

## RESEARCH ARTICLE OPEN ACCESS

# Eukaryote-Wide Distribution of a Family of Longin Domain-Containing GAP Complexes for Small GTPases

Anna M. G. Novák Vanclová<sup>1,2,3</sup>  | Catherine L. Jackson<sup>1</sup>  | Joel B. Dacks<sup>2,3,4,5</sup> 

<sup>1</sup>Université Paris Cité, CNRS, Institut Jacques Monod, Paris, France | <sup>2</sup>Division of Infectious Diseases, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada | <sup>3</sup>Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada | <sup>4</sup>Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic | <sup>5</sup>Department of Genetics, Evolution, & Environment, Centre for Life's Origins and Evolution, University College, London, UK

**Correspondence:** Catherine L. Jackson ([cathy.jackson@ijm.fr](mailto:cathy.jackson@ijm.fr)) | Joel B. Dacks ([dacks@ualberta.ca](mailto:dacks@ualberta.ca))

**Received:** 1 April 2025 | **Revised:** 23 June 2025 | **Accepted:** 10 July 2025

**Funding:** This work was supported by the Centre National de la Recherche Scientifique, France, and grant ANR-20-CE13-0007 from the Agence Nationale de la Recherche, France, to C.L.J. Research in the Dacks Lab is supported by the Natural Sciences and Engineering Research Council of Canada (RES0043758 and RES0046091).

**Keywords:** Arf GTPases | DENN domain | GTPase-activating proteins | LECA | molecular phylogenetics | nutrient signaling | pan-eukaryotic homology search | structural modeling

## ABSTRACT

Arf and Rab family small GTPases and their regulators, GTPase-activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs), play a central role in membrane trafficking. In this study, we focused on a recently reported GAP for Arf (and potentially Rab) proteins, the CSW complex, a part of a small family of longin domain-containing proteins that form complexes with GAP activity. This family also includes folliculin and GATOR1, which are GAPs for the Rag/Gtr GTPases. All three complexes are associated with lysosomes and play a role in nutrient signaling, the latter two being directly involved in the mTOR pathway. The role of CSW is not clear, but in addition to having GAP activity on Arf proteins *in vitro*, its mutation causes severe neurodegenerative diseases. Here we update the reported pan-eukaryotic presence of folliculin and GATOR1, and demonstrate that CSW is also found throughout eukaryotes, though with sporadic distribution. We identify highly conserved motifs in all CSW subunits, some shared with the catalytic subunits of folliculin and GATOR1, that provide new potential avenues for experimental exploration. Remarkably, one such conserved sequence, the “GP” motif, is also found in structurally related longin proteins present in the archaeal ancestor of eukaryotes.

## 1 | Introduction

The small GTPases are a large superfamily of highly conserved proteins that form the functional backbone of regulatory and signaling pathways involved in many cellular processes. Their function is dependent on guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs). GEFs are responsible for the exchange of bound GDP for GTP, converting the GTPase into its active form, while GAPs catalyze GTP

hydrolysis. For the Arf and Rab GTPases, the best-studied catalytic regions of GEFs are the Sec7 domain for the Arf family [1, 2] and the DENN/longin domain for Rab family proteins [3–6]. For Arf proteins, two well-established GAP families have been studied, those containing a Zn finger GAP domain [7, 8] and those with an ELMO domain [9]. For Rab proteins, the TBC domain is the most widespread GAP domain [10, 11]. Unlike the GTPases themselves, generally highly conserved, short, single domain proteins, their GEFs and GAPs usually contain multiple

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Traffic* published by John Wiley & Sons Ltd.

domains in addition to the catalytic domain (e.g., PH, SH3, C2, coiled-coil, BAR, ankyrin-repeat, WD40 or RUN [4, 10, 12]) and are much more varied in their size and architecture.

Arf and Rab GTPases, along with their specific GEFs and GAPs, play a central role in membrane trafficking [1, 13, 14]. An essential function of Arf proteins is in the formation of transport vesicles, while Rab proteins mediate vesicle navigation to the target membrane and fusion with it [1, 13, 14]. Arf and Rab GTPases carry out their functions by recruiting specific proteins to membranes in their active GTP-bound conformation. For example, Arf proteins recruit vesicle coat complexes to membranes to mediate vesicle budding, and Rab proteins recruit vesicle tethering complexes as an early step in vesicle fusion [1, 13]. In addition to vesicular trafficking between organelles, Arf and Rab family proteins also regulate other endomembrane functions, such as organelle-cytoskeleton interactions, organelle movement, and membrane lipid composition [2, 13, 15–17]. This fundamental and complex cellular organization is the staple of eukaryotic cells, as opposed to that in prokaryotes, which do not have extensive internal membrane structures. The closest known prokaryotic relatives of eukaryotes, the Asgard archaea (Asgardarchaeota), have a number of homologs of membrane trafficking proteins previously considered eukaryote-specific [18, 19]. The first cultured Asgardarchaeota organisms were shown to have a simple cell morphology with no observable internal compartmentalization [20, 21]. However, the recent first report of cultivated heimdallarchaeal cells describes infrequently observed internal membrane structures, whose size and frequency increase with higher nutrient concentration [22].

The Arf family, which includes not only Arf, Arf-like, and Sar GTPases in eukaryotes [15], but also the Arf-related proteins (ArfRs) present in the Asgard archaea, is ancient, having arisen before the divergence of eukaryotes from their archaeal ancestor [23, 24]. Given their wide spectrum and high specificity of their interactions, the expansion of the Arf family, and their GEFs and GAPs, likely played a crucial role in the evolution of the endomembrane system, especially the emergence and specialization of its parts [8, 25–27]. All of the above GTPases, GAPs, and GEFs have been shown to be widely distributed, and in fact expanded to multiple paralogues, across eukaryotes, from animals to fungi, plants, algae, and parasites [15]. Studying these proteins and their domains and enzymatic functions in a broader, pan-eukaryotic context is a prerequisite to better understanding of the biology of diverse eukaryotes and early eukaryotic evolution.

One set of potential Arf regulators, however, has not yet been examined systematically using this comparative evolutionary lens. The C9orf72:SMCR8:WDR41 or CSW complex, comprising three subunits, C9orf72, SMCR8, and WDR41, was first described in humans, where mutations in one of its subunits (noncoding region expansions in C9orf72) cause severe neurodegenerative diseases (amyotrophic lateral sclerosis, ALS, and frontotemporal dementia, FTD; [28–30]). It has been identified as a multisubunit GAP and experimentally shown in vitro to act most efficiently upon Arf1, Arf5, and Arf6, and to have some, albeit much weaker, activity toward Rab8 and Rab11 [31–33]. It is considered a member of a small family of longin domain-containing GAP complexes, all of which localize to the lysosomal membrane and play a role in nutrient sensing

[31, 32, 34]. The other two members of this family are the folliculin (FLCN:FNIP) and GATOR1 (NPRL2:NPRL3) complexes that act upon Rag GTPases and work in tandem as a part of the mTOR signaling cascade, a major control hub of cell growth [31, 35–38]. The CSW complex is recruited to the lysosomal membrane via interaction between its subunit WDR41, which contains a beta-propeller domain, and the PQ loop-containing cationic amino acid transporter PQLC2 [35, 39, 40]. This binding is inhibited when the transporter substrate (arginine, histidine, and lysine) is abundant, suggesting that the CSW complex might mediate some kind of cross-talk between Golgi- or endosome-derived membranes and lysosomes under cationic amino acid starvation. The exact function, however, is currently not understood.

The catalytic pocket of the CSW complex is formed by two symmetrically arranged longin domains of different subunits (C9orf72 and SMCR8) with known essential residues of both required for GAP activity (specifically W33-D34 in C9orf72, and R147, the catalytic arginine, referred to as an arginine finger, in SMCR8 [32]). In folliculin and GATOR1, the subunits containing the catalytic arginine are FLCN and NPRL2, respectively, present within their longin domains, with the FNIP2 and NPRL3 subunits also containing longin domains [31]. The evolutionary relationship and degree of sequence homology between these six proteins are yet to be established and examined in detail. However, previous studies group them together based on their domain architecture and secondary structure and consider them related to or a part of the DENN family [5, 31, 34]. Previous studies [34, 41] found evidence for SMCR8/FLCN-type longin proteins and/or FNIP in *Capsaspora*, some fungi, amoebozoans, metamonads, and discobids, and one or both GATOR1 subunits in the previously mentioned groups plus several apicomplexans and stramenopiles, and a few more homologs are currently annotated as such in the Uniprot database (<https://www.uniprot.org/>, [42]). By contrast, little information is available regarding the distribution and function of the CSW proteins outside mammals. Zhang et al. [34] identified a few nonmetazoan homologs using a limited sampling of genomic data available at the time, namely C9orf72 in some choanoflagellates and amoebozoans, *Trichomonas*, and *Naegleria*. It is likely that this unique family of GAPs occurs more widely and could potentially be traceable to the root of a large subsection of the eukaryotic tree of life spanning very diverse organisms, to the last eukaryotic common ancestor (LECA), or even further. An updated, comprehensive eukaryote-wide screening may therefore reveal distribution patterns that could shed some light on the early evolution of the organization and crosstalk between lysosomes and other endomembrane compartments.

