARF family genes corresponding to Arfs and their closest relatives supported the delimitation of the main groups, but also suggested that several sequences initially annotated as Arf1 (based on BLAST searches) may constitute a novel conserved group in unicellular holozoans and several invertebrate lineages (Supplementary figure 11). Specifically, this group is characterized by three unique intron positions, and a focused phylogenetic analysis supported its monophyly and separation from Arf1 and other clades (Supplementary figure 12A). We thus named this novel clade Arl20.

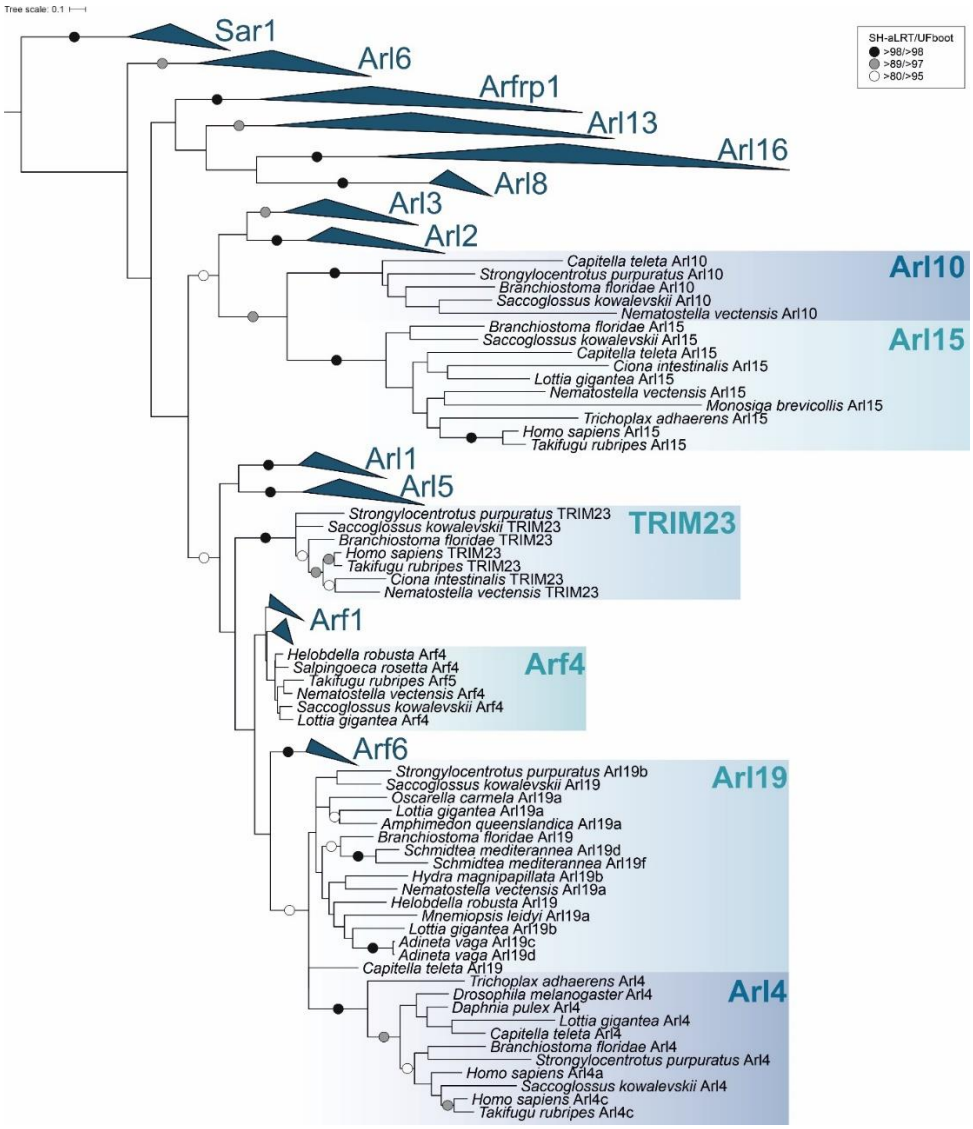

Figure 4. Maximum likelihood phylogenetic tree of the ARF family based on a ScrollSaw dataset in Holozoa. The tree was inferred using IQ-TREE with LG+I+G4 model (the model selected by the program itself) with the ultrafast bootstrap algorithm and the SH-aLRT test (both 10000 replicates), as described under Materials and Methods. Dots at branches represent bootstrap values as indicated in the legend shown in the bottom left. Eukaryotic ancestral paralogs are collapsed as triangles.

Establishment of novel ARF lineages provided a basis for the assignment of sequences excluded by the ScrollSaw protocol by the same approach as described for the ancient eukaryotic paralogs. Moreover, inspection of the exon-intron structures facilitated assignment of some of the problematic genes. For example, *Takifugu rubripes* harbours several standard Arfs of classes I, II and III and one additional Arf-like paralog (TruArfX in Supplementary table 1) with an almost equal similarity to the classes I and II. Both phylogenetic and HMMER-based analyses were inconclusive concerning the origin of this gene, but the exon-intron structure of TruArfX exhibits the pattern typical to the class II Arfs (Supplementary figure 11), supporting annotation of this gene as a divergent representative of the Arf4 group. Combining such different forms of evidence allowed us to annotate the majority of sequences, to establish the phylogenetic distribution of the main groups, and to map their origins and losses onto the holozoan phylogeny (Figure 5; Supplementary table 2).

Altogether we could recognise seven groups that apparently originated after the split of the holozoan lineage from their relatives (Holomycota), that is in the holozoan stem itself (Arf4), at a later step but still before the divergence of Metazoa and their sister group choanoflagellates (Arl15, 19, 20), in the metazoan stem (Arl10), or after the divergence of the deepest metazoan phyla (Arl4, TRIM23). This stepwise build-up of complexity of the ARF family (Figure 5B) contrasts with a somewhat different evolutionary pattern documented for the Rab family, which experienced a wave of expansion concentrated in the metazoan stem lineage (Elias et al., 2012). The novel ARF family members in Holozoa apparently emerged by modification of duplicated copies of specific ancient eukaryotic paralogs, although the exact sources may be difficult to determine. Sequence similarity and some phylogenetic analyses (Figure 4) point to the Arl2/3 clade as the most likely cradle of Arl10 and 15, but the position of these two paralogs is unstable in different phylogenies (Supplementary figure 12B). Evidence is more solid for the origin of Arf4, Arl4, 19, 20 and TRIM23, suggesting these are offshoots stemming from Arf1/6-like ancestors (Figure 4; Supplementary figure 12).

A common origin of Arf1 and Arf4 was already reported (Li et al., 2004; Manolea et al., 2010), but our analysis placed this event before the divergence of ichthyosporeans (i.e., to the common ancestor of Holozoa; Figure 5B), which probably possessed Arf1, Arf4, and Arf6 as single-copy genes. While Arf4 and Arf6 seem to duplicate only sporadically, Arf1 is often present in more than one copy, suggesting a high propensity for duplication; this tendency is in fact seen for eukaryote lineage in general (Supplementary table 2). Phylogenetic analyses usually do not recover Arf1 and Arf4 as supported monophyletic

clades (e.g., Figure 4), which is probably a result of their high sequence similarity reflected also in partial functional overlap of Arf1 and Arf4 (Jackson, 2014; Jackson and Bouvet, 2014). However, their separation is obvious from the comparison of the exon-intron structures of the respective genes (Supplementary figure 11).

Two more holozoan or metazoan GTPase groups are likely evolutionarily derived from the ancestral Arf1 gene, yet have diverged to the point it seems inappropriate to call them "Arfs". One is Arl20, a previously unrecognized group of genes sharing three specific intron positions (Supplementary figure 11). Their relationship to Arf1 cannot be conclusively inferred from our phylogenetic analysis (Supplementary figure 12A), but an intron position shared with Arf1 (and Arf4) and outcomes of similarity searches support this hypothesis. TRIM23 (also called ARD1) is an unusual protein including not only the GTPase domain, but also a block of domains characteristic for the TRIM family (RING-type E3 ubiquitin ligase, a tandem of BBbox domains, and the BBC domain forming a coiled-coil) at the N-terminus. The GTPase domain is highly similar to true Arfs (Vichi et al., 2005) and its specific relationship to Arf1 is obvious from the virtually identical exon-intron structure (of the gene part encoding the GTPase domain; Supplementary figure 11).

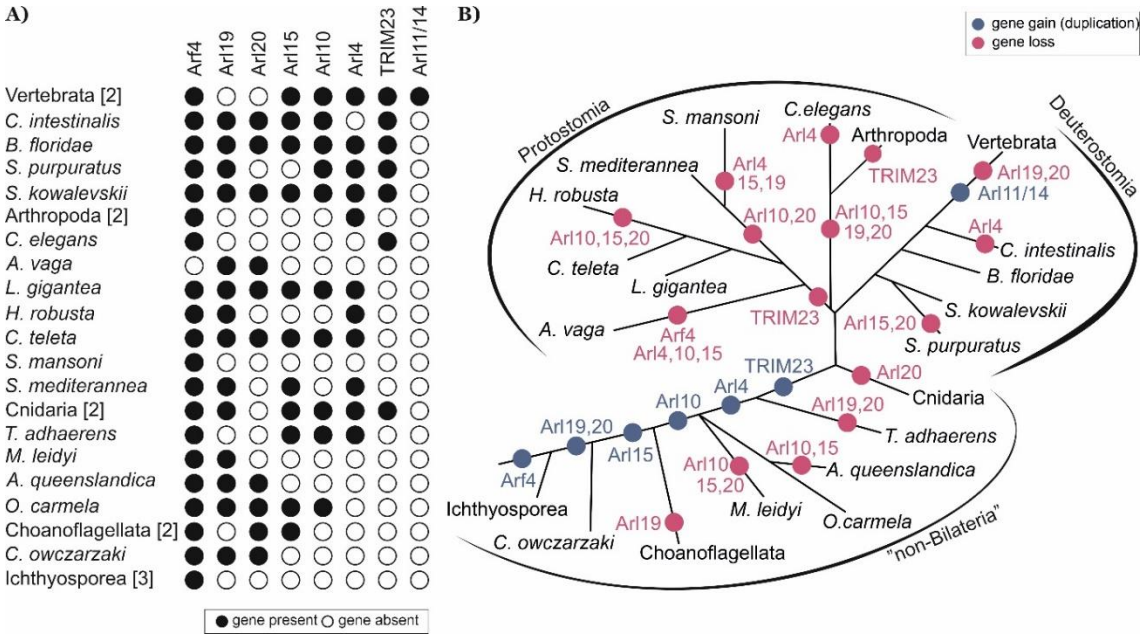

Figure 5. Retention of lineage-specific paralogs of the ARF family in main lineages of Holozoa. (**A**) Black circle: the paralog is present in at least one member of the lineage; white circle: the gene is absent from the lineage (evidenced by genome sequence data). Species with identical distribution are collapsed into higher taxa with the number of species indicated in the square brackets. (**B**) Gene gains (blue circles) and losses (pink circles) mapped onto the holozoan phylogeny.

Arl4 and the Arl19 group newly recognized here constitute an unsupported sister group to Arf6 in our trees (Figure 4). While Arl4 forms a highly supported monophyletic group, its placement disrupts the monophyly of Arl19, perhaps due to an insufficient phylogenetic signal that would unite all Arl19 sequences in the analyses. The origin of Arl4 and Arl19 from Arf6 is conceivable and there are also potential functional links between Arl4 and Arf6; e.g., mammalian Arl4 proteins can recruit the Arf6 GEFs cytohesins to the plasma membrane (Hofmann et al., 2007) and each GTPase can influence actin dynamics (Cotton et al., 2007; Li et al., 2007; Patel et al., 2011). The exon-intron structure of Arl4 and Arl19 are not helpful in unveiling their origin. Only a minority of Arl4 genes contain introns, the intron positions are not conserved between Arl4 genes, and do not match the rest of examined Arf genes (Supplementary figure 11). This suggests that Arl4 may have originated through retroposition (Kaessmann et al., 2009), that is by integration of a reverse-transcribed mRNA into the genome of an early metazoan, with the few non-conserved introns gained secondarily and independently in different metazoan lineages. The exon-intron structure of Arl19 is rather puzzling, as several genes share an intron with Arf1 (Supplementary figure 11), but the whole clade branches off close to Arf6 (Figure 4).

In addition to the aforementioned novel ARF family members broadly conserved across Holozoa or Metazoa, various metazoan lineages exhibit still other novelties suggesting further functional elaboration. Here we focus on vertebrates. First, the vertebrate ARF family complement has been expanded by duplications of Arf1 and Arf4, yielding the well-known two groups of paralogs (Arf1, 2, 3 versus Arf4 and 5). Together with multiple duplications of Arl4, vertebrates are thus endowed with a battery of lineage-specific paralogs that are generally highly similar in sequence and (presumably) function (Supplementary tables 1 and 2). Second, vertebrates have experienced duplication of the Arl10 gene inherited from their invertebrate ancestor, giving rise to two in-paralogs that diverged from each other to such an extent that they were not initially recognized as closely related and which is reflected in their different names: Arl9 and Arl10 (Supplementary figure 12B). Finally, vertebrates encode two divergent ARF family members of a common origin, called Arl11 and Arl14, that seems to have evolved by duplication and divergence from Arl4 (Figure 5; Supplementary figure 13). The functional significance of these novelties is unclear, owing to limited knowledge of the function of the respective proteins in any vertebrate species including humans. It is, however, important to stress that the vertebrate ARF family complement has been sculpted also by gene loss, as vertebrates (in contrast to their sister group tunicates represented in this study by *Ciona intestinalis*) lack Arl19 and Arl20 (Figure 5).