In this study, we took advantage of the rich supply of available sequence data for nonmodel eukaryotic groups, including previously obscure and recently discovered deep-branching lineages, and performed a pan-eukaryotic search for each of the subunits of the CSW complex, other longin GAPs, and their potential interaction partners to observe patterns in their distribution and co-occurrence, and attempt to trace their point of origin and evolutionary history using phylogenetic methods. Structural modeling and primary sequence analysis confirmed the conservation of known functionally critical amino acid positions and identified additional highly conserved amino acid residues which likely have functional importance. Intriguingly, some of

these residues are also shared with longin domain proteins encoded in asgardarchaeal metagenomes.

## 2 | Results and Discussion

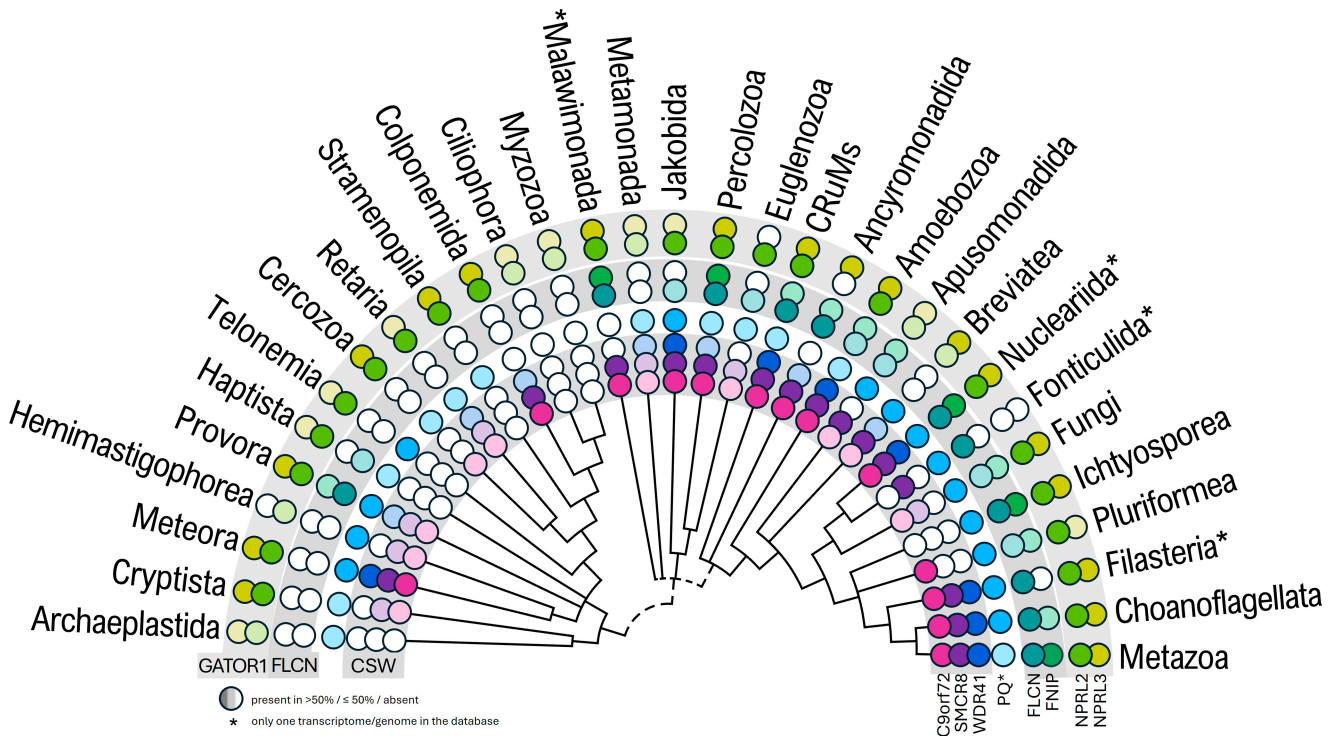
### 2.1 | Longin GAP Complexes Have Wide but Patchy Distribution

We investigated the distribution of CSW, folliculin, and GATOR1 complex proteins across all major eukaryotic lineages using an enriched and modified TCS database of genomes and transcriptomes and sensitive sequence homology detection (see Section 4, Table S1). The resulting homolog datasets, including e-values of the hits, taxonomical annotation, and other notes can be found in Tables S2–S8.

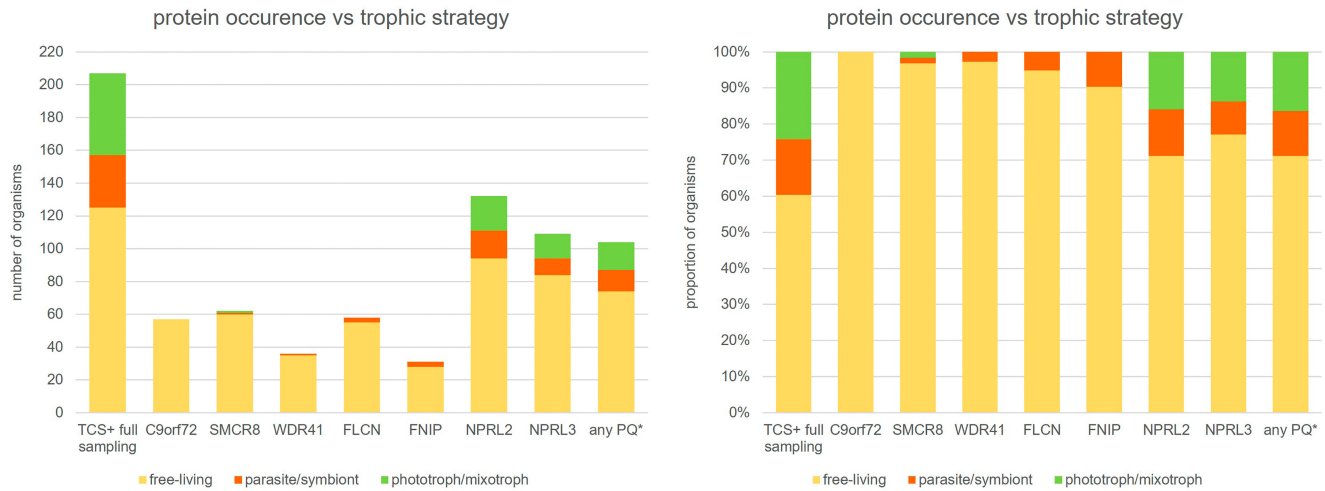
We report that subunits of the CSW complex, which have been studied in humans and annotated in various animals in the Uniprot database, are present in many other eukaryotic lineages. However, their distribution is patchy as it seems to be completely missing from a number of major lineages, especially among Diaphoretickes (i.e., Archaeplastida, Haptista, Telonemia, Stramenopila, Ciliophora, and Myzozoa; Figure 1) and as it is not universally conserved in lineages in which it has been identified. The distribution of the folliculin complex exhibits a similar pattern, with its depletion among Diaphoretickes being even more noticeable (present or partially present only

in some Provora and Haptista). GATOR1, on the other hand, is at least partially retained in all lineages (with the only exception of Fonticulida which is, however, represented by a single, possibly incomplete, dataset). In the context of the currently accepted topology of the eukaryotic tree of life (outlined in figure 1, [43–45]), including the recently proposed position of its root [46], these findings suggest that some form of all three complexes was already present in LECA and that its patchy distribution is the result of multiple secondary losses. The scenario explaining this distribution pattern could be simplified by the recently proposed changes to the eukaryotic tree, which place Provora, *Meteora*, and Hemimastigophorea (i.e., some of the groups where CSW and/or folliculin is retained to some degree) as sister to Diaphoretickes instead of nested within them [47]. If genuine, this relationship would strongly suggest there was a single instance of loss of CSW and folliculin complexes in the common ancestor of Diaphoretickes and a few cases of secondary gain of some subunits via LGT in isolated lineages of this group (e.g., Colponemida).

Little is known about the function of these complexes outside model organisms and, in the case of CSW, even the function in humans remains unclear. The common denominator, however, seems to be a role in nutrient signaling [31]. Trophic strategy may therefore represent a contributing factor to the evolutionary pressures that led to the loss or retention of these proteins in different lineages. While not universally conserved, GATOR1 proteins seem to occur at a relatively similar



**FIGURE 1** | The distribution of the three longin GAP complexes across major eukaryotic lineages. From the center outward, CSW: C9orf72 (magenta), SMCR8 (purple), and WDR41 (blue), any PQ-loop protein homologous to PQLC2 (i.e., potential interaction partners of CSW; sky blue), folliculin: FLCN (teal) and FNIP2 (cold green); GATOR1: NRPL2 (warm green) and NPRL3 (chartreuse). Darker hue of the respective color indicates the protein is present in majority of the investigated datasets, paler hue indicates it is present sporadically, while white denotes no reliable evidence of its presence. The tree topology is conservatively synthesized from the recent pan-eukaryotic phylogenies [43–45], the possible position of the root is along the branches denoted by dashed lines. The asterisks denote evolutionarily significant but poorly sampled lineages that were represented by a single genome or transcriptome in our database.



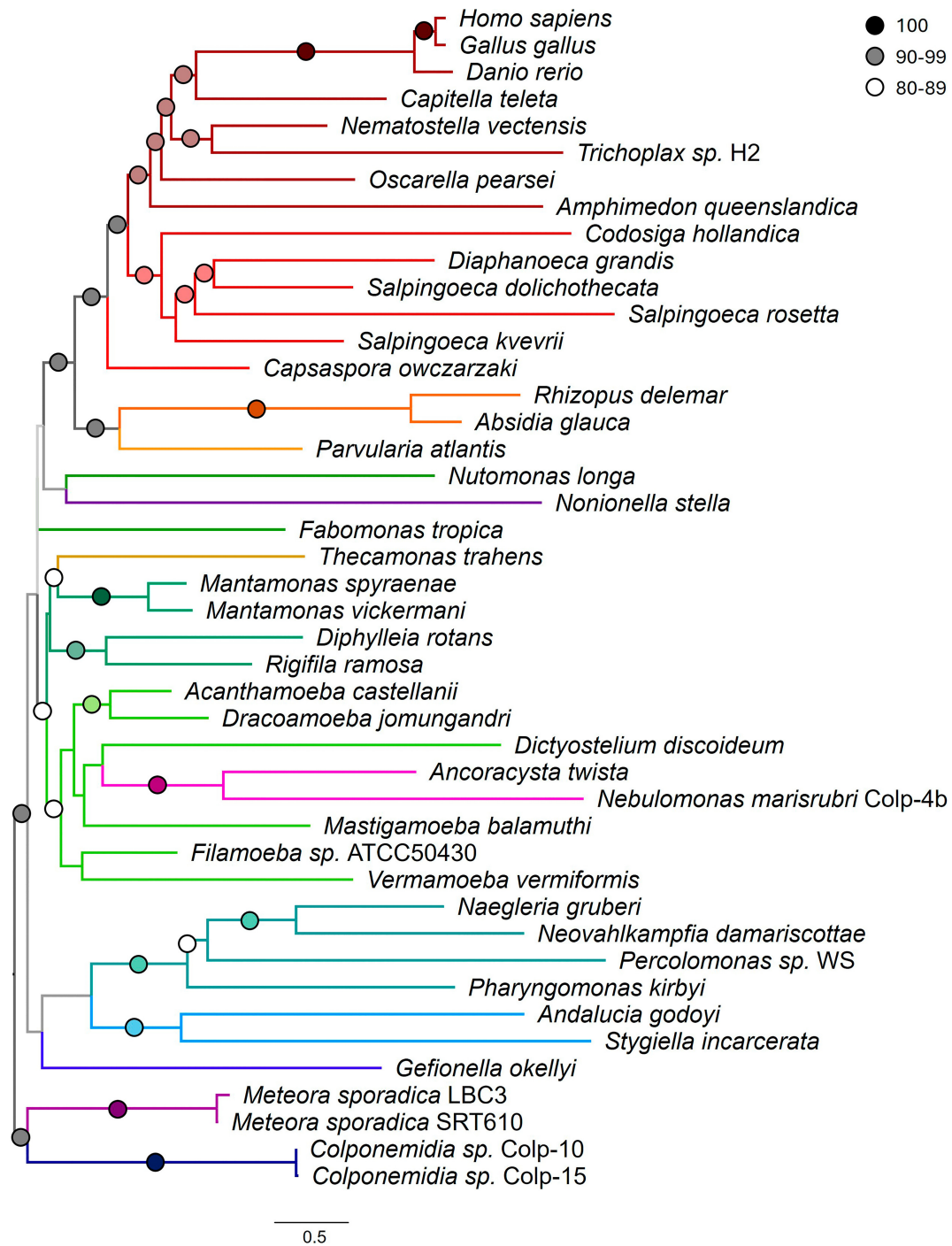
**FIGURE 2** | Occurrence of the proteins of interest across eukaryotes broken down by trophic strategy. All organisms in our database were sorted into three categories: Phototrophic or mixotrophic (green), parasitic or symbiotic (orange), and free-living heterotrophic (yellow) (Table S1). The composition of the database in regard to these categories is shown as the first column in both bar plots (as counts and percentages, respectively). The remaining columns shows their ratio in the subsets of organisms in which the respective protein has been identified. GATOR1 subunits and PQLC2/YPQ are more widespread and seem to be distributed equally across trophic categories. CSW and folliculin complex proteins, on the other hand, occur more sporadically and show a bias toward free-living heterotrophic organisms.

rate in free-living heterotrophs (178 of 250 possible orthologues in 125 taxa, 71%), parasitic/symbiotic heterotrophs (27 of 64 possible orthologues in 32 taxa, 42%), and phototrophs/mixotrophs (36 of 100 possible orthologues in 50 sampled taxa or 36%) (Figure 2, Table S1). This is not the case for CSW and folliculin complex proteins, whose presence is biased toward free-living heterotrophs (41% and 33% of possible retained orthologues, respectively, compared to 2% and 9% retained in parasites/symbionts and 1% and 0% in phototrophs/mixotrophs). For example, while all three subunits are conserved in most free-living Amoebozoa, the only parasitic representative in our sampling, *Entamoeba histolytica*, lacks all of them. Similarly, the same subunits are absent from phototrophic/mixotrophic organisms with the exception of an SMCR8 homolog in *Euglena gracilis*.

Given their roles in amino acid sensing and signaling (demonstrated for folliculin and GATOR1 which are involved in the mTOR pathway, and presumed for CSW due to its interaction with cationic amino acid transporter PQLC2 and recruitment to lysosomes upon amino acid starvation [39]), this may reflect the difference in amino acid cost and availability under different evolutionary strategies. Photo- and mixotrophs generally synthesize most of their amino acids themselves. Parasites, on the other hand, often live in relative nutrient abundance and can therefore afford to streamline their energy metabolism at the cost of energetic efficiency or supplement it with alternative energy sources, including amino acids [48, 49]. Both of these situations may drive changes in amino acid signaling pathways, potentially resulting in fewer steps and proteins involved. At the same time, the possibility of some of these results being false negatives and caused by limited sampling or transcriptome incompleteness should not be dismissed (BUSCO scores of all datasets are given in Table S1). It should also be noted that the final dataset for WDR41 may be more affected by false negative results due to the stricter e-value threshold used in homology mining (see Section 4).

In conclusion, our analysis indicates that all three longin domain-containing GAP complexes were already present in LECA, at least in some form. GATOR1 is the most widely distributed among eukaryotes, across taxonomic groups and trophic strategies, although some secondary losses have occurred. CSW and folliculin, on the other hand, appear to be somewhat confined to the nondiaphoretic half of the eukaryotic tree of life, which includes animals, fungi, and amoebozoa, in addition to a number of protist groups. They are notably absent from plants and other phototrophs across eukaryotic diversity and more frequently found in free-living heterotrophs than in parasites and symbionts. This pattern of conservation could shed light on their functions, given the role of all three complexes in nutrient sensing in mammalian cells.

We also searched for homologs of PQLC2 in order to investigate its co-occurrence with the CSW complex. Our results show that they occur throughout eukaryotes, often in two or more divergent paralogues and in a patchy pattern. Our phylogenetic analyses suggest complex relationships between these homologs and an unclear history of duplications resulting in multiple differentially retained clades (Figures S1 and S2, Table S9, Data S1). These also include YPQ2 and YPQ3, vacuolar arginine/histidine transporters reported in yeast [50] which underlines the notion that the identified proteins likely have various localizations that cannot be discerned without experimental evidence. We were not able to identify a notable pattern of co-occurrence between specific clade(s) and the CSW complex or WDR41 (i.e., the subunit interacting with PQLC2 in human). However, the possibility of interaction between the human CSW complex and other, heterologous PQLC2 homologs, such as the yeast YPQs, may be worth experimental investigation and may have broader implications for the CSW complex function and its evolution throughout eukaryotes. For instance, the existence of organisms possessing CSW but lacking any identifiable PQLC2 homolog (e.g., many of the sampled Amoebozoa, some CRuMs, Ancyromonadida, and Colponemidia; see