The emergence of other major eukaryotic clades was accompanied by limited evolutionary novelty in the ARF family

Given the identification of multiple novel ARF family paralogs in Holozoa/Metazoa, we also applied the ScrollSaw protocol to other eukaryote groups to uncover possible lineage-specific innovations. Interestingly, while gene duplications specific to terminal organismal lineages are common in the ARF family, only three higher-level taxa – rhodophytes, glaucophytes, and Chloroplastida – seem to have evolved novel family members by gene duplication in their stem lineages (Figure 3B). The genome of red algal ancestors underwent massive reductive evolution (Yoon et al., 2017), which is reflected also by their highly reduced set of Rab GTPases (Petrželková and Eliáš, 2014) as well as of ARF family proteins (Figure 3; Supplementary table 2). Somewhat opposite to this trend, a novel ARF family member, here denoted ArlRhodo, is shared by distantly related rhodophyte taxa and apparently emerged before the radiation of the whole group. Their origin remains elusive, as the phylogenetic analysis placed ArlRhodo as a separate clade of the ARF family with no specific affinities to any of the ancestral clades (Supplementary figure 14). By contrast, the glaucophyte innovation, in fact represented by multiple paralogs in individual glaucophyte species, can clearly be traced as a highly divergent offshoot of Arl13 (Supplementary figure 15).

The only previously documented innovation of the ARF family specific for a major eukaryotic group other than metazoans is the plant ArfB (Vernoud et al., 2003). It was proposed to be an Arf6 ortholog (Li et al., 2004), and indeed our phylogenetic analysis places ArfB as sister group to Arf6. However, this topology is not statistically supported and can be an artefact resulting from the apparently rapid initial evolution of the ArfB gene reflected by the long stem branch subtending the ArfB subtree (Supplementary figure 16). Moreover, ArfB genes share one intron position with Arf1, but none with Arf6 (Supplementary figure 4). Hence, we leave the origin of the ArfB group as unresolved for the time being. This not withstanding, the timing of the ArfB emergence coincides with a duplication of the ARF GEF BIG in the Chloroplastida (Pipaliya et al., 2019). The duplication of the ArfB paralogs in embryophytes also coincides with the duplication of GBF1 proteins in that same lineage. As both of these GEFs act on Arf1-derived paralogs in metazoans at least, this lends itself to the hypothesis that ArfB is derived from Arf1. It raises the hypothesis that the one of the BIG duplicates acts specifically on ArfB in green algae. It also means that the ArfB and BIG and GBF1 duplicates should all be included in any GTPase activity assays aimed at understanding how this network functions in plant cells.

Extensive molecular tinkering in the evolution of membrane attachment mechanisms in the ARF family

It is currently understood that a large fraction of ARF family members act within endomembrane traffic pathways through their actions on the surface of source membranes (Gillingham and Munro, 2007). This necessitates specific, and (typically) transient, membrane attachment, typically relying on specific PTMs, employed by different ARF family members. Our analyses illuminate the origins of the previously described means of membrane association, but also extend their diversity to mechanisms previously unknown in the family (summarized in Figure 6).

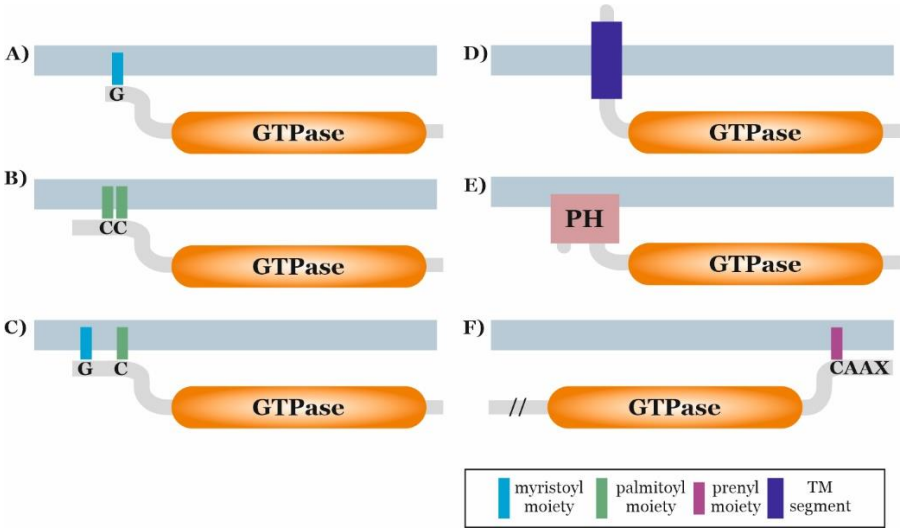

Figure 6. Membrane attachment mechanisms of ARF family proteins. Examples of different broadly conserved mechanisms of membrane attachments of ARF family members are depicted. (**A**) N-terminally myristoylated glycine residues, common for Arfs and several Arf-like proteins. (**B**) One or two S-palmitoylated cysteine residues near the N-terminus, typical for Arl16 and also common in Arl13. (**C**) N-terminally myristoylated glycine residue coupled with S-palmitoylated cysteine residue near the N-terminus, typical for ArlRhodo. (**D**) N-terminal transmembrane region, typical for SRβ and Arl10. (**E**) N-terminally accreted PH domain, present in divergent Arf-like proteins in kinetoplastids and choanoflagellates. (**F**) Prenylation motif (CaaX) at the C-terminus of certain eustigmatophyte-specific ARF family members (characterized also by a long N-terminal extension, in the figure marked with "//"). Supplementary table 1 lists all identified ARF family proteins predicted to be N-myristoylated or S-palmitoylated, or to contain a transmembrane region or PH domain.

N-terminal myristoylation (N-myristoylation) is the most common lipid modification mediating the reversible membrane attachment of ARF family proteins (Kahn et al., 1988; Liu et al., 2009). Two necessary prerequisites for N-myristoylation are the glycine residue at

the second position of the protein and specific sequence motif downstream that is recognised by the myristoyl transferase catalysing the addition of the myristate moiety to the N-terminal glycine (Duronio et al., 1991; Resh, 1999). Once acted upon by N-myristoyl transferase, the myristate group is attached through an amide bond that is permanent for the life of the protein. Reversibility in membrane association is tightly linked to the activation status of the ARF family protein, as the myristoylated N-terminal α-helix is accommodated in a hydrophobic channel when the protein is inactive (GDP-bound) but becomes solvent exposed in response to activation (GTP-binding), resulting in its propensity to bury the freed myristate in a lipid bilayer (Pasqualato et al., 2002; Seidel et al., 2004; Liu et al., 2009; Liu et al., 2010).

Using dedicated bioinformatic tools (see Materials and Methods), we predicted this post-translational modification for the majority of the proteins representing the ancestral eukaryotic paralogs Arf1, Arf6, Arl1, and Arl5 (Table 1; Supplementary table 1; Supplementary table 4), in keeping with previous experimental data from yeast and mammalian proteins (Kahn et al., 1988; D'Souza-Schorey and Stahl, 1995; Lee et al., 1997; Lin et al., 2002). Virtually all Arf6, Arl1 and Arl5 proteins possess the conserved glycine residue at the second position, and the negligible minority of those not predicted as N-myristoylation targets may be false negatives. From almost 450 Arf1 genes investigated, 40 do not possess the expected glycine residue and cannot be modified by myristoylation in a standard manner. We note that a recent study found N-myristoyltransferase capable of acylating lysine in the third position (Dian et al., 2020), though the predicting algorithms employed here did not include this possibility. All of these 40 Arf1 genes are accompanied by two or more Arf1 genes that are N-myristoylated in the given organism (Supplementary table 1), so they apparently represent lineage-specific paralogs with a changed behaviour towards membranes. The newly recognised Arl18 paralog, though not closely related to the previous four paralogs, also seems to be ancestrally myristoylated, as all genes contain a glycine residue at the second position and the majority of them are predicted as N-myristoylated (Table 1; Supplementary table 1; Supplementary table 4). Interestingly, the Arl18 sister group Arl8 seems to ancestrally lack glycine at the second position (Supplementary table 1; Table 1) and the only putatively N-myristoylated Arl8 can be found in rhizarians, suggesting secondary acquisition of the myristoylation motif in this lineage. The majority of Arl2 and Arl3 proteins do harbour a glycine residue at the second position, but N-terminal myristoylation is predicted only for a few Arl3 proteins (Table 1; Supplementary table 1; Supplementary table 4) and these may be false positives, considering

the experimental evidence for the lack of N-myristoylation in representative Arl3 proteins (Sharer et al., 2002; Setty et al., 2004). The Arl6 group is clearly heterogeneous, including members that certainly are not myristoylated as well as members that likely have this modification. Thus, the evolutionary course leading to the distribution of N-myristoylation in different ARF family members is not always clear. One possibility is an early origin of this modification in an ancestor of all the clades with N-myristoylated members, followed by its multiple secondary losses. However, multiple independent acquisitions is certainly a likely, and mutually non-exclusive, alternative.

Ancient ARF family

	Seq	Gly2	Myr	Palm	TM
	∑	%			
Arf1	439	90.9	71.3	11.4	0.5
Arf6	64	98.4	82.8	4.7	0
Arl1	116	100	94.8	0.9	0
Arl2	115	91.3	0	1.7	0
Arl3	106	89.6	8.5	6.6	1.9
Arl5	82	100	78	14.6	0
Arl6	87	77	20.7	12.6	1.1
Arl8	106	21.7	5.7	0.9	0
Arl13	70	24.3	11.4	55.7	0
Arl16	50	6	2	54	2
Arl17	26	26.9	11.5	15.4	0
Arl18	28	100	60.7	21.4	0
Arfrp1	112	1.8	0.9	9.8	0
Sar1	160	8.8	0	1.9	0.6
Sar1L	12	0	0	0	0
SRβ	117	4.3	0.9	20.5	100

Holozoa lineages-specific ARF family

	Seq	Gly2	Myr	Palm	TM
	∑	%			
Arf4	30	100	100	0	0
Arl4	20	95	60	5	0
Arl10	15	13.3	0	20.0	93.3
Arl11/14	5	100	100	40	0
Arl15	14	14.3	7.1	92.9	0
Arl19	42	100	78.6	19	0
Arl20	19	89.5	36.8	5	0
TRIM23	9	0	0	33.3	0

Archaeplastida lineages-specific ARF family

	Seq	Gly2	Myr	Palm	TM
	∑	%			
Chloroplastida ArfB	16	100	25	0	0
Rhodophyta ArlRhodo	8	100	100	62.5	12.5
Glaucophyta Arl13L	6	16.7	0	0	0

Table 1. Acylation and the presence of a transmembrane region in the ARF family. The table provides a summary of the results of prediction of N-myristoylation, S-palmitoylation and presence of the transmembrane (TM) region in particular subgroups of the ARF family. For each subgroup (group of orthologs), the number of sequences (Seq) and the percentages of sequences with glycine residues at the second position (Gly2), sequences predicted as N-myristoylated (Myr), sequences predicted as S-palmitoylated on at least one cysteine residue (Palm), and sequences with predicted transmembrane region (TM) are given. Values above 50% are highlighted in pink. For complete data see Table S1 and Table S4. These predictions were made as described under Materials and Methods.

S-palmitoylation (i.e., addition of a palmitoyl moiety to one or more cysteine residues) also mediates protein association with membranes, though unlike N-myristoylation there are enzymes capable of reversing this acylation making it a more transient modification (Zhou and Cox, 2014). We again employed a suite of dedicated algorithms to predict the presence of this modification in ARF family members, as described under Materials and Methods. Arl15 proteins typically harbour several N-terminal cysteine residues, usually predicted as S-palmitoylated (Supplementary figure 17), and approximately half of the Arl13 and Arl16 sequences analysed also contain one or more putative S-palmitoylated cysteine residues in their N-terminal region (Table 1; Supplementary table 1; Supplementary table 4). S-palmitoylation of Arl13 from *C. elegans* and mammals has been confirmed experimentally and demonstrated as crucial not only for the proper localization of the proteins, but also for stability and function (Cevik et al., 2010; Roy et al., 2017). In a few cases, such as in the red algae-specific paralog ArlRhodo, S-palmitoylation seems to accompany N-myristoylation (Supplementary table 1, Table 1), similar to various other proteins, including GTPases (e.g., some Gα proteins; Zhou and Cox, 2014).