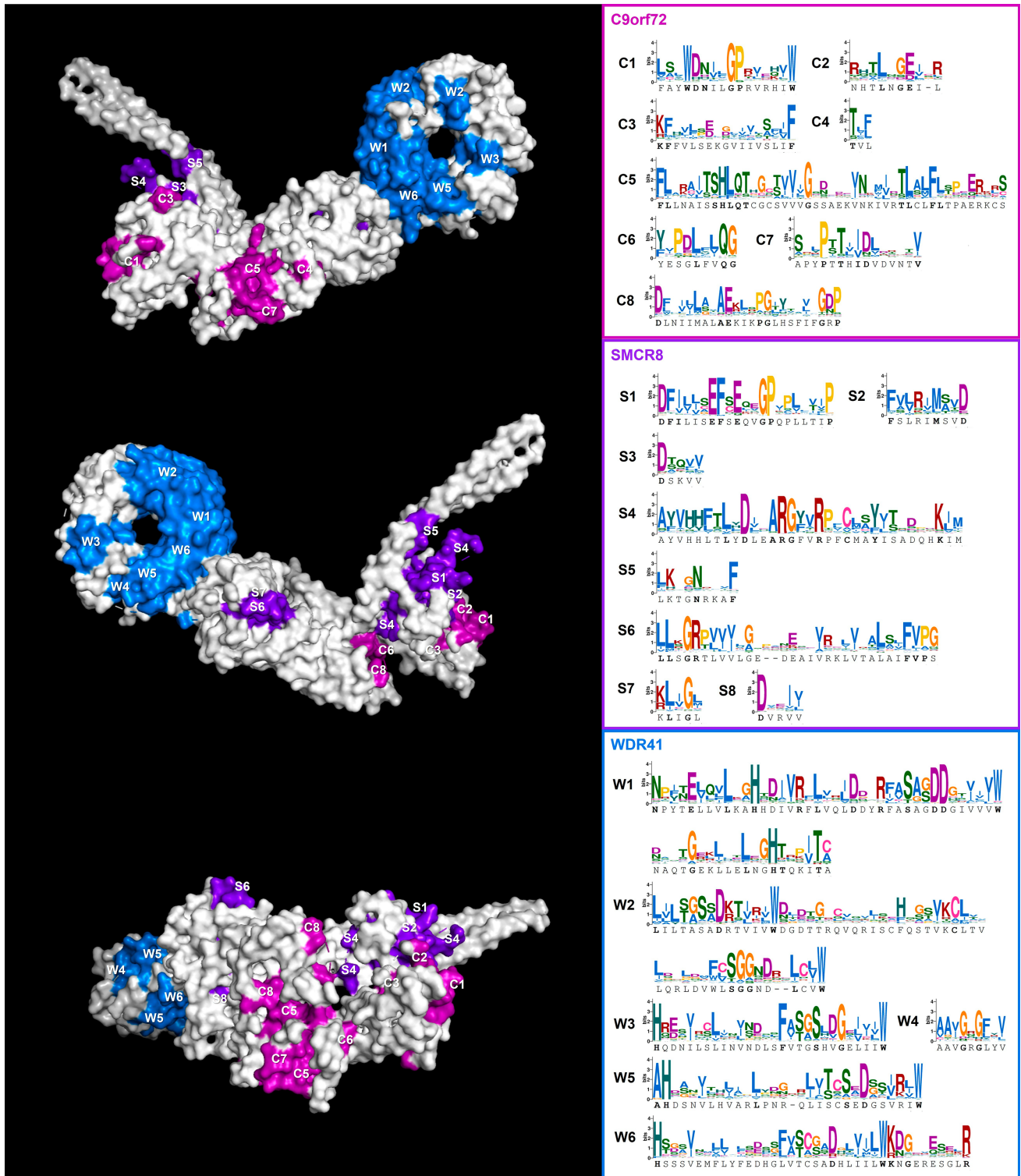


**FIGURE 3** | Concatenated phylogeny of the three subunits of the CSW complex. The dark, light, and colorless circles denote bootstrap support of 100, 90–99, and 80–89, respectively, the branches without circles have lower support. The colors delineate main taxonomic groups: Holozoa (shades of red), Holomycota (shades of orange), Apusomonadida (gold), Amoebozoa (light green), CRuMs (green-cyan), Percolozoa (dark cyan), Jakobida (sky blue), Malawimonada (blue), Provora (pink), and other Diaphoretickes (shades of purple, navy) (see Figure S7 for detailed color legend).

Table S1) suggests their CSW complex is able to interact with an alternative partner, be it on lysosome or elsewhere, or that it performs its function exclusively in a soluble form. These results show that the receptor on membranes for CSW in humans is also widely distributed among eukaryotes, but is not always present in organisms that have CSW complex proteins, suggesting that other receptors might exist. Identification of such alternative membrane receptors could lead to new insights into the functions of the CSW complex.

## 2.2 | The Inter-Relationships Between CSW, Folliculin, and GATOR1 Proteins

The six longin proteins of CSW, folliculin, and GATOR1 complexes were previously reported as a distinct family of DENN domain-containing proteins based on domain organization [34] and structural and functional similarity [31, 32]. In order to verify this notion and further investigate their relationships in a broader context, we mainly focused on the longin domains



**FIGURE 4** | Reference CSW complex structure with conserved regions highlighted. Conserved sequence motifs were manually selected from the pan-eukaryotic alignments of the three subunits and are visualized as sequence logos on the right and shown mapped onto the human structure shown in three different orientations; the three subunits are distinguished by colors and letter IDs. Regions C1–C3 and S1–S5 are parts of the longin domains of their respective proteins, while C4–C8 and S6–S8 occur downstream of it. The conserved regions of WDR41 shown are all equivalent to, or part of, the WD repeats with the exception of W4 which falls between the fourth and fifth repeat. The exact positions and other details are given in Table S13.

themselves and employed several parallel approaches (based on structure modeling, sequence alignment and similarity, and phylogenetics; see Section 4).

Our analyses clearly identified significant conservation of structure and/or sequence in particular regions of particular proteins (see Section 2.4). Although we were not able to

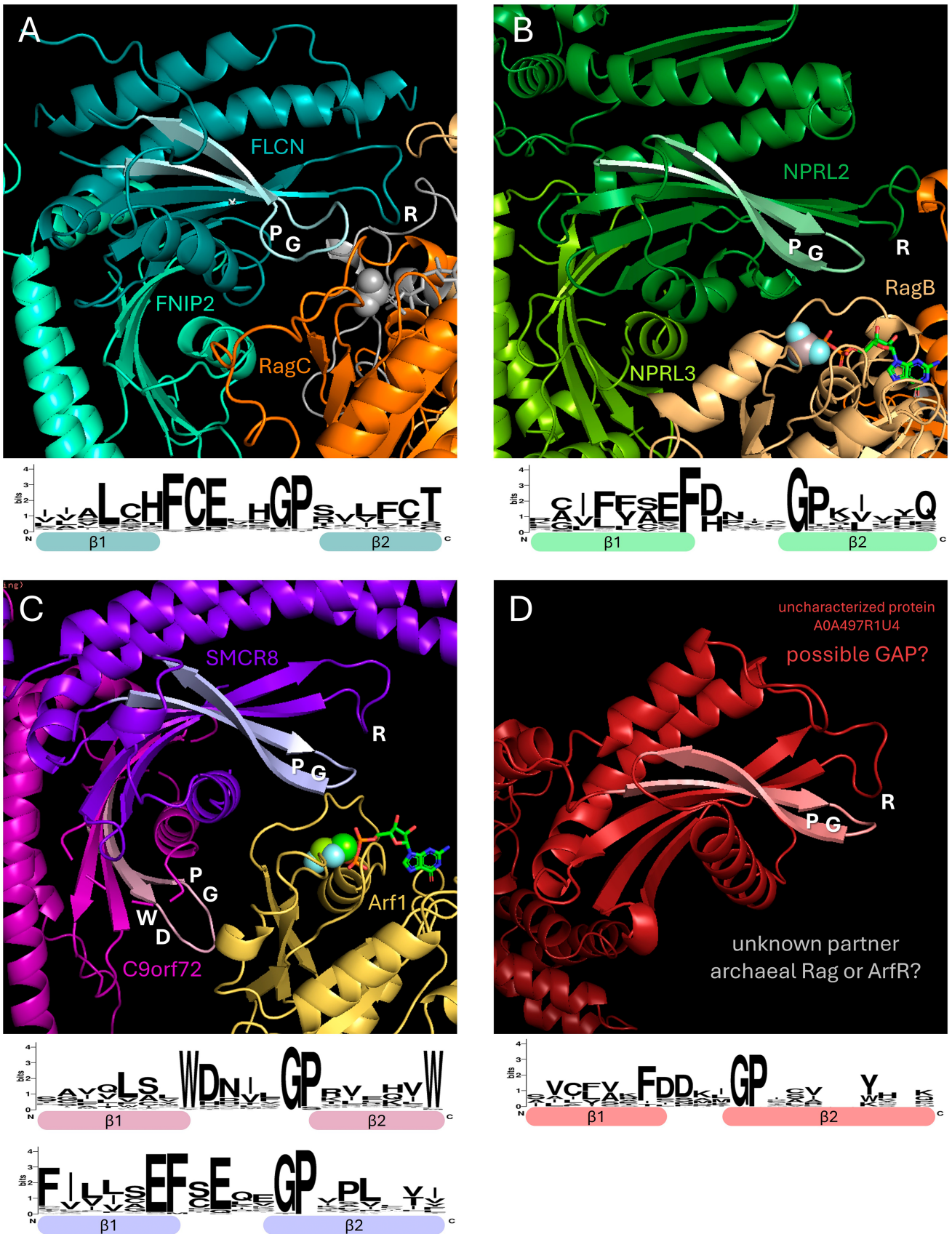


FIGURE 5 | Legend on next page.