In addition to employing covalently attached saturated fatty acids, proteins also can be permanently (absent proteolytic cleavage) anchored in the membrane via a transmembrane domain. Of the proteins investigated here, this was previously demonstrated for SRβ, a protein anchored in the ER membrane via its N-terminal transmembrane region (Keenan et al., 2001) that appears to be conserved in all SRβ sequences investigated (Table 1). An N-terminal transmembrane region was independently acquired by the Metazoa-specific Arl10 (see above) and several other ARF family members in various eukaryotes (Table 1; Supplementary table 1). In some cases, we could confirm conservation of such putative N-terminally anchored GTPases in a broader organism clade beyond the species primarily targeted by our analysis, as is the case of divergent putative Arf1 paralogs from *Bigelowiella natans* and other chlorarachniophytes (Supplementary figure 18A) and from *Pavlova pinguis* and other haptophytes of the class Pavlovophyceae (Supplementary figure 18B). Another mode of membrane attachment utilized by some ARF family members is accretion of specific membrane-binding domains (Figure 6). This is exemplified by unusual proteins from choanoflagellates and trypanosomatids that contain an N-terminal phosphoinositide-binding PH domain (Lemmon, 2007) connected to the ARF family GTPase domain by a long linker region (Figure 6; Supplementary table 1). Finally, the eustigmatophyte *Vischeria* sp. encodes a unique ARF family protein with a long N-terminal extension lacking any detectable

conserved protein domain or functional motif and with a C-terminal tail ending with the amino acid sequence CSIM (Figure 6; Supplementary table 1), which is reminiscent of the so-called CaaX motif (or box) directing prenylation of the cysteine residue in diverse proteins (Fu and Casey, 1999). A similar protein, including this motif, is encoded by additional eustigmatophytes (not shown), and two different prediction programs proposed the cysteine residue to be prenylated (see Material and Methods for details). C-terminal prenylation is a common modification ensuring membrane attachment of GTPases belonging to Rab, Ras and Rho families (Zhou and Cox, 2014), but to our knowledge it has not been reported previously for an ARF family protein.

The well-studied mammalian members of the ARF family are subject to other post-translational modifications (e.g., see Phosphosite Plus; https://www.phosphosite.org/), though these either lack consensus motifs that prevent predicting their existence in other organisms or have no known functional consequences, or both. One exception to this is N-terminal acetylation of Arl8, which has been shown to be important for its association with lysosomal membranes (Hofmann and Munro, 2006). Similarly, in *S. cerevisiae* the Arfrp1 protein (unfortunately named Arl3p only in this organism) is also acetylated and this is required for its association with Golgi membranes (Behnia et al., 2004). Future development of appropriate prediction tools, perhaps combined with dedicated biochemical investigations, will be instrumental in grasping the full breath and evolutionary conservation of PTMs in the ARF family.

In summary, the use of several different means of membrane attachment is consistent with ARF family proteins acting predominantly on a membrane surface, and the diversity of various membrane attachment mechanisms exhibited by this family is surprisingly extensive and reminiscent of what has been described for the distantly related GTPase Rheb (Záhonová et al., 2018).

Extensive diversity of multi-domain ARF family members

The existence of the PH domain-containing ARF family proteins or the aforementioned multi-domain TRIM23 protein (Vichi et al., 2005) counter the paradigm of ARF family members being limited to single (GTPase) domain proteins with only short N- and C-terminal extensions. In fact, our analyses challenge this dogma further, as we observe a much greater number of different protein architectures involving combinations of the GTPase domain of the ARF family with other functional domains.

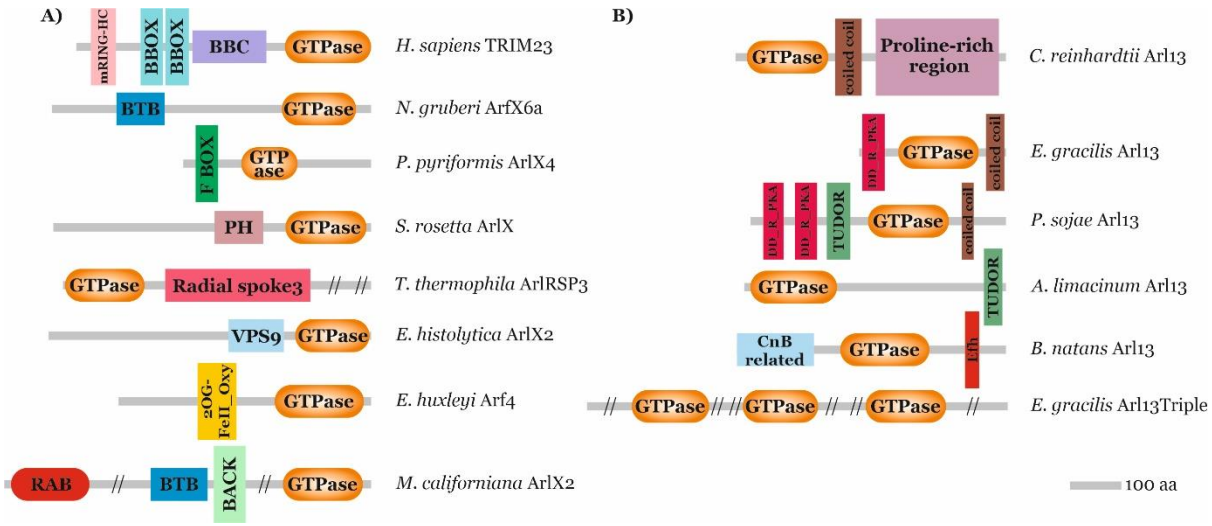

Figure 7. Multi-domain architectures of ARF family proteins. (**A**) Examples of lineage-specific ARF family proteins with extra domains accreted to the GTPase domain. Sequence IDs of the proteins listed are provided in Supplementary table 1. (**B**) Variation in the domain architecture of Arl13 proteins across the eukaryote diversity. The Arl13 from *Chlamydomonas reinhardtii* represents the most common and presumably ancestral state.

The novel, presumably ancestral eukaryotic, Arl17 group characterized by combining an Arf-related domain with varying numbers of tandemly arrayed copies of a novel uncharacterized domain (Figure 2) was introduced above. Additional domain architectures are found in proteins that generally seem to be lineage-specific innovations restricted to particular taxa; some examples are provided in Figure 7. Similar to TRIM23, some include domains linked to ubiquitination, namely the BTB domain or the F-box domain (see Genschik et al., 2013), indicating recurrent recruitment of ARF family members into ubiquitin-dependent regulatory circuits. Ciliates exhibit a unique protein with an ARF family GTPase domain fused to a segment homologous to radial spoke protein 3 (RSP3), a component of radial spokes in the axoneme (see Wirschell et al., 2008). This predicts ciliary localization of this protein, and indeed, it is among the proteins detected in the ciliary proteome of *Tetrahymena thermophila* (Smith et al., 2005). *Entamoeba histolytica* possesses a protein with a divergent C-terminal ARF family domain preceded by the VPS9 domain. The latter domain is known to act as a GEF of the endosomal Rab GTPase Rab5 (Ishida et al., 2016), so this protein may be part of a pathway with multiple sequentially acting GTPases similar to regulatory GTPase cascades known from mammalian or yeast cells (Jones et al., 1999; Mizuno-Yamasaki et al., 2012). Another unique domain combination occurs in one of the Arf paralogs in the haptophyte *Emiliania huxleyi*, which is fused to the C-terminus of a

block including a domain of the 2OG-Fe(II) oxygenase superfamily. It is possible that the GTPase domain regulates the enzyme activity of the N-terminal part of the protein. The ARF family domain can combine also with other Ras superfamily GTPase domains, as demonstrated by a protein from *Malawimonas californiana* with an N-terminal Rab domain and a C-terminal Arf domain linked by a region containing detectable BTB and BACK domains (Figure 7A).

Tinkering with protein domains in ARF family proteins can be encountered in a different evolutionary context than the emergence of lineage-specific paralogs. In the case of Arl13, domains were acquired or lost without gene duplication, resulting in differences in domain architectures between orthologous Arl13 genes. Previously characterized orthologs from mammals and *Chlamydomonas reinhardtii* exhibit a poorly conserved C-terminal extension that includes a region forming a coiled-coil followed by a proline-rich region (Hori et al., 2008; Miertzschke et al., 2014; Figure 7B). Inspection of the large collection of Arl13 sequences amassed for this study revealed that this arrangement is distributed broadly across the eukaryote phylogeny and likely ancestral. However, some species depart in various way from this structure, e.g. by lacking the proline-rich region or the coiled-coil. Recently, Zhang et al. (2018) identified a non-canonical Arl13 gene from *Trypanosoma brucei* containing the DD_RI_PKA domain (Dimerization/Docking domain of the Regulatory subunit of protein kinase A (PKA)) that is essential for targeting of *T. brucei* Arl13 to the cilium. Our analysis revealed that the same protein architecture is present also in *Euglena gracilis*, suggesting it is a synapomorphic character for the whole Euglenozoa phylum (Figure 7B). Meanwhile, a subset of Stramenopiles (oomycetes and ochrophytes) independently acquired DD_RI_PKA domain as two tandemly arrayed copies (Figure 7B). DD_RI_PKA mediates interaction of PKA with A-kinase-anchoring proteins (AKAPs), which regulate PKA localization in the cell (Sarma et al., 2010). Given the ciliary function of Arl13 (see above), we speculate that the DD_RI_PKA domain in some Arl13 proteins interacts with a cilium-localized AKAP, such as the aforementioned RSP3 protein (Gaillard et al., 2001; Jivan et al., 2009). In contrast, mammalian Arl13b contains the simpler VxP motif in the large C-terminal domain that is required for ciliary localization (Higginbotham et al., 2012; Cevik et al., 2013; Gigante et al., 2020). DD_RI_PKA domains in *Phytophthora sojae* Arl13 are followed by the TUDOR domain, known for the ability to bind to the methylated lysine and/or arginine residues (Botuyan and Mer, 2016). The TUDOR domain was independently accreted also to the C-terminus of the Arl13 from *Aurantiochytrium limacinum* (Figure 7B). Another notable variant is encountered in Arl13 from *B. natans* (Figure 7B) and other chlorarachniophytes

(Supplementary figure 19), which exhibit a novel form of the C-terminal extension including the Ca^{2+}-binding EF-hand motif. Interestingly, the N-terminus of chlorarachniophyte Arl13 proteins appears to be related to calcineurin B, a Ca^{2+}-binding regulatory subunit of the protein phosphatase calcineurin (Guerini, 1997). It thus seems likely that Arl13 function is regulated by Ca^{2+} in chlorarachniophytes. Exceptional is an *E. gracilis* gene (co-occurring in this species with a typical Arl13 gene) that we named Arl13Triple, as it is composed of a tandem triplication of a divergent Arl13-reated GTPase domain (Figure 7B). The varying domain architecture of Arl13 in different eukaryotes points to a substantial degree of functional divergence of this key ciliary component.

Insights into the early radiation of the ARF family

In the analysis of protein family evolution, resolution between the sub-family paralogues is a tremendously informative result as it allows the inference of cellular evolution of the associated organellar compartments. However, such resolution has been difficult to obtain for many families. The ScrollSaw methodology was a step forward in obtaining resolution for datasets with many paralogues and short sequence length; e.g., Rabs and TBC proteins (Rab GAPs; Elias et al., 2012; Gabernet-Castello et al., 2013). Here, our application of the ScrollSaw methodology also yielded a partially resolved backbone topology (Figure 1). We observed the robust sisterhood of Arl8 and Arl18 and of these both to Arl16. We also observed the sisterhood of Arl2 and Arl3 plus the moderately supported node uniting Arf1 with Arf6. Most notably, there was a strongly resolved node separating the grouping of Arf1, Arf6, Arls 1, 2, 3 and 5 together, from the remainder of the sub-families.

This resolution provides the basis for several key inferences about the ancestral role of the ARF family progenitor and some implications about the role of these proteins during eukaryogenesis. What is clearly ancestral is a GTPase that changes conformation to relocate from the cytosol to a membrane and which binds other proteins as effector. Given the widespread role of ARF family members, this could mean a role in membrane-traffic. However, with at least one resolved node separating the best-known family members Arf1 and Sar1, a simple scenario of a single primordial GTPase that nucleates a primordial vesicle coat-forming complex is ruled out. This suggests that the proto-coatomer hypothesis (Devos et al., 2004) may well need to be modified to take a more complicated scenario, including possible convergence, parallel evolution, and even merging of architectures into account (Dacks and Robinson, 2017; Field and Rout, 2019).

ArfR is the archaeal GTPase group progenitor to the eukaryotic ARF family

One of the key, yet poorly understood, aspects of the ARF family evolution is its origin. A decade ago it was proposed that this eukaryotic family was founded by an archaebacterial representative of a broader prokaryotic GTPase group typified by MglA (Dong et al., 2007). However, this hypothesis was later challenged by realization that MglA-related proteins have a unique modification of the G3 motif of the GTPase domain, predicting a unique catalytic mechanism and making them unlikely ancestors of any branch of the eukaryotic Ras superfamily, which exhibit the canonical form of the G3 motif (Wuichet and Søgaard-Andersen, 2014). Most recently, an analysis of GTPases encoded in the metagenomically reconstructed genome of *Lokiarchaeum* sp. GC14_75, the first discovered member of the archaebacterial superphylum Asgard, revealed the presence of a group of genes noticeably similar to eukaryotic ARF family proteins (Klinger et al., 2016). These GTPases were denoted ArfL, but their direct relationship to the ARF family was not recovered by phylogenetic analysis.

The recent publication of reconstructed genomes from additional Asgard archaebacteria (Zaremba-Niedzwiedzka et al., 2017) enabled us to reinvestigate the origin of the eukaryotic ARF family. The new Asgard genomes proved to be similarly rich in Ras superfamily genes as the first lokiarchaeal genome (Supplementary table 5), and most of them could be assigned by phylogenetic analyses into the previously delineated groups including ArfL (data not shown). However, a group of the genes from representatives of Heimdallarchaeota and Lokiarchaeota proved to be phylogenetically closer to the eukaryotic ARF family, including SRβ proteins, than any previously analysed prokaryotic genes (Figure 8). Interestingly, this group included one gene (GenBank accession number KKK44255.1) from *Lokiarchaeum* sp. GC14_75 that was for unclear reasons excluded from the previous analysis of the Ras superfamily gene complement in this archaebacterium (Klinger et al., 2016), so its specific relationship to the eukaryotic proteins was missed. In addition, the new group includes one gene from *Candidatus* Korarchaeum cryptofilum, a member of the TACK superphylum sister to the Asgard clade that was sequenced years ago (Elkins et al., 2008). Again, this sequence was not included in previous analyses, so its significance has not been recognized. Finally, by searching metagenomic assemblies in the NCBI database with the *Ca.* K. cryptofilum sequence as a query we found three sequences that fall into the new archaebacterial GTPase group and closest to the korarchaeal genes, suggesting that they may come from other members of the TACK superphylum.

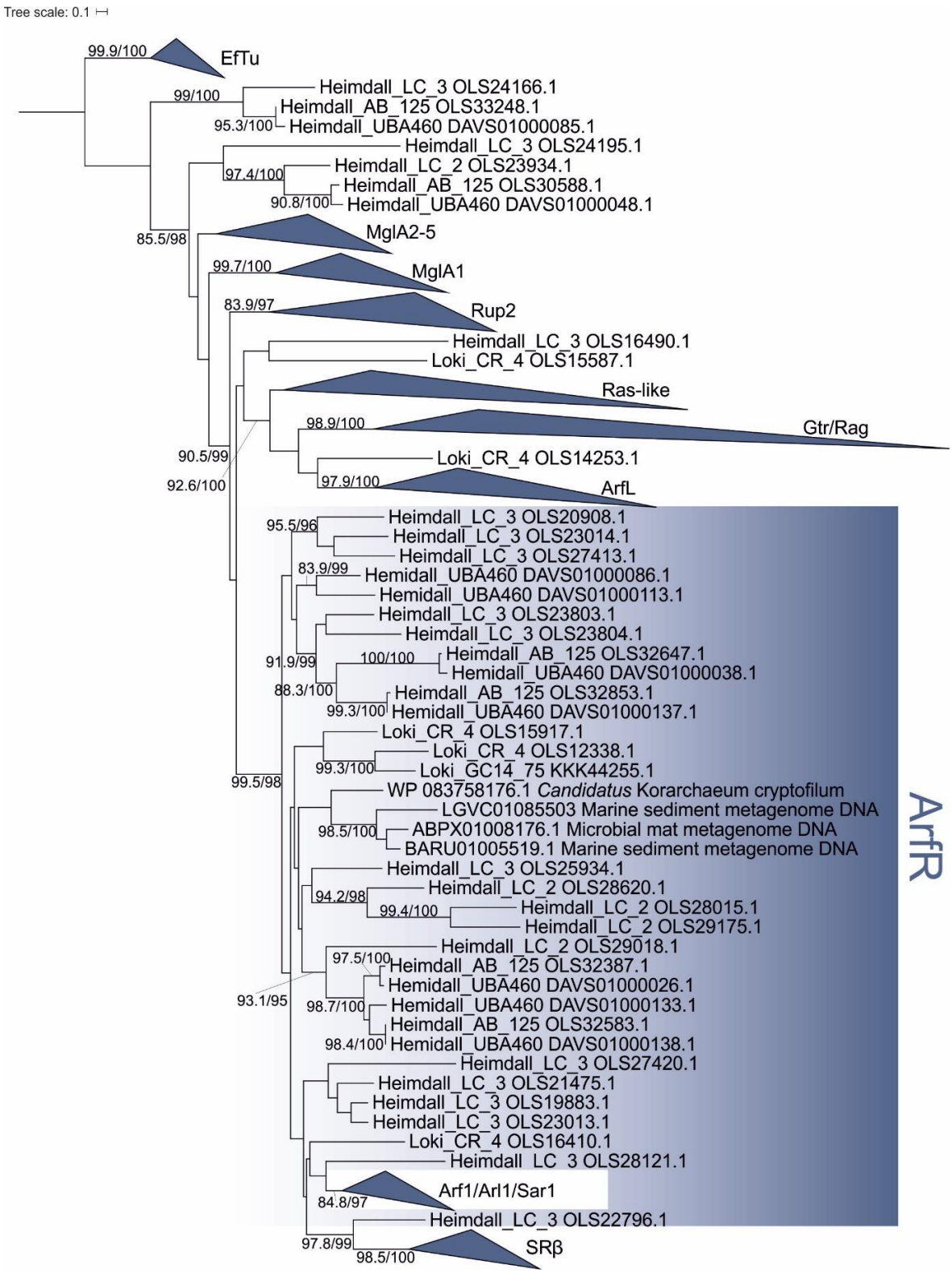

Figure 8. A new group of archaebacterial proteins specifically related to the ARF family. The tree was inferred using IQ-TREE with LG+F+I+G4 model (the model selected by the program itself) with the ultrafast bootstrap algorithm and the SH-aLRT test (both 10000 replicates). Numbers at branches correspond to SH-aLRT support values (right side of the slash) and ultrafast bootstrap support values (left side of the slash) and are shown when >80/95. The set of sequences included in the analysis was based on the sampling utilized previously by Klinger et al. (2016), which was expanded by addition of

Ras superfamily sequences from new genome assemblies from the Asgard superphylum (Zaremba-Niedzwiedzka et al., 2017; Parks et al., 2017) and from *Candidatus* Korarachaeum cryptofilum. The three sequences from metagenome assemblies at NCBI that were identified as best hits of the *Ca.* K. cryptofilum sequence WP_083758176.1 and their taxonomic provenance is unknown, although they most likely also come from archaebacteria. The genome assembly denoted "Heimdall_UBA460" was published as "archaeon UBA460" without providing classification into a specific archaebacterial group (see GenBank accession number DAVS00000000.1). Its affiliation to Heimdallarchaeota is indicated both by the Genome Taxonomy Database (GTDB; http://gtdb.ecogenomic.org/) and high similarity of its genes to the Heimdall_AB_125 genome. For the sake of simplicity, GTPase clades observed and defined in the previous study (Klinger et al., 2016) were collapsed. Note the novel group denoted "ArfR" (Arf-related), consisting of archaebacterial GTPases specifically related to eukaryotic Arf, Arl, Sar1 and SRβ proteins.

The existence of the GTPases in the TACK and Asgard superphyla affiliated to the eukaryotic ARF family with convincing statistical support suggest that not the previously identified ArfL group, but these newly discovered genes represent the actual archaebacterial stock from which the eukaryotic ARF family evolved. Interestingly, this new group of archaebacterial GTPases, which we propose to be called ArfR (Arf-related), exhibits sporadic distribution in archaebacterial phyla, being absent from all sequenced members of Cren-, Bathy-, Thaum-, Aig-, and Verstraetearchaeota in the TACK superphylum and from the genome data reconstructed so far for Thor- and Odinarchaeota in the Asgard superphylum. ArfR thus joins a growing list of "'eukaryotic signature proteins" that show a similarly patchy distribution across different archaebacterial lineages (Eme et al., 2017). In addition, the tree presented in Figure 8 suggests an intriguing possibility that the divergence between the "core" of the ARF family and the more loosely affiliated SRβ lineage occurred before the FECA, owing to the existence of a gene in a Heimdallarchaeota member that branches with reasonable support as a sister lineage of SRβ. More detailed analyses concentrating on the archaebacterial ArfR genes, ideally including data from additional Asgard and/or TACK genomes, are needed to corroborate this hypothesis.

Conclusions

Our comprehensive analysis of an extensive, well-curated dataset of ARF family proteins has provided evolutionary insights and raised questions to be addressed by future molecular cell biological exploration. The identification of 16 ancient ARF family paralogs both extends the inferred complexity of LECA and sets a framework of what components can be expected to be acting when delving into cellular function in diverse eukaryotes. By contrast the

identification of expanded complements, including novel paralogs, e.g. the metazoan Arl19 and Arl20, provide specific new candidates for investigation in some of the best explored and heavily utilized cell biological model systems. The diversity of domain architecture challenges the paradigm of this family strictly as small GTPases and begs probing of new protein-protein interactions. And finally, the resolved relationships and newly identified archaeal origins, particularly within the Asgard archaea have given us tremendous new insight into the earliest stages of eukaryotic origins.

Materials and Methods

Building and curation of the ARF family dataset

ARF family GTPases were searched in genome and/or transcriptome assemblies from 114 eukaryotic species selected such as to cover as many main eukaryotic lineages as possible (sequence identifiers, source databases, and further comments are provided in Supplementary table 1). The selection of taxa reflected the availability of relevant data as of 2018, when the sampling was frozen to obtain a final sequence dataset for all the subsequent analyses. As a result, several main eukaryote lineages, for which genome or transcriptome data became available more recently (e.g. the CRuMs supergroup, Telonemia, Rhodelphidia etc.), are not represented in our dataset. ARF family sequences were identified using BLAST and its variants (Altschul et al., 1997). Each organism-specific dataset was queried with reference members of the family and significant hits were evaluated by reverse BLAST searches against an in-house extensively curated taxonomically-rich database of GTPases. Query sequences being more similar to previously annotated members of the ARF family (including SRβ) were kept for further analysis. Existing protein sequence predictions were carefully evaluated and in a many cases revised by modifying the predicted exon-intron structure of the underlying gene model based on information from transcriptomic data or comparison to homologous sequences. To identify genes potentially missing from existing genome annotations, tblastn searches of nucleotide sequence data were carried out and gene models were created anew for previously missed genes. If possible, truncated sequences were completed using EST/TSA data or by iteratively recruiting raw genomic/transcriptomic sequencing reads. Revised, newly predicted or extended sequences are provided in Supplementary table 1. Putative pseudogenes (except for the human Arf2 pseudogene sequence, which can be reconstructed) as well as extremely divergent sequences with

disrupted ARF family motif(s) were not included into the dataset and are not listed in Supplementary table 1.

Each gene was initially annotated by considering results of blastp searches against our comprehensive database of Ras superfamily proteins (iteratively updated by adding sequences newly annotated in the course of the study). In most cases, the blastp output enabled unambiguous assignment of the query sequence into one of the previously delineated ortholog groups or into novel orthogroups that emerged during the study. Sequences most similar to true Arfs, yet difficult to assign into the Arf1 or Arf6 groups or being visibly divergent were provisionally annotated as "ArfX". Still more divergent ARF family members that did not show an apparently consistent affinity to a particular ARF family orthogroup when examined by BLAST searches were provisionally annotated as "ArlX". The annotation of some of the ArfX and ArlX sequences was subsequently revised after the employment of the ScrollSaw protocol described below.

Sequence data from the glaucophyte *Gloeochaete wittrockiana* were included despite the fact that we noticed contamination of both available transcriptome assemblies (MMETSP0308 and MMETSP1089; https://www.imicrobe.us/#/projects/104) by sequences from an amoebozoan. The putative contaminants were identified by careful examination of individual sequences and excluded from the dataset. Another potential contaminant (the contig PCB_a545736;2 K: 25), showing a high similarity to Arl4 genes from primates, was noticed in the transcriptome assembly from the breviate *Pygsuia biforma* and removed from analyses.

The ScrollSaw protocol and phylogenetic analyses

A master multiple alignment was built for the identified ARF family members, excluding short incomplete sequences and also all SRβ sequences, as this group is noticeably different from the core of the ARF family and many of its members tend to be rather divergent in their sequence. Altogether, the master alignment included 1931 sequences. It was built iteratively, starting with separate alignments for each group of sequences initially assigned to the same (potential) orthologous group using the on-line program MAFFT (version 7), with default parameters (Katoh and Standley, 2013). All alignments were checked by eye and further edited manually using BioEdit (Hall, 1999). The set of separate alignments was merged into one large alignment using the on-line Merge function of MAFFT. Divergent (ArlX) sequences were added to the alignment at the end. The final alignment was then manually trimmed to remove poorly conserved and unreliably aligned positions. After removing

redundancies, the alignment comprised 1891 non-identical sequences and 148 aligned positions (all falling within the GTPase domain shared across the family).

The alignment was subjected to an analysis essentially following the previously published ScrollSaw protocol (Elias et al., 2012). The sequences were divided according to the source species into 13 taxonomic groups covering the known diversity of eukaryotes: Holozoa, Holomycota, Apusomonadida, Breviatea, Amoebozoa, Malawimonadida, Planomonadida, Discoba, Metamonada, Archaeplastida, Cryptista, Haptista, and SAR. The sizes of the groups differ substantially, since many evolutionarily important lineages were represented only by a small number of species with genomic or transcriptomic data available at the time when we initiated the study (in the case of Breviatea and Planomonadida by only a single species). The master alignment was then subsampled by keeping only sequences from each possible pair of the taxa listed above, corresponding to 78 combinations. For each of the 78 alignments, genetic distances between the sequences were inferred using the maximum likelihood (ML) method (with the WAG+Γ+I substitution model) implemented in Tree-Puzzle 5.3 (Schmidt et al., 2002). Each resulting distance matrix was analysed using a custom Python script to identify the so-called minimal-distance pairs. A minimal-distance pair consists of two sequences from the two different taxonomic group compared that have mutually minimal distances when distances to sequences from the other taxon are considered. Minimal-distance pairs from all 78 pairwise taxon comparisons were gathered and redundancies were removed, resulting in a set of 568 sequences. To further reduce the complexity of the dataset we then removed all sequences that formed only one minimal-distance pair in all 78 pairwise taxon comparisons combined. This step yielded the final full ScrollSaw dataset comprising 354 sequences. A reduced ScrollSaw variant was prepared by removing sequences from the majority of metamonads exhibiting generally divergent genes, including *Monocercomonoides exilis*, *Giardia intestinalis*, *Spironucleus* spp. and *Trichomonas vaginalis*.