**FIGURE 5** | Conservation of the GP motif and arginine finger of the eukaryotic longin domain-containing GAP complexes and structurally similar Asgardarchaeota longins. Detail of the catalytic pocket of human folliculin (A), GATOR1 (B), and CSW (C) complexes, including their interaction partners (RagC, RagB, and Arf1, respectively) and bound GDP, with arginine fingers, GP motifs, and WD motif of C9orf72 highlighted. Sequence logos for the GP motif-containing regions denoted by lighter color shade (first and second beta sheets of the longin domain and the loop between them) based on our pan-eukaryotic sampling are shown underneath. A reference heimdallarchaeal FLCN/SMCR8-type longin domain-containing protein and a logo based on 14 other structurally related Asgard longins with previously published predicted structures are shown in (D), underscoring their remarkable similarity to the eukaryotic longins with GAP activity. The subunit color coding is as per Figure 1 (FLCN: teal, FNIP2: cold green, NRPL2: warm green, NPRL3: chartreuse, C9orf72: magenta, SMCR8: purple) with additional colors for the client GTPases (RagB: wheat, RagC: orange, Arf1: yellow) and Asgard longin protein (red).

consistently recover the proposed family as a clearly defined clade with resolved inner relationships, we identified several interesting homology relationships and cases of significant conservation of structure and/or sequence. All versus all BLAST results revealed sequence homology of SMCR8 and FLCN in the longin domain. Conversely, C9orf72 exhibits some homology to SMCR8 and FNIP2 outside the longin domain and weak similarity to NPRL2 in the longin domain. NPRL2 and NPRL3, on the other hand, robustly retrieve each other in all searches (Figure S3, Table S10). Interestingly, this is not corroborated by pairwise alignment-based analysis of sequence identities and similarities of the full protein sequences (Table S11), further suggesting the homologous regions are scattered and full-length alignment of different GAP longin proteins is problematic. Moreover, when analyzed based on predicted secondary structure, different patterns emerge (Figures S4 and S5). For instance, the abovementioned pairs of longin domains based on sequence similarity (NPRL2/NPRL3 and SMCR8/FLCN) do not exhibit prominent secondary structure similarity, and the GAP longin family as a whole seems less distinct from some other DENN proteins in our sampling. Our results indicate that among the members of the GAP longin family, specific protein regions have clear relationships between them but that complexity has been introduced, potentially by instances of gene fusions, evolutionary convergence, or simply artefacts stemming from paucity of phylogenetic signal.

### 2.3 | Phylogenetic Signal of CSW Supports Its Presence in LECA and Vertical Inheritance

In order to better understand the evolutionary history of the individual complex subunits, we built single-gene unrooted trees for each of the proteins of interest (Figures S6.1–7) and a tree based on a concatenated matrix of the three subunits of the CSW complex (Figure 3). While the single-gene tree topologies have generally low support, major well-sampled taxonomic groups, such as Metazoa, Fungi, Stramenopila, Percolozoa, etc., cluster in most cases. The concatenated tree for CSW subunits has low backbone support, but its overall topology is consistent with the expected relationships between the sampled taxonomic groups. Opisthokonts are recovered as a monophyletic group with resolved and well-supported inner topology, and so are discobids. Amoebozoa form a well-supported monophyletic cluster with Provora, represented by *Ancoracysta twistata* and *Nebulomonas marisrubri*, nested inside (the attraction of these groups is also apparent in the single gene trees, Figures S6.1 and S6.3). This represents the only well-supported phylogenetic relationship incongruent with pan-eukaryotic topology and a likely case of

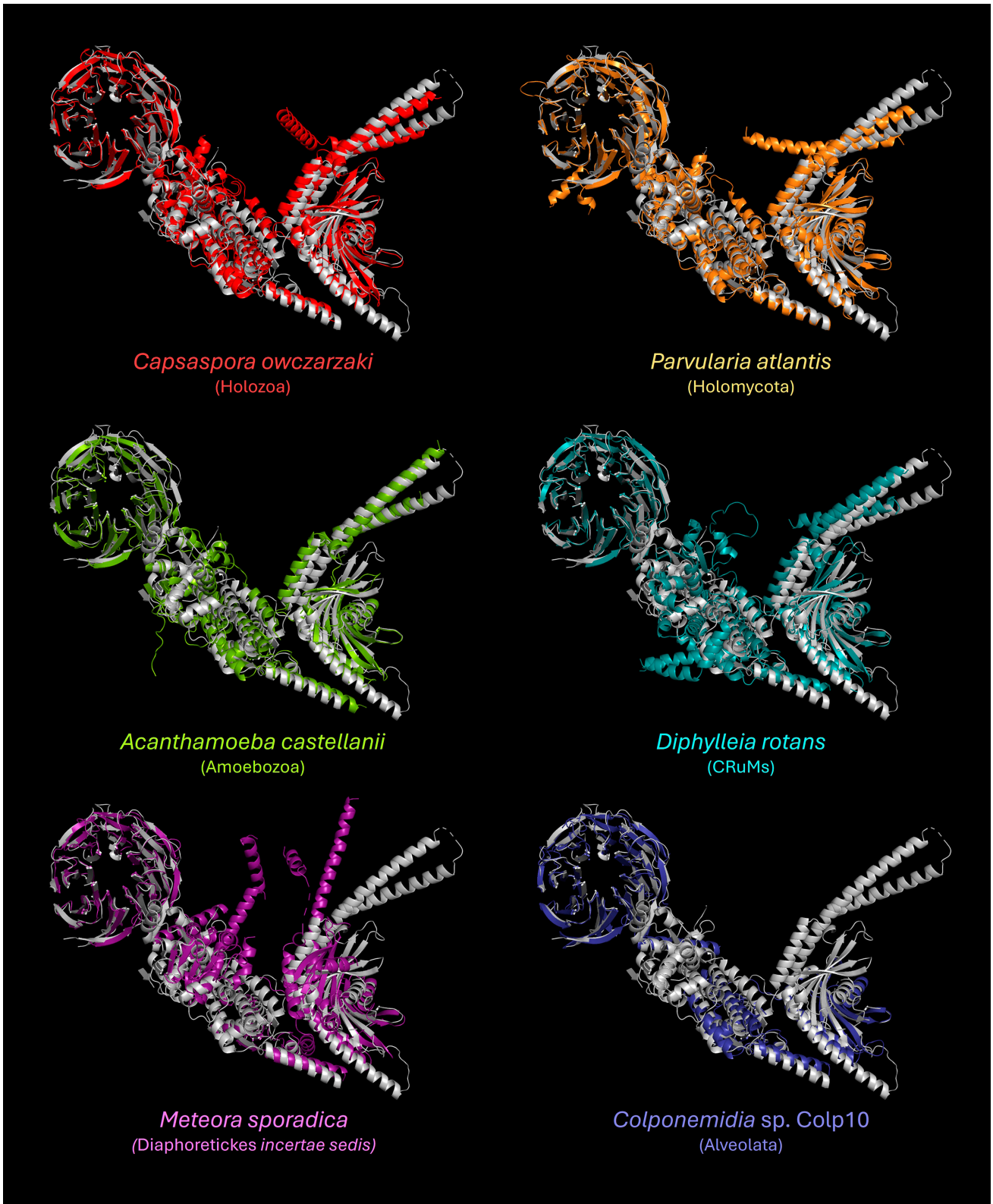
lateral gene transfer (LGT) from Amoebozoa to Provora, possibly related to the phagotrophic predatory life strategy of the latter. We conclude this is unlikely to be the result of contamination, as the two provorans cluster together but form separate branches of regular length. This provides a new context for the notable absence of the folliculin complex in all diaphoretickes except Provora, begging the question of whether their folliculin could also be the result of LGT. It is also worth noting that Provora were recently proposed to be sister to other diaphoretickes [47]. Either of these scenarios could greatly reduce the number of secondary losses required to reconcile its observed distribution with its origin in LECA.

### 2.4 | Eukaryote-Wide Comparison Reveals Novel Motifs of Interest

Thanks to the pan-eukaryotic sampling of the CSW proteins, we were able to gain a broader perspective on the relative conservation of their structural and sequence motifs (Figure 4). The essential residues identified by previous studies in humans, namely the W33-D34 of C9orf72 and the arginine finger (R147) of SMCR8, are almost universally conserved across other eukaryotes (C1 and S4 in Figure 4), suggesting the catalytic function is conserved despite the significant divergence in other parts of the protein chains. In addition, we identified a number of other widely conserved residues or motifs that were not previously reported or studied and that may represent potential targets of future experiments aiming to gain a deeper understanding of the role of the CSW complex in human cells as well as other organisms.

Sites of notable sequence conservation are not limited to the longin domain of C9orf72 and SMCR8, but also occur in the downstream part of the protein chain forming the middle section of the complex. The significance of these sites, for example whether some of them (e.g., the outward facing C5, C7, and S6 in Figure 4) could represent binding interfaces for additional molecules, remains to be determined by mutation experiments.

The most remarkable example is the GP motif in the catalytic longin domain that occurs in both C9orf72 and SMCR8 (G38-P39 and G68-P69 in the reference human sequences; C1 and S1 regions in Figure 4), as well as FLCN and NPRL2 but not in FNIP2, NPRL3, and other non-GAP longin proteins such as adaptins, TRAPP complex subunits, or DENND family proteins. Structurally, it represents either the C-terminal part of the loop immediately preceding the second beta sheet of the longin domain or the N-terminal part of this beta sheet.