The two variants of the final ScrollSaw dataset were used for inferring the ML phylogenetic trees using the program IQ-TREE (Nguyen et al., 2015). The substitution model (LG+I+G4) was selected by the program itself based on specific optimality criteria. Branch support was assessed by the SH-aLRT test (Guindon et al., 2010) and the ultrafast bootstrap approximation (Minh et al., 2013). The branch support of the reduced dataset was further examined by MrBayes 3.2 (Ronquist et al., 2011) using the CIPRES Science Gateway (Miller et al., 2010) with the following settings: prset aamodelpr=fixed(WAG); lset rates=gamma Ngammacat=4 mcmc ngen=1000000 printfreq=10000 samplefreq=1000 nchains=4

burnin=80. A number of additional alignments for specific dedicated analyses, derived by subsampling the master alignment or aligning the selected sequences *de novo* (using MAFFT with subsequent manual editing as described above) were used for ML phylogenetic inference using the same or similar approach. The alignments were in most cases trimmed according to the mask applied to the master alignment. Smaller phylogenetic analysis with only a subset of paralogous groups were trimmed either manually or by stand-alone version of trimAl (version 1.2rev57; option automated1; Capella-Gutierrez et al., 2009) in order to retrieve more positions for the ML phylogenetic analysis. The stand-alone version of IQ-TREE or the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/; Trifinopoulos et al., 2016) were used for the analyses. The substitution models were selected by the model selection program implemented in the IQ-TREE. Branch support was assessed by the SH-aLRT test and the ultrafast bootstrap approximation.

Annotation of sequences

The full and reduced ScrollSaw datasets were used as a basis for annotation of the rest of the sequences. The identity of individual sequences or their groups was tested by adding them to the reduced ScrollSaw dataset and inferring a ML tree with IQ-TREE. The scrutinized sequences were assigned to a particular ancestral eukaryotic paralog and annotated accordingly if they clustered together with reference representatives of the given paralog group and the relationship was supported by SH-aLRT and ultrafast bootstrap values of ≥80 and 95, respectively. Not all genes could be annotated by this approach, hence the HMMER package (stand alone version 3.0; hmmer.org) was employed as an alternative. The aligned full ScrollSaw dataset was divided into 14 separate alignments, each representing one ancestral paralog (Arf1, 6, Arl1, 2, 3, 5, 6, 8, 13, 16, 18, Arfrp1, Sar1 and Sar1L). A profile HMM was constructed for each alignment using hmmbuild and a database of profile HMMs was created using hmmpress. The unannotated sequences were then used as queries in hmmscan searches against the database and the "best 1 domain" score difference between the first and the second best hits was determined. If this difference was equal to or higher than 20, the sequence was annotated according to the best hit. Sequences annotated based on the phylogenetic analyses or hmmscan searches are marked by asterisk (*) in the column "Conclusively annotated" in the Supplementary table 1. Proteins representing SRβ and Arl17, which were not represented by reference sequences in the ScrollSaw dataset, were unequivocally identified owing to the distinct characters of these sequence groups, which makes them easy to recognise by BLAST-based similarity searches (SRβ) or by considering

the presence of the novel conserved C-terminal domain (Arl17; see the main text). All SRβ and Arl17 proteins are therefore also considered as conclusively annotated. A single truncated sequence (Arl17b gene from *Chromera velia*) lacked the C-terminal extension with the characteristic C-terminal domain, but was assigned to the Arl17 group based on its close sequence similarity to undisputed Arl17 sequences. A combination of BLAST searches, ML phylogenetic analyses and comparison of exon-intron structure was used to obtain the most likely annotation of the sequences that could not be conclusively annotated by the aforementioned approaches. Several sequences were annotated as Arf1/6, as they showed affinity to the Arf1/6 clade, but it was impossible to decide whether they originated from ancestral Arf1 or Arf6 paralogs. Only 160 out of more than 2000 ARF family sequences analysed could not be annotated with any confidence, so they remained unassigned to any ancestral paralog (Supplementary table 1; Supplementary table 2).

Taxon-specific ScrollSaw analyses and annotation of lineage-specific paralogs

To detect lineage-specific paralogs, we applied the ScrollSaw protocol separately to sets of sequences from the following main eukaryote taxa: Chloroplastida, Rhodophyta, Glaucophyta, Cryptista, Haptista, SAR, Discoba, Metamonada, Amoebozoa, Holomycota and Holozoa. Lineages represented by only one or two species (Apusomonadida, Breviatea, Planomonadida, and Malawimonadida) were not included. The ScrollSaw protocol and phylogenetic analyses of the resulted datasets were performed generally as described above for the whole dataset. For each main eukaryote taxon analysed, species representing it were assigned to predefined monophyletic subgroups specified in Supplementary table 3. Sequences from these species were extracted from the trimmed master alignment of the ARF family protein (except for sequences from Holozoa, which were aligned *de novo* using MAFFT and then trimmed according to the mask used for the whole dataset), the ScrollSaw protocol was applied to identify minimal-distance pairs, and ML phylogenetic trees were calculated on the filtered sequences. In contrast to the pan-eukaryotic ScrollSaw analysis, sequences that formed only one minimal-distance pair were not omitted (except for the analysis of the Holozoa dataset, where the criterion of the sequence belonging to at least two minimal-distance pairs was kept). The ML trees were inspected to identify robustly supported clades that would define conserved paralogs ancestral for the focal eukaryotic taxon but different from the previously defined ancestral eukaryote paralogs. In the case of Holozoa, paralogs specific for individual subgroups were considered, too. Further representatives of these paralogs (i.e., specific orthologs of the constituent sequences identified in the ScrollSaw

trees) were then identified among the sequences that did not pass the ScrollSaw step by a combination of BLAST searches, phylogenetic analyses and (in case of Holozoa) HMMER-based comparisons. Candidates for ancestral taxon-specific paralogs were detected only in Chloroplastida, Rhodophyta, and Glaucophyta, as described in detail in the main text.

Prediction of transmembrane regions and post-translation modifications

The presence of transmembrane (TM) regions in ARF family proteins was examined using the online TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). In the case of sequences with suspicious TM absence or presence (i.e., when the result was untypical for the respective ARF family subgroup), the on-line tool TMpred (https://www.ch.embnet.org/software/TMPRED_form.html) was additionally employed. The predictions are listed in Supplementary table 1. N-terminal myristoylation of sequences with the glycine residue at the second position was evaluated using the on-line ExPASy Myristoylator tool (http://web.expasy.org/myristoylator/; Bologna et al., 2004), NMT - The MYR Predictor (http://mendel.imp.ac.at/myristate/SUPLpredictor.htm; Maurer-Stroh et al., 2002), and the stand-alone version of GPS-Lipid (v1.0, http://lipid.biocuckoo.org/index.php; Xie et al., 2016). Only those proteins predicted as N-terminally myristoylated by at least two tools were considered as significant candidates. Setting of all tools was default except for NMT - The MYR Predictor where only N-terminal glycine residues were considered, and fungal sequences were predicted with the “Fungi specific” option. In case of GPS-Lipid, the threshold was set to “low”. Possible S-palmitoylation was predicted using SeqPalm (http://lishuyan.lzu.edu.cn/seqpalm/; Li et al., 2015), the stand-alone version of CKSAAP-Palm programme (http://doc.aporc.org/wiki/CKSAAP-Palm; Wang et al., 2009), PalmPred (http://proteininformatics.org/mkumar/palmpred/index.html; Kumari et al., 2014), stand-alone version of GPS-Lipid, and WAP-Palm (http://bioinfo.ncu.edu.cn/WAP-Palm.aspx; Shi et al., 2013). Only those sites predicted as S-palmitoylated by at least three tools were considered as significant candidates. Setting of all tools was default except for GPS-Lipid with the threshold set to “high”. Complete results from all tools are showed in Supplementary table 4, consensual results are included in Supplementary table 1. Possible prenylation was assessed only for VisArfX from the alga *Vischeria* sp., as it is the only protein from our dataset with a typical C-terminal prenylation motif. The online programs iPreny-PseAAC (http://app.aporc.org/iPreny-PseAAC/index.html; Xu et al., 2017) and GPS-Lipid were used with default settings; both tools predicted VisArfX as a prenylated protein.

Other sequence analyses

Intron positions were investigated in four groups of ARF family genes (Sar1/Sar1L; Arl8/Arl18; Arfs and the GTPase domain of Arl17; Arfs and Arf-like in Holozoa) as a means to illuminate the origin and relationships of these genes. The positions of introns (including their phases) were mapped onto a multiple alignment of respective protein sequences using a custom Java script. The multiple sequence alignments were constructed *de novo* using MAFFT, inspected visually and adjusted manually whenever necessary (Sar1/Sar1L, Arl8/Arl18, Arf, and Arf-like in holozoans). For the analysis of Arf and Arl17 genes, the respective protein sequences were extracted from the master alignment. Sequences with no introns in the coding sequence or represented only by transcriptomic data were omitted. A manually curated dataset of gene exon-intron structures was used as the input for the intron positions mapping. For presentation purposes, regions corresponding to unconserved N- and C- termini of the sequences were trimmed and long sequence-specific insertions were collapsed.

To highlight the pattern of protein sequence conservation, CHROMA (ver. 1.0 Goodstadt and Ponting, 2001) was used for processing some of the multiple sequence alignments presented (novel C-terminal domain of Arl17, C-terminus of Arl15, Arl13 and ArfL from chlorarachniophytes, and ArfX from Pavlovophyceae). Sequence logos of the Walker B motif were obtained using the on-line tool WebLogo 3 (http://weblogo.threeplusone.com/create.cgi; Crooks et al., 2004) from the multiple sequence alignment of the respective sequences after removing sequence-specific insertions present in a few sequences.

Conserved protein domains and other structural features in ARF family proteins were identified using searches of Pfam (http://pfam.xfam.org/; Finn et al., 2016), the Conserved Domains database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; Marchler-Bauer et al., 2017), and the SMART database (http://smart.embl-heidelberg.de/; Letunic and Bork, 2018). Domain predictions provided by the three tools were compared and spurious results (low-significance with only a single tool) were ignored. The identity of unusual N-terminal extensions present in some Arl13 proteins were evaluated using HHpred (https://toolkit.tuebingen.mpg.de/; Söding et al., 2005). Multiple sequence alignments of the different forms on the N-terminal extensions conserved within different taxa were used as queries in the HHpred searches. In case of the N-terminal extension conserved in Arl13 proteins from Euglenozoa, the sampling was expanded beyond the focal set of taxa (including only three euglenozoans) by adding to the alignment several additional euglenozoan Arl13

sequences to improve the representativeness of the alignment. Similarly, additional chlorarachniophyte Arl13 sequences were identified and aligned with the sole representative in the focal dataset (that from *B. natans*), and additional stramenopile (oomycete and ochrophyte) Arl13 sequences with the same conserved N-terminal extension as the stramenopile sequences in the focal set were included to increase the sensitivity of the analysis. Some TRIM23 sequences were predicted by the standard tools to contain only one BBOX domain rather than the two common in most members of this group, but inspection of a multiple sequence alignment revealed high similarity of all sequences in the respective region, suggesting that all TRIM23 sequences likely conform to the same domain architecture with two BBOX domains.

Acknowledgements

We thank Eunsoo Kim (American Museum of Natural History, New York) for an access to a genome assembly of *Goniomonas avonlea* prior to publication and Vladimír Hampl (Charles University in Prague) for his permission to use sequences from an unpublished genome assembly of *Paratrimastix pyriformis*. This work was supported by Czech Science Foundation grant 18-13458S; ERD Funds, project OPVVV CZ.02.1.01/0.0/0.0/16_019/0000759 (Centre for research of pathogenicity and virulence of parasites); and the infrastructure grant CZ.1.05/2.1.00/19.0388 („Přístroje IET“). Work in the Kahn lab is supported by a grant from the National Institutes of Health (NIH R35GM122568). Work in the Dacks Lab is supported by grants from the Natural Sciences and Engineering Research Council of Canada (RES0021028, RES0043758, RES0046091). JBD is the Canada Research Chair (Tier Ii) in Evolutionary Cell Biology.

Competing interests

No competing interests declared.

References

Al-Bassam, J. (2017). Revisiting the tubulin cofactors and Arl2 in the regulation of soluble αβ-tubulin pools and their effect on microtubule dynamics. Molecular Biology of the Cell, 28(3), 359–363. https://doi.org/10.1091/mbc.E15-10-0694

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Anantharaman, V., Abhiman, S., de Souza, R. F., & Aravind, L. (2011). Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system. Gene, 475(2), 63–78. https://doi.org/10.1016/j.gene.2010.12.001

Barlow, L. D., Nývltová, E., Aguilar, M., Tachezy, J., & Dacks, J. B. (2018). A sophisticated, differentiated Golgi in the ancestor of eukaryotes. BMC Biology, 16(1), 27. https://doi.org/10.1186/s12915-018-0492-9

Behnia, R., Panic, B., Whyte, J. R. C., & Munro, S. (2004). Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. Nature Cell Biology, 6(5), 405–413. https://doi.org/10.1038/ncb1120

Bologna, G., Yvon, C., Duvaud, S., & Veuthey, A.-L. (2004). N-Terminal myristoylation predictions by ensembles of neural networks. Proteomics, 4(6), 1626–1632. https://doi.org/10.1002/pmic.200300783

Bosgraaf, L., & Van Haastert, P. J. M. (2003). Roc, a Ras/GTPase domain in complex proteins. Biochimica et Biophysica Acta, 1643(1–3), 5–10. https://doi.org/10.1016/j.bbamcr.2003.08.008

Botuyan, M. V., & Mer, G. (2016). Chapter 8—Tudor Domains as Methyl-Lysine and Methyl-Arginine Readers. In O. Binda & M. E. Fernandez-Zapico (Ed.), Chromatin Signaling and Diseases (s. 149–165). Academic Press. https://doi.org/10.1016/B978-0-12-802389-1.00008-3

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics (Oxford, England), 25(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Casanova, J. E. (2007). Regulation of Arf activation: The Sec7 family of guanine nucleotide exchange factors. Traffic (Copenhagen, Denmark), 8(11), 1476–1485. https://doi.org/10.1111/j.1600-0854.2007.00634.x

Cevik, S., Hori, Y., Kaplan, O. I., Kida, K., Toivenon, T., Foley-Fisher, C., Cottell, D., Katada, T., Kontani, K., & Blacque, O. E. (2010). Joubert syndrome Arl13b functions at ciliary membranes and stabilizes protein transport in Caenorhabditis elegans. The Journal of Cell Biology, 188(6), 953–969. https://doi.org/10.1083/jcb.200908133

Cevik, S., Sanders, A. A. W. M., Van Wijk, E., Boldt, K., Clarke, L., van Reeuwijk, J., Hori, Y., Horn, N., Hetterschijt, L., Wdowicz, A., Mullins, A., Kida, K., Kaplan, O. I., van Beersum, S. E. C., Man Wu, K., Letteboer, S. J. F., Mans, D. A., Katada, T., Kontani, K., Ueffing, M., Roepman, R., Kremer, H., Blacque, O. E. (2013). Active transport and diffusion barriers restrict Joubert Syndrome-associated ARL13B/ARL-13 to an Inv-like ciliary membrane subdomain. PLoS Genetics, 9(12), e1003977. https://doi.org/10.1371/journal.pgen.1003977

Clark, J., Moore, L., Krasinskas, A., Way, J., Battey, J., Tamkun, J., & Kahn, R. A. (1993). Selective amplification of additional members of the ADP-ribosylation factor (ARF) family: Cloning of additional human and Drosophila ARF-like genes. Proceedings of the National Academy of Sciences of the United States of America, 90(19), 8952–8956. https://doi.org/10.1073/pnas.90.19.8952

Cotton, M., Boulay, P.-L., Houndolo, T., Vitale, N., Pitcher, J. A., & Claing, A. (2007). Endogenous ARF6 interacts with Rac1 upon angiotensin II stimulation to regulate membrane ruffling and cell migration. Molecular Biology of the Cell, 18(2), 501–511. https://doi.org/10.1091/mbc.e06-06-0567

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. Genome Research, 14(6), 1188–1190. https://doi.org/10.1101/gr.849004

Dacks, J. B., Field, M. C., Buick, R., Eme, L., Gribaldo, S., Roger, A. J., Brochier-Armanet, C., & Devos, D. P. (2016). The changing view of eukaryogenesis—Fossils, cells, lineages

and how they all come together. Journal of Cell Science, 129(20), 3695–3703. https://doi.org/10.1242/jcs.178566

Dacks, J. B., & Robinson, M. S. (2017). Outerwear through the ages: Evolutionary cell biology of vesicle coats. Current Opinion in Cell Biology, 47, 108–116. https://doi.org/10.1016/j.ceb.2017.04.001

Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č., Lang, B. F., & Eliáš, M. (2015). Bacterial proteins pinpoint a single eukaryotic root. Proceedings of the National Academy of Sciences of the United States of America, 112(7), E693–E699. https://doi.org/10.1073/pnas.1420657112

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A., & Rout, M. P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biology, 2(12), e380. https://doi.org/10.1371/journal.pbio.0020380

Dian, C., Pérez-Dorado, I., Rivière, F., Asensio, T., Legrand, P., Ritzefeld, M., Shen, M., Cota, E., Meinnel, T., Tate, E. W., & Giglione, C. (2020). High-resolution snapshots of human N-myristoyltransferase in action illuminate a mechanism promoting N-terminal Lys and Gly myristoylation. Nature Communications, 11(1), 1132. https://doi.org/10.1038/s41467-020-14847-3

Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M. C., & Pereira-Leal, J. B. (2011). Thousands of Rab GTPases for the Cell Biologist. PLOS Computational Biology, 7(10), e1002217. https://doi.org/10.1371/journal.pcbi.1002217

Donaldson, J. G., & Jackson, C. L. (2011). Arf Family G Proteins and their regulators: Roles in membrane transport, development and disease. Nature Reviews Molecular Cell Biology, 12(6), 362–375. https://doi.org/10.1038/nrm3117

Dong, J.-H., Wen, J.-F., & Tian, H.-F. (2007). Homologs of eukaryotic Ras superfamily proteins in prokaryotes and their novel phylogenetic correlation with their eukaryotic analogs. Gene, 396(1), 116–124. https://doi.org/10.1016/j.gene.2007.03.001

D'Souza-Schorey, C., & Stahl, P. D. (1995). Myristoylation is required for the intracellular localization and endocytic function of ARF6. Experimental Cell Research, 221(1), 153–159. https://doi.org/10.1006/excr.1995.1362

Duronio, R. J., Rudnick, D. A., Adams, S. P., Towler, D. A., & Gordon, J. I. (1991). Analyzing the substrate specificity of Saccharomyces cerevisiae myristoyl-CoA:protein N-myristoyltransferase by co-expressing it with mammalian G protein alpha subunits in Escherichia coli. Journal of Biological Chemistry, 266(16), 10498–10504.

Elias, M., & Archibald, J. M. (2009). The RJL family of small GTPases is an ancient eukaryotic invention probably functionally associated with the flagellar apparatus. Gene, 442(1–2), 63–72. https://doi.org/10.1016/j.gene.2009.04.011

Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M. C., & Dacks, J. B. (2012). Sculpting the endomembrane system in deep time: High resolution phylogenetics of Rab GTPases. Journal of Cell Science, 125(10), 2500–2508. https://doi.org/10.1242/jcs.101378

Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., & Stetter, K. O. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. Proceedings of the National Academy of Sciences of the United States of America, 105(23), 8102–8107. https://doi.org/10.1073/pnas.0801980105

Eme, L., Spang, A., Lombard, J., Stairs, C. W., & Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. Nature Reviews Microbiology, 15(12), nrmicro.2017.133. https://doi.org/10.1038/nrmicro.2017.133

Fansa, E. K., & Wittinghofer, A. (2016). Sorting of lipidated cargo by the Arl2/Arl3 system. Small GTPases, 7(4), 222–230. https://doi.org/10.1080/21541248.2016.1224454

Field, M. C., & Rout, M. P. (2019). Pore timing: The evolutionary origins of the nucleus and nuclear pore complex. F1000Research, 8, F1000 Faculty Rev-369. https://doi.org/10.12688/f1000research.16402.1

Fielding, A. B., Schonteich, E., Matheson, J., Wilson, G., Yu, X., Hickson, G. R. X., Srivastava, S., Baldwin, S. A., Prekeris, R., & Gould, G. W. (2005). Rab11-FIP3 and FIP4 interact with Arf6 and the exocyst to control membrane traffic in cytokinesis. EMBO Journal, 24(19), 3389–3399. https://doi.org/10.1038/sj.emboj.7600803

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Research, 44(D1), D279-285. https://doi.org/10.1093/nar/gkv1344

Fisher, S., Kuna, D., Caspary, T., Kahn, R. A., & Sztul, E. (2020). ARF family GTPases with links to cilia. American Journal of Physiology. Cell Physiology, 319(2), C404–C418. https://doi.org/10.1152/ajpcell.00188.2020

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E. G. J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthélémy, R.-M., Bast, J., Bazykin, G. A., Chabrol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., Gouret, P., Hallatschek, O., Hecox-Lea, B., Labadie, K., Lejeune, B., Piskurek, O., Poulain, J., Rodriguez, F., Ryan, J. F., Vakhrusheva, O. A., Wajnberg, E., Wirth, B., Yushenova, I., Kellis, M., Kondrashov, A. S., Mark Welch, D. B., Pontarotti, P., Weissenbach, J., Wincker, P., Jaillon, O., Van Doninck, K. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature, 500(7463), 453–457. https://doi.org/10.1038/nature12326

Francis, J. W., Goswami, D., Novick, S. J., Pascal, B. D., Weikum, E. R., Ortlund, E. A., Griffin, P. R., & Kahn, R. A. (2017a). Nucleotide Binding to ARL2 in the TBCD·ARL2·β-Tubulin Complex Drives Conformational Changes in β-Tubulin. Journal of Molecular Biology, 429(23), 3696–3716. https://doi.org/10.1016/j.jmb.2017.09.016

Francis, J. W., Newman, L. E., Cunningham, L. A., & Kahn, R. A. (2017b). A Trimer Consisting of the Tubulin-specific Chaperone D (TBCD), Regulatory GTPase ARL2, and β-Tubulin Is Required for Maintaining the Microtubule Network. Journal of Biological Chemistry, 292(10), 4336–4349. https://doi.org/10.1074/jbc.M116.770909

Francis, J. W., Turn, R. E., Newman, L. E., Schiavon, C., & Kahn, R. A. (2016). Higher order signaling: ARL2 as regulator of both mitochondrial fusion and microtubule dynamics allows integration of 2 essential cell functions. Small GTPases, 7(4), 188–196. https://doi.org/10.1080/21541248.2016.1211069

Fu, H. W., & Casey, P. J. (1999). Enzymology and biology of CaaX protein prenylation. Recent Progress in Hormone Research, 54, 315–342; discussion 342-343.

Funakoshi, Y., Hasegawa, H., & Kanaho, Y. (2011). Regulation of PIP5K activity by Arf6 and its physiological significance. Journal of Cellular Physiology, 226(4), 888–895. https://doi.org/10.1002/jcp.22482

Gabernet-Castello, C., O'Reilly, A. J., Dacks, J. B., & Field, M. C. (2013). Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. Molecular Biology of the Cell, 24(10), 1574–1583. https://doi.org/10.1091/mbc.E12-07-0557

Gaillard, A. R., Diener, D. R., Rosenbaum, J. L., & Sale, W. S. (2001). Flagellar radial spoke protein 3 is an A-kinase anchoring protein (AKAP). Journal of Cell Biology, 153(2), 443–448. https://doi.org/10.1083/jcb.153.2.443

Genschik, P., Sumara, I., & Lechner, E. (2013). The emerging family of CULLIN3-RING ubiquitin ligases (CRL3s): Cellular functions and disease implications. EMBO Journal, 32(17), 2307–2320. https://doi.org/10.1038/emboj.2013.173

Gigante, E. D., Taylor, M. R., Ivanova, A. A., Kahn, R. A., & Caspary, T. (2020). ARL13B regulates Sonic hedgehog signaling from outside primary cilia. ELife, 9. https://doi.org/10.7554/eLife.50434

Gillingham, A. K., & Munro, S. (2007). The small G proteins of the Arf family and their regulators. Annual Review of Cell and Developmental Biology, 23, 579–611. https://doi.org/10.1146/annurev.cellbio.23.090506.123209

Goodstadt, L., & Ponting, C. P. (2001). CHROMA: Consensus-based colouring of multiple alignments for publication. Bioinformatics, 17(9), 845–846. https://doi.org/10.1093/bioinformatics/17.9.845

Gotthardt, K., Lokaj, M., Koerner, C., Falk, N., Gießl, A., & Wittinghofer, A. (2015). A G-protein activation cascade from Arl13B to Arl3 and implications for ciliary targeting of lipidated proteins. ELife, 4. https://doi.org/10.7554/eLife.11859

Guerini, D. (1997). Calcineurin: Not just a simple protein phosphatase. Biochemical and Biophysical Research Communications, 235(2), 271–275. https://doi.org/10.1006/bbrc.1997.6802

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology, 59(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41, 95–98.

Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., & Millar, A. H. (2007). SUBA: The Arabidopsis Subcellular Database. Nucleic Acids Research, 35(Database issue), D213–D218. https://doi.org/10.1093/nar/gkl863

Higginbotham, H., Eom, T.-Y., Mariani, L. E., Bachleda, A., Hirt, J., Gukassyan, V., Cusack, C. L., Lai, C., Caspary, T., & Anton, E. S. (2012). Arl13b in primary cilia regulates the migration and placement of interneurons in the developing cerebral cortex. Developmental Cell, 23(5), 925–938. https://doi.org/10.1016/j.devcel.2012.09.019

Hofmann, I., & Munro, S. (2006). An N-terminally acetylated Arf-like GTPase is localised to lysosomes and affects their motility. Journal of Cell Science, 119(Pt 8), 1494–1503. https://doi.org/10.1242/jcs.02958

Hofmann, I., Thompson, A., Sanderson, C. M., & Munro, S. (2007). The Arl4 family of small G proteins can recruit the cytohesin Arf6 exchange factors to the plasma membrane. Current Biology, 17(8), 711–716. https://doi.org/10.1016/j.cub.2007.03.007

Hori, Y., Kobayashi, T., Kikko, Y., Kontani, K., & Katada, T. (2008). Domain architecture of the atypical Arf-family GTPase Arl13b involved in cilia formation. Biochemical and Biophysical Research Communications, 373(1), 119–124. https://doi.org/10.1016/j.bbrc.2008.06.001

Houghton, F. J., Bellingham, S. A., Hill, A. F., Bourges, D., Ang, D. K. Y., Gemetzis, T., Gasnereau, I., & Gleeson, P. A. (2012). Arl5b is a Golgi-localised small G protein involved in the regulation of retrograde transport. Experimental Cell Research, 318(5), 464–477. https://doi.org/10.1016/j.yexcr.2011.12.023

Cheung, P.-Y. P., & Pfeffer, S. R. (2016). Transport Vesicle Tethering at the Trans Golgi Network: Coiled Coil Proteins in Action. Frontiers in Cell and Developmental Biology, 4, 18. https://doi.org/10.3389/fcell.2016.00018

Ishida, M., E Oguchi, M., & Fukuda, M. (2016). Multiple Types of Guanine Nucleotide Exchange Factors (GEFs) for Rab Small GTPases. Cell Structure and Function, 41(2), 61–79. https://doi.org/10.1247/csf.16008

Ivanova, A. A., Caspary, T., Seyfried, N. T., Duong, D. M., West, A. B., Liu, Z., & Kahn, R. A. (2017). Biochemical characterization of purified mammalian ARL13B protein indicates that it is an atypical GTPase and ARL3 guanine nucleotide exchange factor (GEF). Journal of Biological Chemistry, 292(26), 11091–11108. https://doi.org/10.1074/jbc.M117.784025

Jackson, C. L. (2014). Arf Proteins and Their Regulators: At the Interface Between Membrane Lipids and the Protein Trafficking Machinery. In A. Wittinghofer (Ed.), Ras Superfamily Small G Proteins: Biology and Mechanisms 2: Transport (s. 151–180). Springer International Publishing. https://doi.org/10.1007/978-3-319-07761-1_8

Jackson, C. L., & Bouvet, S. (2014). Arfs at a glance. Journal of Cell Science, 127(Pt 19), 4103–4109. https://doi.org/10.1242/jcs.144899

Jivan, A., Earnest, S., Juang, Y.-C., & Cobb, M. H. (2009). Radial spoke protein 3 is a mammalian protein kinase A-anchoring protein that binds ERK1/2. Journal of Biological Chemistry, 284(43), 29437–29445. https://doi.org/10.1074/jbc.M109.048181

Jones, S., Jedd, G., Kahn, R. A., Franzusoff, A., Bartolini, F., & Segev, N. (1999). Genetic interactions in yeast between Ypt GTPases and Arf guanine nucleotide exchangers. Genetics, 152(4), 1543–1556.

Kaessmann, H., Vinckenbosch, N., & Long, M. (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. Nature Reviews Genetics, 10(1), 19–31. https://doi.org/10.1038/nrg2487

Kahn, R. A., Goddard, C., & Newkirk, M. (1988). Chemical and immunological characterization of the 21-kDa ADP-ribosylation factor of adenylate cyclase. Journal of Biological Chemistry, 263(17), 8282–8287. http://www.jbc.org/content/263/17/8282

Kahn, R. A. (2009). Toward a model for Arf GTPases as regulators of traffic at the Golgi. FEBS Letters, 583(23), 3872–3879. https://doi.org/10.1016/j.febslet.2009.10.066

Kahn, R. A., Bruford, E., Inoue, H., Logsdon, J. M., Nie, Z., Premont, R. T., Randazzo, P. A., Satake, M., Theibert, A. B., Zapp, M. L., & Cassel, D. (2008). Consensus nomenclature for the human ArfGAP domain-containing proteins. Journal of Cell Biology, 182(6), 1039–1044. https://doi.org/10.1083/jcb.200806041

Kahn, R. A., Cherfils, J., Elias, M., Lovering, R. C., Munro, S., & Schurmann, A. (2006). Nomenclature for the human Arf family of GTP-binding proteins: ARF, ARL, and SAR proteins. Journal of Cell Biology, 172(5), 645–650. https://doi.org/10.1083/jcb.200512057

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010

Keenan, R. J., Freymann, D. M., Stroud, R. M., & Walter, P. (2001). The signal recognition particle. Annual Review of Biochemistry, 70, 755–775. https://doi.org/10.1146/annurev.biochem.70.1.755

Khatter, D., Sindhwani, A., & Sharma, M. (2015). Arf-like GTPase Arl8: Moving from the periphery to the center of lysosomal biology. Cellular Logistics, 5(3). https://doi.org/10.1080/21592799.2015.1086501

Klinger, C. M., Spang, A., Dacks, J. B., & Ettema, T. J. G. (2016). Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. Molecular Biology and Evolution, 33(6), 1528–1541. https://doi.org/10.1093/molbev/msw034

Klöpper, T. H., Kienle, N., Fasshauer, D., & Munro, S. (2012). Untangling the evolution of Rab G proteins: Implications of a comprehensive genomic analysis. BMC Biology, 10(1), 71. https://doi.org/10.1186/1741-7007-10-71

Kumari, B., Kumar, R., & Kumar, M. (2014). PalmPred: An SVM based palmitoylation prediction method using sequence profile information. PloS One, 9(2), e89246. https://doi.org/10.1371/journal.pone.0089246

Lee, F. J., Huang, C. F., Yu, W. L., Buu, L. M., Lin, C. Y., Huang, M. C., Moss, J., & Vaughan, M. (1997). Characterization of an ADP-ribosylation factor-like 1 protein in Saccharomyces cerevisiae. Journal of Biological Chemistry, 272(49), 30998–31005. https://doi.org/10.1074/jbc.272.49.30998

Leipe, D. D., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. Journal of Molecular Biology, 317(1), 41–72. https://doi.org/10.1006/jmbi.2001.5378

Lemmon, M. A. (2007). Pleckstrin homology (PH) domains and phosphoinositides. Biochemical Society Symposium, 74, 81–93. https://doi.org/10.1042/BSS0740081

Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. Nucleic Acids Research, 46(D1), D493–D496. https://doi.org/10.1093/nar/gkx922

Li, C.-C., Chiang, T.-C., Wu, T.-S., Pacheco-Rodriguez, G., Moss, J., & Lee, F.-J. S. (2007). ARL4D recruits cytohesin-2/ARNO to modulate actin remodeling. Molecular Biology of the Cell, 18(11), 4420–4437. https://doi.org/10.1091/mbc.e07-02-0149

Li, S., Li, J., Ning, L., Wang, S., Niu, Y., Jin, N., Yao, X., Liu, H., & Xi, L. (2015). In Silico Identification of Protein S-Palmitoylation Sites and Their Involvement in Human Inherited Disease. Journal of Chemical Information and Modeling, 55(9), 2015–2025. https://doi.org/10.1021/acs.jcim.5b00276

Li, Y., Kelly, W. G., Logsdon, J. M., Schurko, A. M., Harfe, B. D., Hill-Harfe, K. L., & Kahn, R. A. (2004). Functional genomic analysis of the ADP-ribosylation factor family of GTPases: Phylogeny among diverse eukaryotes and function in C. elegans. FASEB Journal, 18(15), 1834–1850. https://doi.org/10.1096/fj.04-2273com

Lin, C.-Y., Li, C.-C., Huang, P.-H., & Lee, F.-J. S. (2002). A developmentally regulated ARF-like 5 protein (ARL5), localized to nuclei and nucleoli, interacts with heterochromatin protein 1. Journal of Cell Science, 115(Pt 23), 4433–4445. https://doi.org/10.1242/jcs.00123

Liu, Y., Kahn, R. A., & Prestegard, J. H. (2009). Structure and Membrane Interaction of Myristoylated ARF1. Structure (London, England : 1993), 17(1), 79–87. https://doi.org/10.1016/j.str.2008.10.020

Liu, Y., Kahn, R. A., & Prestegard, J. H. (2010). Dynamic structure of membrane-anchored Arf*GTP. Nature Structural & Molecular Biology, 17(7), 876–881. https://doi.org/10.1038/nsmb.1853

Makiuchi, T., & Nozaki, T. (2014). Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. Biochimie, 100, 3–17. https://doi.org/10.1016/j.biochi.2013.11.018

Manolea, F., Chun, J., Chen, D. W., Clarke, I., Summerfeldt, N., Dacks, J. B., & Melançon, P. (2010). Arf3 Is Activated Uniquely at the trans-Golgi Network by Brefeldin A-inhibited Guanine Nucleotide Exchange Factors. Molecular Biology of the Cell, 21(11), 1836–1849. https://doi.org/10.1091/mbc.E10-01-0016

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., & Bryant, S. H. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Research, 43(Database issue), D222-226. https://doi.org/10.1093/nar/gku1221

Maurer-Stroh, S., Eisenhaber, B., & Eisenhaber, F. (2002). N-terminal N-myristoylation of proteins: Prediction of substrate proteins from amino acid sequence. Journal of Molecular Biology, 317(4), 541–557. https://doi.org/10.1006/jmbi.2002.5426

Melville, D. B., Studer, S., & Schekman, R. (2020). Small sequence variations between two mammalian paralogs of the small GTPase SAR1 underlie functional differences in coat protein complex II assembly. Journal of Biological Chemistry, 295(25), 8401–8412. https://doi.org/10.1074/jbc.RA120.012964

Meza, I., Talamás-Rohana, P., & Vargas, M. A. (2006). The Cytoskeleton of Entamoeba histolytica: Structure, Function, and Regulation by Signaling Pathways. Archives of Medical Research, 37(2), 234–243. https://doi.org/10.1016/j.arcmed.2005.09.008

Miertzschke, M., Koerner, C., Spoerner, M., & Wittinghofer, A. (2014). Structural insights into the small G-protein Arl13B and implications for Joubert syndrome. Biochemical Journal, 457(2), 301–311. https://doi.org/10.1042/BJ20131097

Miller, E. A., & Barlowe, C. (2010). Regulation of coat assembly—Sorting things out at the ER. Current Opinion in Cell Biology, 22(4), 447–453. https://doi.org/10.1016/j.ceb.2010.04.003

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE), 1–8. https://doi.org/10.1109/GCE.2010.5676129

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution, 30(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Mizuno-Yamasaki, E., Rivera-Molina, F., & Novick, P. (2012). GTPase networks in membrane traffic. Annual Review of Biochemistry, 81, 637–659. https://doi.org/10.1146/annurev-biochem-052810-093700

More, K., Klinger, C. M., Barlow, L. D., & Dacks, J. B. (2020). Evolution and Natural History of Membrane Trafficking in Eukaryotes. Current Biology, 30(10), R553–R564. https://doi.org/10.1016/j.cub.2020.03.068

Mourão, A., Nager, A. R., Nachury, M. V., & Lorentzen, E. (2014). Structural basis for membrane targeting of the BBSome by ARL6. Nature Structural & Molecular Biology, 21(12), 1035–1041. https://doi.org/10.1038/nsmb.2920

Nabais, C., Peneda, C., & Bettencourt-Dias, M. (2020). Evolution of centriole assembly. Current Biology: CB, 30(10), R494–R502. https://doi.org/10.1016/j.cub.2020.02.036

Neuwald, A. F. (2007). Galpha Gbetagamma dissociation may be due to retraction of a buried lysine and disruption of an aromatic cluster by a GTP-sensing Arg Trp pair. Protein Science, 16(11), 2570–2577. https://doi.org/10.1110/ps.073098107

Newman, L. E., Schiavon, C. R., Turn, R. E., & Kahn, R. A. (2017). The ARL2 GTPase regulates mitochondrial fusion from the intermembrane space. Cellular Logistics, 7(3), e1340104. https://doi.org/10.1080/21592799.2017.1340104

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

Panic, B., Whyte, J. R. C., & Munro, S. (2003). The ARF-like GTPases Arl1p and Arl3p Act in a Pathway that Interacts with Vesicle-Tethering Factors at the Golgi Apparatus. Current Biology, 13(5), 405–410. https://doi.org/10.1016/S0960-9822(03)00091-5

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature Microbiology, 2(11), 1533–1542. https://doi.org/10.1038/s41564-017-0012-7

Pasqualato, S., Renault, L., & Cherfils, J. (2002). Arf, Arl, Arp and Sar proteins: A family of GTP-binding proteins with a structural device for „front–back" communication. EMBO Reports, 3(11), 1035–1041. https://doi.org/10.1093/embo-reports/kvf221

Patel, M., Chiang, T.-C., Tran, V., Lee, F.-J. S., & Côté, J.-F. (2011). The Arf family GTPase Arl4A complexes with ELMO proteins to promote actin cytoskeleton remodeling and reveals a versatile Ras-binding domain in the ELMO proteins family. Journal of Biological Chemistry, 286(45), 38969–38979. https://doi.org/10.1074/jbc.M111.274191

Pereira-Leal, J. B. (2008). The Ypt/Rab family and the evolution of trafficking in fungi. Traffic (Copenhagen, Denmark), 9(1), 27–38. https://doi.org/10.1111/j.1600-0854.2007.00667.x

Petrželková, R., & Eliáš, M. (2014). Contrasting patterns in the evolution of the Rab GTPase family in Archaeplastida. Acta Societatis Botanicorum Poloniae, 83(4), 303–315. https://doi.org/10.5586/asbp.2014.052

Pipaliya, S. V., Schlacht, A., Klinger, C. M., Kahn, R. A., & Dacks, J. (2019). Ancient complement and lineage-specific evolution of the Sec7 ARF GEF proteins in eukaryotes. Molecular Biology of the Cell, 30(15), 1846–1863. https://doi.org/10.1091/mbc.E19-01-0073