**FIGURE 6** | CSW complex secondary structure comparison between diverse eukaryotes. Structures modeled by AlphaFold (in color) are aligned to cryo-EM-based reference structure of the human CSW complex (white). SMCR8 subunit of *Colponemidia* sp. Colp10 is not shown due to its high divergence and misalignment. WDR41 subunit and the longin domains of the catalytic subunits exhibit higher structure conservation, especially across the nondiaphoretickes (*Capsaspora*, *Parvularia*, *Acanthamoeba*, and *Diphyllia*) homologs, while the middle section of the complex is more divergent or possibly completely disrupted in some cases (i.e., SMCR8 of *Meteora* and Colp10).

As such, it clearly occupies a central position in the catalytic pocket and spatially neighbors the arginine finger, an essential motif present and universally conserved not only on SMCR8 but also FLCN and NRPL3 (Figure 5A–C). The GP motif of C9orf72 is positioned symmetrically to the one in SMCR8 and is closely adjacent to the abovementioned essential and highly conserved WD motif. This analysis, based on structural predictions, strongly supports the notion that the GP motif plays an important role in the function of all three complexes and, given its specific presence in this family and absence from other longins, may be at least partially responsible for them carrying out GAP activity instead of GEF activity, which is more typical of longin proteins. Together, these analyses make this novel motif an attractive prospective target for future experiments, for example, to test the effect of its mutation on enzyme kinetics and specificity.

## 2.5 | Secondary Structure of CSW Complexes From Diverse Eukaryotes

A number of studies have provided extensive structural information for the GATOR1 and folliculin complexes, including comparative structural analyses of yeast and human complexes [31, 35–37], but no comparative structural analyses have been reported for CSW. As a first step to obtaining information about the structures of the CSW complex across eukaryotes, we selected a small subset of taxonomically representative organisms with all three subunits identified, modeled their secondary structures using AlphaFold (Figures S8–S10), and aligned them to human reference (Figure 6) and to each other (Table S12). This sampling included *Capsaspora owczarzaki*, a protist lineage closely related to animals, *Parvularia atlantis*, a nuclearioid amoeba closely related to fungi, *Acanthamoeba castellanii*, an amoebozoan, *Diphylleia rotans*, a deep-branching flagellate distantly related to Amoebozoa and Opisthokonta, *Meteora sporadica*, a deep lineage of Diaphoretickes, and *Colponemidia* sp. Colp10, an alveolate. In this dataset, WDR41 and C9orf72 seem to be structurally conserved across the different organisms (TM-scores generally between 0.8 and 0.9), but significant divergence is observed in the case of SMCR8 where the best alignments do not exceed a TM-score of 0.8 while the lowest-scoring ones fall below 0.5 (Table S12). This is congruent with the fact that these models by themselves are of lower overall confidence, usually exhibiting low PAE scores outside the longin domain (Figure S9). The structure of SMCR8 of *Capsaspora*, *Parvularia*, *Acanthamoeba*, and *Diphylleia* seems to be well-conserved around the catalytic site but diverges, to various degrees, in the second and third alpha helices of the longin domain and further downstream. In *Meteora* and Colp10, the two representatives of Diaphoretickes in our sample, the structure of SMCR8 may be even more disrupted, even in the catalytic longin domain, which brings up the question of whether this subunit can function in a complex with C9orf72 and carry out the same function as its counterparts in other lineages. A functionally divergent SMCR8 is plausible, considering that the CSW complex is not present in the vast majority of Diaphoretickes, possibly due to secondary loss following a divergence in or loss of function in their common ancestor. Such a proposition would, however, require more detailed analysis, including in vitro experiments, for example, to test whether the divergent subunits form a complex and to determine their cellular localization.

Nevertheless, our results indicate that the secondary structures of the longin domains of C9orf72 and SMCR8, which form the catalytic pocket, and the WDR41 subunit, which mediates interaction with lysosome, are relatively highly conserved across diverse nondiaphoretickes, which suggests retention of the GAP and lysosome-interacting function of their CSW complex.

## 2.6 | Conserved Sequence Motifs Connect Eukaryotic GAP Longins to a Specific Class of Asgard Longins

The conservation of the GP motif and arginine finger in components of all three of the eukaryotic complexes prompted us to look for deeper evolutionary origins of the GAP catalytic subunits. Asgard archaea are known to encode expanded complements of longin proteins [26, 51], and indeed there are sequences and predicted structures previously deposited in the Uniprot and AlphaFold databases that exhibit high structural similarity to GAP longins, in some cases explicitly annotated as “Folliculin/SMCR8 longin domain-containing” (Figure 5D; protein sequences published in [19, 52–56] can be found in Data S1). Strikingly, the GP motif and arginine finger in the same structural arrangement are also present in these longins. This conservation and their notable presence in Heimdallarchaea, the proposed closest Asgard relatives of eukaryotes [18, 57], further support their annotation as homologous and specifically related to the eukaryotic GAP longins. It also raises the possibility that these Asgard longin domains and the eukaryotic GAP longins share a common ancestor, and that they may potentially share GAP activity. CSW and folliculin/GATOR1 act as GAPs for Arf and Rag proteins, respectively, and Asgard archaea have previously been demonstrated to possess both Arf-related (ArfR) [23] and Rag/Gtr GTPases [23, 26, 58] which begs the question of whether either, or possibly both of these families could represent interaction partners of the GP motif-containing Asgard longins. Furthermore, if these longins are indeed evolutionarily related and could carry out similar enzymatic function, the fact that the eukaryotic members of this hypothetical family are all specifically associated with lysosomes could be of great significance in regard to the possible origin and ancestral forms of the eukaryotic endomembrane system, some traits of which are already present in Asgards, and particularly, Heimdallarchaea [20, 22, 58, 59].

## 3 | Conclusion

We searched for homologs of the subunits of the longin domain-containing GAP complexes with an updated database containing genomes and transcriptomes of representatives spanning most currently known lineages of eukaryotes. We report that homologs of the GATOR1 proteins are present in some representatives of all groups, suggesting that the complex was already present in LECA. This result is congruent with the universal presence of the TOR signaling pathway in eukaryotes [60, 61]. On the other hand, we only find evidence of the folliculin and CSW complexes in roughly half of the examined lineages and note that both seem to be absent in the vast majority of diaphoretickes. This distribution is even more prominent in the case of folliculin. Moreover, both complexes exhibit patchy distribution favoring free-living heterotrophic organisms as opposed to

phototrophs/mixotrophs and parasites/symbionts, likely reflecting the very different dynamics of nutrient availability under different lifestyles. Nevertheless, considering the currently accepted pan-eukaryotic tree topology, including the recently proposed position of its root within “excavates,” some form of all three complexes was also likely present in LECA but later lost in multiple lineages. The disparity in the occurrence of GATOR1 and folliculin is unexpected, given their complementary function in the mTOR signaling pathway in humans, but could relate to GATOR1’s interaction with other partners that may not be conserved throughout eukaryotes [41].

The longin GAP complexes were previously proposed to be related to one another and to form a small protein family [31, 32, 34] and our results regarding their secondary structures and conserved sequence motifs confirm this classification. Longin domains have historically been difficult to phylogenetically resolve [12] and our results are no exception. We suspect that this is partially due to the long time since the divergence of the protein families and the diverse roles that longin-domain containing proteins play in the cell, but also that their evolution could be underpinned by multiple cases of domain rearrangement or fusion and evolutionary convergence, which are difficult to trace. These factors make it challenging to reconstruct all of the details of their exact phylogenetic relationships. This work has also raised potential new research avenues to understanding the mechanism of the longin GAP complexes. We report a new, highly conserved GP motif present in the catalytic site of SMCR8, C9orf72, FLCN, and NPRL2, that is, specific to the longin domains with GAP activity. We predict that the GP motif might be essential for GAP activity given its conserved position close to the catalytic arginine. Moreover, we identified this motif in a class of structurally similar longins previously reported in Heimdallarchaeia, the proposed closest relatives of eukaryotes. Further studies of this and the other signature motifs we identified should provide new insight into the functions of the longin GAP complexes, not only as cellular components playing their role in their respective organisms but also as evolutionary entities whose relationships and histories may have broader implications for eukaryotic evolution.

## 4 | Materials and Methods

### 4.1 | Homolog Mining

As the base of all pan-eukaryotic homology searches reported in this study, we used the TCS database, a curated subset of Eukprot [62], comprising 196 genomes and transcriptomes, supplemented by several recently published genomes for some of the undersampled lineages: *Mantamonas spyrae* and *Mantamonas vickermanii* (CRuMs) [63], *Proteromonas lacerate* (Opalineae) [64], *Diplonema papillatum* (Diplonemida) [65], *N. marisrubri*, *Nibbleromonas quarantinus*, *Nibbleromonas curacaus*, and *Ubyssesya fretuma* (Nebulida) [66], and *M. sporadica* (deep lineage related to Hemimastigophorea) [44]; the resulting database consists of 207 genomes and transcriptomes and will be hereafter referred to as TCS+.

Homologs of C9orf72, SMCR8, WDR41, PQLC2, FLCN, FNIP, NPRL2, and NPRL3 were searched in TCS+ using HMMer.