Resh, M. D. (1999). Fatty acylation of proteins: New insights into membrane targeting of myristoylated and palmitoylated proteins. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1451(1), 1–16. https://doi.org/10.1016/S0167-4889(99)00075-0

Rojas, A. M., Fuentes, G., Rausell, A., & Valencia, A. (2012). The Ras protein superfamily: Evolutionary tree and role of conserved amino acids. Journal of Cell Biology, 196(2), 189–201. https://doi.org/10.1083/jcb.201103008

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Rosa-Ferreira, C., Christis, C., Torres, I. L., & Munro, S. (2015). The small G protein Arl5 contributes to endosome-to-Golgi traffic by aiding the recruitment of the GARP complex to the Golgi. Biology Open, 4(4), 474–481. https://doi.org/10.1242/bio.201410975

Roy, D., & Lohia, A. (2004). Sequence divergence of Entamoeba histolytica tubulin is responsible for its altered tertiary structure. Biochemical and Biophysical Research Communications, 319(3), 1010–1016. https://doi.org/10.1016/j.bbrc.2004.05.079

Roy, K., Jerman, S., Jozsef, L., McNamara, T., Onyekaba, G., Sun, Z., & Marin, E. P. (2017). Palmitoylation of the ciliary GTPase ARL13b is necessary for its stability and its role in cilia formation. Journal of Biological Chemistry, 292(43), 17703–17717. https://doi.org/10.1074/jbc.M117.792937

Sarma, G. N., Kinderman, F. S., Kim, C., von Daake, S., Chen, L., Wang, B.-C., & Taylor, S. S. (2010). Structure of D-AKAP2:PKA RI complex: Insights into AKAP specificity and selectivity. Structure (London, England : 1993), 18(2), 155–166. https://doi.org/10.1016/j.str.2009.12.012

Sato, K., & Nakano, A. (2007). Mechanisms of COPII vesicle formation and protein sorting. FEBS Letters, 581(11), 2076–2082. https://doi.org/10.1016/j.febslet.2007.01.091

Seidel, R. D., Amor, J. C., Kahn, R. A., & Prestegard, J. H. (2004). Conformational changes in human Arf1 on nucleotide exchange and deletion of membrane-binding elements. Journal of Biological Chemistry, 279(46), 48307–48318. https://doi.org/10.1074/jbc.M402109200

Setty, S. R. G., Shin, M. E., Yoshino, A., Marks, M. S., & Burd, C. G. (2003). Golgi Recruitment of GRIP Domain Proteins by Arf-like GTPase 1 Is Regulated by Arf-like

GTPase 3. Current Biology, 13(5), 401–404. https://doi.org/10.1016/S0960-9822(03)00089-7

Setty, S. R. G., Strochlic, T. I., Tong, A. H. Y., Boone, C., & Burd, C. G. (2004). Golgi targeting of ARF-like GTPase Arl3p requires its Nalpha-acetylation and the integral membrane protein Sys1p. Nature Cell Biology, 6(5), 414–419. https://doi.org/10.1038/ncb1121

Sharer, J. D., Shern, J. F., Van Valkenburgh, H., Wallace, D. C., & Kahn, R. A. (2002). ARL2 and BART Enter Mitochondria and Bind the Adenine Nucleotide Transporter. Molecular Biology of the Cell, 13(1), 71–83. https://doi.org/10.1091/mbc.01-05-0245

Shi, S.-P., Sun, X.-Y., Qiu, J.-D., Suo, S.-B., Chen, X., Huang, S.-Y., & Liang, R.-P. (2013). The prediction of palmitoylation site locations using a multiple feature extraction method. Journal of Molecular Graphics & Modelling, 40, 125–130. https://doi.org/10.1016/j.jmgm.2012.12.006

Schlacht, A., & Dacks, J. B. (2015). Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. Genome Biology and Evolution, 7(4), 1098–1109. https://doi.org/10.1093/gbe/evv045

Schmidt, H. A., Strimmer, K., Vingron, M., & von Haeseler, A. (2002). TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics (Oxford, England), 18(3), 502–504. https://doi.org/10.1093/bioinformatics/18.3.502

Schwartz, T., & Blobel, G. (2003). Structural Basis for the Function of the β Subunit of the Eukaryotic Signal Recognition Particle Receptor. Cell, 112(6), 793–803. https://doi.org/10.1016/S0092-8674(03)00161-2

Schweitzer, J. K., Sedgwick, A. E., & D'Souza-Schorey, C. (2011). ARF6-mediated endocytic recycling impacts cell movement, cell division and lipid homeostasis. Seminars in Cell & Developmental Biology, 22(1), 39–47. https://doi.org/10.1016/j.semcdb.2010.09.002

Smith, J. C., Northey, J. G. B., Garg, J., Pearlman, R. E., & Siu, K. W. M. (2005). Robust Method for Proteome Analysis by MS/MS Using an Entire Translated Genome: Demonstration on the Ciliome of Tetrahymena thermophila. Journal of Proteome Research, 4(3), 909–919. https://doi.org/10.1021/pr050013h

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research, 33(Web Server issue), W244–W248. https://doi.org/10.1093/nar/gki408

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L., & Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature, 521(7551), 173–179. https://doi.org/10.1038/nature14447

Stephen, L. A., & Ismail, S. (2016). Shuttling and sorting lipid-modified cargo into the cilia. Biochemical Society Transactions, 44(5), 1273–1280. https://doi.org/10.1042/BST20160122

Surkont, J., & Pereira-Leal, J. B. (2016). Are There Rab GTPases in Archaea? Molecular Biology and Evolution, 33(7), 1833–1842. https://doi.org/10.1093/molbev/msw061

Sztul, E., Chen, P.-W., Casanova, J. E., Cherfils, J., Dacks, J. B., Lambright, D. G., Lee, F.-J. S., Randazzo, P. A., Santy, L. C., Schürmann, A., Wilhelmi, I., Yohe, M. E., & Kahn, R. A. (2019). ARF GTPases and their GEFs and GAPs: Concepts and challenges. Molecular Biology of the Cell, 30(11), 1249–1271. https://doi.org/10.1091/mbc.E18-12-0820

Tamkun, J. W., Kahn, R. A., Kissinger, M., Brizuela, B. J., Rulka, C., Scott, M. P., & Kennison, J. A. (1991). The arflike gene encodes an essential GTP-binding protein in

Drosophila. Proceedings of the National Academy of Sciences of the United States of America, 88(8), 3120–3124. https://doi.org/10.1073/pnas.88.8.3120

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research, 44(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Turn, R. E., East, M. P., Prekeris, R., & Kahn, R. A. (2020). The ARF GAP ELMOD2 acts with different GTPases to regulate centrosomal microtubule nucleation and cytokinesis. Molecular Biology of the Cell, 31(18), 2070–2091. https://doi.org/10.1091/mbc.E20-01-0012

van Dam, T. J. P., Bos, J. L., & Snel, B. (2011). Evolution of the Ras-like small GTPases and their regulators. Small GTPases, 2(1), 4–16. https://doi.org/10.4161/sgtp.2.1.15113

Van Valkenburgh, H., Shern, J. F., Sharer, J. D., Zhu, X., & Kahn, R. A. (2001). ADP-ribosylation factors (ARFs) and ARF-like 1 (ARL1) have both specific and shared effectors: Characterizing ARL1-binding proteins. Journal of Biological Chemistry, 276(25), 22826–22837. https://doi.org/10.1074/jbc.M102359200

Vernoud, V., Horton, A. C., Yang, Z., & Nielsen, E. (2003). Analysis of the small GTPase gene superfamily of Arabidopsis. Plant Physiology, 131(3), 1191–1208. https://doi.org/10.1104/pp.013052

Vetter, I. R. (2014). The Structure of the G Domain of the Ras Superfamily. In Ras Superfamily Small G Proteins: Biology and Mechanisms 1 (s. 25–50). Springer, Vienna. https://doi.org/10.1007/978-3-7091-1806-1_2

Vichi, A., Moss, J., & Vaughan, M. (2005). ADP-ribosylation factor domain protein 1 (ARD1), a multifunctional protein with ubiquitin E3 ligase, GAP, and ARF domains. Methods in Enzymology, 404, 195–206. https://doi.org/10.1016/S0076-6879(05)04019-X

Vlahou, G., Eliáš, M., von Kleist-Retzow, J.-C., Wiesner, R. J., & Rivero, F. (2011). The Ras related GTPase Miro is not required for mitochondrial transport in Dictyostelium discoideum. European Journal of Cell Biology, 90(4), 342–355. https://doi.org/10.1016/j.ejcb.2010.10.012

Wang, X.-B., Wu, L.-Y., Wang, Y.-C., & Deng, N.-Y. (2009). Prediction of palmitoylation sites using the composition of k-spaced amino acid pairs. Protein Engineering, Design & Selection, 22(11), 707–712. https://doi.org/10.1093/protein/gzp055

Wilson, G. M., Fielding, A. B., Simon, G. C., Yu, X., Andrews, P. D., Hames, R. S., Frey, A. M., Peden, A. A., Gould, G. W., & Prekeris, R. (2005). The FIP3-Rab11 protein complex regulates recycling endosome targeting to the cleavage furrow during late cytokinesis. Molecular Biology of the Cell, 16(2), 849–860. https://doi.org/10.1091/mbc.e04-10-0927

Wirschell, M., Zhao, F., Yang, C., Yang, P., Diener, D., Gaillard, A., Rosenbaum, J. L., & Sale, W. S. (2008). Building a radial spoke: Flagellar radial spoke protein 3 (RSP3) is a dimer. Cell Motility and the Cytoskeleton, 65(3), 238–248. https://doi.org/10.1002/cm.20257

Wuichet, K., & Søgaard-Andersen, L. (2014). Evolution and Diversity of the Ras Superfamily of Small GTPases in Prokaryotes. Genome Biology and Evolution, 7(1), 57–70. https://doi.org/10.1093/gbe/evu264

Xie, Y., Zheng, Y., Li, H., Luo, X., He, Z., Cao, S., Shi, Y., Zhao, Q., Xue, Y., Zuo, Z., & Ren, J. (2016). GPS-Lipid: A robust tool for the prediction of multiple lipid modification sites. Scientific Reports, 6, 28249. https://doi.org/10.1038/srep28249

Xu, Y., Wang, Z., Li, C., & Chou, K.-C. (2017). iPreny-PseAAC: Identify C-terminal Cysteine Prenylation Sites in Proteins by Incorporating Two Tiers of Sequence Couplings into PseAAC. Medicinal Chemistry (Shariqah (United Arab Emirates)), 13(6), 544–551. https://doi.org/10.2174/1573406413666170419150052

Yang, Y.-K., Qu, H., Gao, D., Di, W., Chen, H.-W., Guo, X., Zhai, Z.-H., & Chen, D.-Y. (2011). ARF-like protein 16 (ARL16) inhibits RIG-I by binding with its C-terminal domain in a GTP-dependent manner. Journal of Biological Chemistry, 286(12), 10568–10580. https://doi.org/10.1074/jbc.M110.206896

Yoon, H. S., Nelson, W., Lindstrom, S. C., Boo, S. M., Pueschel, C., Qiu, H., & Bhattacharya, D. (2017). Rhodophyta. In J. M. Archibald, A. G. B. Simpson, & C. H. Slamovits (Ed.), Handbook of the Protists (s. 89–133). Springer International Publishing. https://doi.org/10.1007/978-3-319-28149-0_33

Yu, C.-J., & Lee, F.-J. S. (2017). Multiple activities of Arl1 GTPase in the trans-Golgi network. Journal of Cell Science, 130(10), 1691–1699. https://doi.org/10.1242/jcs.201319

Yutin, N., Wolf, M. Y., Wolf, Y. I., & Koonin, E. V. (2009). The origins of phagocytosis and eukaryogenesis. Biology Direct, 4, 9. https://doi.org/10.1186/1745-6150-4-9

Zahn, C., Hommel, A., Lu, L., Hong, W., Walther, D. J., Florian, S., Joost, H.-G., & Schürmann, A. (2006). Knockout of Arfrp1 leads to disruption of ARF-like1 (ARL1) targeting to the trans-Golgi in mouse embryos and HeLa cells. Molecular Membrane Biology, 23(6), 475–485. https://doi.org/10.1080/09687860600840100

Záhonová, K., Petrželková, R., Valach, M., Yazaki, E., Tikhonenkov, D. V., Butenko, A., Janouškovec, J., Hrdá, Š., Klimeš, V., Burger, G., Inagaki, Y., Keeling, P. J., Hampl, V., Flegontov, P., Yurchenko, V., & Eliáš, M. (2018). Extensive molecular tinkering in the evolution of the membrane attachment mode of the Rheb GTPase. Scientific Reports, 8(1), 5239. https://doi.org/10.1038/s41598-018-23575-0

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., & Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature, 541(7637), 353–358. https://doi.org/10.1038/nature21031

Zhang, Y., Huang, Y., Srivathsan, A., Lim, T. K., Lin, Q., & He, C. Y. (2018). The unusual flagellar-targeting mechanism and functions of the trypanosome ortholog of the ciliary GTPase Arl13b. Journal of Cell Science, 131(17):jcs219071. https://doi.org/10.1242/jcs.219071

Zhou, B., & Cox, A. D. (2014). Posttranslational Modifications of Small G Proteins. In Ras Superfamily Small G Proteins: Biology and Mechanisms 1 (s. 99–131). Springer, Vienna. https://doi.org/10.1007/978-3-7091-1806-1_5