The HMM profiles for all genes were prepared by gradual addition of more divergent homologs to the initial seed datasets of human references and other reviewed sequences from Uniprot. This included BLAST searches against non-mammal metazoan datasets in Genbank, and subsequent HMMer searches against datasets for Holozoa and Amorphea. In the case of C9orf72 and SMCR8, this initial search was also performed based on partial profiles for three DENN subdomains (the first of which, termed uDENN, corresponds to the longin domain), as annotated in the reference human sequences, to increase sensitivity. All TCS+ HMMer hits with  $e$ -value  $< 0.01$  for C9orf72, SMCR8, WDR41, and PQLC2 were considered preliminary candidates while only the best hits were considered for FLCN, FNIP, NPRL2, and NPRL3. Reverse BLAST searches against the human reference genome were then performed for all preliminary datasets, and only sequences that returned their respective reference gene among the top three hits were included in the final candidate datasets. The  $e$ -value threshold of these searches was set to  $1e-20$  for WDR41 and  $1e-5$  for the rest as the latter value proved too loose for a satisfactory distinction between WDR41 and other WD repeat proteins. All candidates with detailed annotations, including the ones that did not pass the reverse BLAST check or those later discarded, as well as the final high-confidence candidate datasets are available in Tables S2–S9 and Data S1.

### 4.2 | Phylogenetic Analyses

To further validate these datasets and identify and remove redundant sequences, suspect contaminations, and low-quality sequences, preliminary phylogenetic trees (not shown) were constructed and manually inspected. The datasets for individual proteins of interest were investigated one by one but also in several combinations in order to verify the distinction between related sequences (e.g., C9orf72 vs. FLCN). Alignments were constructed by MAFFT (v7.487 July 25, 2021; [67]) and trimmed by TrimAl (v1.4.rev15 December 17, 2013; [68]), and maximum likelihood trees were built by IQ-TREE2 (v2.1.3 August 24, 2021; [69]) with automatic model selection and 1000 ultrafast bootstraps. Sequences that were removed during this analysis are annotated as such in Tables S2–S9.

Unrooted single-gene trees were subsequently built for each of the proteins of interest to assess the inner relationships of the identified homologs. An additional rooted tree was constructed for the PQLC2 dataset with PQLC1 as an outgroup to confirm the correctness of their distinction. The sequences were aligned by MAFFT and trimmed by BMGE (v2.0; [70]), and trees were built by IQ-TREE2 with LG + C20 + G4 matrix and 1000 ultrafast bootstraps. A concatenated phylogenetic matrix of the three CSW complex subunits was also prepared from a subset of homologs selected based on the presence of at least two subunits in the given organism, absence of divergent paralogues, and sequence completeness (annotated in Tables S2–S4) and analyzed using the same method with C60 matrix. Aligned datasets, trimmed matrices, log files, and trees in nexus format can be found in Data S1.

The sequence identity and similarity of the CSW subunits were investigated, and sequence logos were prepared for manually

selected conserved regions using WebLogo [71]. The selected sequences are annotated in Table S13.

### 4.3 | Structure and Sequence Comparison Between Various Longin Proteins

We employed several approaches to investigate the relationship between C9orf72, SMCR8, FLCN, FNIP, NPRL2, and NPRL3 in terms of their sequence and structure and to compare them to other longin domain-containing proteins. For these analyses, we used a test dataset comprising six well-supported homologs from diverse eukaryotes per protein for the six proteins of interest plus the following longin protein sampling: AP1-3 sigma subunits, AP2-4 mu subunits, BET5/TRAPPC1, MON1, SEC22, DENND1, DENND10, and AVL9. Regions corresponding to the longin domains were extracted from the sequences based on cursory secondary structure prediction by SWISS-MODEL [72], creating three parallel datasets of full-length proteins, longin regions, and nonlongin regions (if present), respectively (Table S10, Figure S3, Data S1). All versus all BLAST searches were performed on each of the datasets to identify even partial sequence homology between different longin proteins (the hits retrieving the same protein were disregarded, as well as hits with *e*-value above 1). Additionally, sequence identity and similarity were calculated from a pairwise MAFFT alignment for every pair of full-length proteins using a custom script (Table S11, Data S1).

A subset of the longin region sequences (three representatives per protein) was modeled by AlphaFold [73] on the ColabFold platform (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>, [74]). A guide tree and structure-based sequence alignment of these longin domains was then produced using FoldMason (<https://search.foldseek.com/foldmason>, [75]), and a maximum likelihood tree was additionally constructed based on the structure-based alignment by IQ-TREE2 with LG + C20 + G4 matrix and 1000 ultrafast bootstraps (Figures S4 and S5, Data S1).

### 4.4 | Secondary Structure of the CSW Complex

To compare the structures of the full-length CSW proteins, secondary structures of C9orf72, SMCR8, and WDR41 were predicted by AlphaFold for a subset of taxonomically representative organisms in which all three subunits were identified (*C. owczarzaki*, *P. atlantis*, *A. castellanii*, *D. rotans*, *M. sporadica* LBC3, and Colponemidia strain Colp10). The models were aligned to a reference cryo-EM-based model of the human CSW complex [76], and their noninformative and disordered regions were manually removed using PyMol (<http://www.pymol.org/pymol>). Structural homology of these models was investigated, and the structure-guided alignments and guide trees were generated using FoldMason [75]; all pairwise structure alignments were also generated and scored using TM-align (<https://zhanggrouop.org/TM-align/>) [77]. The FoldMason alignments and trees are available in Data S1, PAE and pLDTT plots for the models are shown in Figures S8–S10, and TM-score matrices for the structure alignments are given in Table S12.

### Acknowledgements

This study was supported by the Centre National de la Recherche Scientifique, France, and grant ANR-20-CE13-0007 from the Agence Nationale de la Recherche, France, to C.L.J. Research in the Dacks Lab is supported by the Natural Sciences and Engineering Research Council of Canada (RES0043758 and RES0046091).

### Ethics Statement

The authors have nothing to report.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

No new materials nor code were generated for this manuscript. All data is included in [Supporting Information](#).

### References

1. J. G. Donaldson and C. L. Jackson, "ARF Family G Proteins and Their Regulators: Roles in Membrane Transport, Development and Disease," *Nature Reviews. Molecular Cell Biology* 12 (2011): 362–375, <https://doi.org/10.1038/nrm3117>.
2. E. Sztul, P.-W. Chen, J. E. Casanova, et al., "ARF GTPases and Their GEFs and GAPs: Concepts and Challenges," *Molecular Biology of the Cell* 30 (2019): 1249–1271, <https://doi.org/10.1091/mbc.E18-12-0820>.
3. M. Ishida, E. Oguchi, and M. Fukuda, "Multiple Types of Guanine Nucleotide Exchange Factors (GEFs) for Rab Small GTPases," *Cell Structure and Function* 41 (2016): 61–79, <https://doi.org/10.1247/csf.16008>.
4. T. P. Levine, R. D. Daniels, L. H. Wong, A. T. Gatta, A. Gerondopoulos, and F. A. Barr, "Discovery of New Longin and Roadblock Domains That Form Platforms for Small GTPases in Ragulator and TRAPP-II," *Small GTPases* 4 (2013): 62–69, <https://doi.org/10.4161/sgtp.24262>.
5. A. L. Marat, H. Dokainish, and P. S. McPherson, "DENN Domain Proteins: Regulators of Rab GTPases," *Journal of Biological Chemistry* 286 (2011): 13791–13800, <https://doi.org/10.1074/jbc.R110.217067>.
6. S. Yoshimura, A. Gerondopoulos, A. Linford, D. J. Rigden, and F. A. Barr, "Family-Wide Characterization of the DENN Domain Rab GDP-GTP Exchange Factors," *Journal of Cell Biology* 191 (2010): 367–381, <https://doi.org/10.1083/jcb.201008051>.
7. H. Inoue and P. A. Randazzo, "Arf GAPs and Their Interacting Proteins," *Traffic* 8 (2007): 1465–1475, <https://doi.org/10.1111/j.1600-0854.2007.00624.x>.
8. A. Schlacht, E. K. Herman, M. J. Klute, M. C. Field, and J. B. Dacks, "Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System," *Cold Spring Harbor Perspectives in Biology* 6 (2014): a016048, <https://doi.org/10.1101/cshperspect.a016048>.
9. M. P. East, J. B. Bowzard, J. B. Dacks, and R. A. Kahn, "ELMO Domains, Evolutionary and Functional Characterization of a Novel GTPase-Activating Protein (GAP) Domain for Arf Protein Family GTPases," *Journal of Biological Chemistry* 287 (2012): 39538–39553, <https://doi.org/10.1074/jbc.M112.417477>.
10. J. L. Bos, H. Rehmann, and A. Wittinghofer, "GEFs and GAPs: Critical Elements in the Control of Small G Proteins," *Cell* 129 (2007): 865–877, <https://doi.org/10.1016/j.cell.2007.05.018>.
11. M. A. M. Frasa, K. T. Koessmeier, M. R. Ahmadian, and V. M. M. Braga, "Illuminating the Functional and Structural Repertoire of Human TBC/RABGAPs," *Nature Reviews. Molecular Cell Biology* 13 (2012): 67–73, <https://doi.org/10.1038/nrm3267>.

12. N. De Franceschi, K. Wild, A. Schlacht, J. B. Dacks, I. Sinning, and F. Filippini, "Longin and GAF Domains: Structural Evolution and Adaptation to the Subcellular Trafficking Machinery," *Traffic* 15 (2014): 104–121, <https://doi.org/10.1111/tra.12124>.
13. Y. Homma, S. Hiragi, and M. Fukuda, "Rab Family of Small GTPases: An Updated View on Their Regulation and Functions," *FEBS Journal* 288 (2021): 36–55, <https://doi.org/10.1111/febs.15453>.
14. E. Mizuno-Yamasaki, F. Rivera-Molina, and P. Novick, "GTPase Networks in Membrane Traffic," *Annual Review of Biochemistry* 81 (2012): 637–659, <https://doi.org/10.1146/annurev-biochem-052810-093700>.
15. C. L. Jackson, J. Ménétrey, M. Sivia, J. B. Dacks, and M. Eliáš, "An Evolutionary Perspective on Arf Family GTPases," *Current Opinion in Cell Biology* 85 (2023): 102268, <https://doi.org/10.1016/j.ceb.2023.102268>.
16. I. Kjos, K. Vestre, N. A. Guadagno, M. Borg Distefano, and C. Progida, "Rab and Arf Proteins at the Crossroad Between Membrane Transport and Cytoskeleton Dynamics," *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research* 1865 (2018): 1397–1409, <https://doi.org/10.1016/j.bbamcr.2018.07.009>.
17. F.-L. Li and K.-L. Guan, "The Arf Family GTPases: Regulation of Vesicle Biogenesis and Beyond," *BioEssays* 45 (2023): 2200214, <https://doi.org/10.1002/bies.202200214>.
18. L. Eme, D. Tamarit, E. F. Caceres, et al., "Inference and Reconstruction of the Heimdallarchaeal Ancestry of Eukaryotes," *Nature* 618 (2023): 992–999, <https://doi.org/10.1038/s41586-023-06186-2>.
19. K. Zaremba-Niedzwiedzka, E. F. Caceres, J. H. Saw, et al., "Asgard archaea Illuminate the Origin of Eukaryotic Cellular Complexity," *Nature* 541 (2017): 353–358, <https://doi.org/10.1038/nature21031>.
20. B. Avci, K. Panagiotou, M. Albertsen, T. J. G. Ettema, A. Schramm, and K. U. Kjeldsen, "Peculiar Morphology of Asgard Archaeal Cells Close to the Prokaryote-Eukaryote Boundary," *MBio* 16 (2025): e00327-25, <https://doi.org/10.1128/mbio.00327-25>.
21. H. Imachi, M. K. Nobu, N. Nakahara, et al., "Isolation of an Archaeon at the Prokaryote-Eukaryote Interface," *Nature* 577 (2020): 519–525, <https://doi.org/10.1038/s41586-019-1916-6>.
22. H. Imachi, M. K. Nobu, S. Ishii, et al., "Eukaryotes' Closest Relatives Are Internally Simple *Syntrophic archaea*." 2025, <https://doi.org/10.1101/2025.02.26.640444>.
23. R. Vargová, R. Chevreau, M. Alves, et al., "The Asgard Archaeal Origins of Arf Family GTPases Involved in Eukaryotic Organelle Dynamics," *Nature Microbiology* 10 (2025): 495–508, <https://doi.org/10.1038/s41564-024-01904-6>.
24. J. Zhu, R. Xie, Q. Ren, et al., "Asgard Arf GTPases Can Act as Membrane-Associating Molecular Switches With the Potential to Function in Organelle Biogenesis," *Nature Communications* 16 (2025): 2622, <https://doi.org/10.1038/s41467-025-57902-7>.
25. L. D. Barlow and J. B. Dacks, "Seeing the Endomembrane System for the Trees: Evolutionary Analysis Highlights the Importance of Plants as Models for Eukaryotic Membrane-Trafficking," *Seminars in Cell & Developmental Biology* 80 (2018): 142–152, <https://doi.org/10.1016/j.semcdb.2017.09.027>.
26. C. M. Klinger, A. Spang, J. B. Dacks, and T. J. G. Ettema, "Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks," *Molecular Biology and Evolution* 33 (2016): 1528–1541, <https://doi.org/10.1093/molbev/msw034>.
27. K. More, C. M. Klinger, L. D. Barlow, and J. B. Dacks, "Evolution and Natural History of Membrane Trafficking in Eukaryotes," *Current Biology* 30 (2020): R553–R564, <https://doi.org/10.1016/j.cub.2020.03.068>.
28. M. DeJesus-Hernandez, I. R. Mackenzie, B. F. Boeve, et al., "Expanded GGGGCC Hexanucleotide Repeat in Non-Coding Region of C9ORF72 Causes Chromosome 9p-Linked Frontotemporal Dementia and Amyotrophic Lateral Sclerosis," *Neuron* 72 (2011): 245–256, <https://doi.org/10.1016/j.neuron.2011.09.011>.
29. T. P. Levine, R. D. Daniels, A. T. Gatta, L. H. Wong, and M. J. Hayes, "The Product of C9orf72, a Gene Strongly Implicated in Neurodegeneration, Is Structurally Related to DENN Rab-GEFs," *Bioinformatics* 29 (2013): 499–503, <https://doi.org/10.1093/bioinformatics/bts725>.
30. A. E. Renton, E. Majounie, A. Waite, et al., "A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD," *Neuron* 72 (2011): 257–268, <https://doi.org/10.1016/j.neuron.2011.09.010>.
31. R. M. Jansen and J. H. Hurley, "Longin Domain GAP Complexes in Nutrient Signalling, Membrane Traffic and Neurodegeneration," *FEBS Letters* 597 (2023): 750–761, <https://doi.org/10.1002/1873-3468.14538>.
32. M.-Y. Su, S. A. Fromm, J. Remis, D. B. Toso, and J. H. Hurley, "Structural Basis for the ARF GAP Activity and Specificity of the C9orf72 Complex," *Nature Communications* 12 (2021): 3786, <https://doi.org/10.1038/s41467-021-24081-0>.
33. D. Tang, J. Sheng, L. Xu, C. Yan, and S. Qi, "The C9orf72-SMCR8-WDR41 Complex Is a GAP for Small GTPases," *Autophagy* 16 (2020): 1542–1543, <https://doi.org/10.1080/15548627.2020.1779473>.
34. D. Zhang, L. M. Iyer, F. He, and L. Aravind, "Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease," *Frontiers in Genetics* 3 (2012): 283, <https://doi.org/10.3389/fgene.2012.00283>.
35. Z. Cui, A. M. N. Joiner, R. M. Jansen, and J. H. Hurley, "Amino Acid Sensing and Lysosomal Signaling Complexes," *Current Opinion in Structural Biology* 79 (2023): 102544, <https://doi.org/10.1016/j.sbi.2023.102544>.
36. R. E. Lawrence, S. A. Fromm, Y. Fu, et al., "Structural Mechanism of a Rag GTPase Activation Checkpoint by the Lysosomal Folliculin Complex," *Science* 366 (2019): 971–977, <https://doi.org/10.1126/science.aax0364>.
37. G. Y. Liu and D. M. Sabatini, "mTOR at the Nexus of Nutrition, Growth, Ageing and Disease," *Nature Reviews. Molecular Cell Biology* 21 (2020): 183–203, <https://doi.org/10.1038/s41580-019-0199-y>.
38. J. Meng and S. M. Ferguson, "GATOR1-Dependent Recruitment of FLCN-FNIP to Lysosomes Coordinates Rag GTPase Heterodimer Nucleotide Status in Response to Amino Acids," *Journal of Cell Biology* 217 (2018): 2765–2776, <https://doi.org/10.1083/jcb.201712177>.
39. J. Amick, A. K. Tharkeshwar, G. Talaia, and S. M. Ferguson, "PQLC2 Recruits the C9orf72 Complex to Lysosomes in Response to Cationic Amino Acid Starvation," *Journal of Cell Biology* 219 (2019): e201906076, <https://doi.org/10.1083/jcb.201906076>.
40. G. Talaia, J. Amick, and S. M. Ferguson, "Receptor-Like Role for PQLC2 Amino Acid Transporter in the Lysosomal Sensing of Cationic Amino Acids," *Proceedings of the National Academy of Sciences* 118 (2021): e2014941118, <https://doi.org/10.1073/pnas.2014941118>.
41. S. Dokudovskaya, F. Waharte, A. Schlessinger, et al., "A Conserved Coatmer-Related Complex Containing Sec13 and Seh1 Dynamically Associates With the Vacuole in *Saccharomyces cerevisiae*," *Molecular & Cellular Proteomics* 10 (2011): M110.006478, <https://doi.org/10.1074/mcp.M110.006478>.
42. The UniProt Consortium, "UniProt: The Universal Protein Knowledgebase in 2025," *Nucleic Acids Research* 53 (2025): D609–D617, <https://doi.org/10.1093/nar/gkae1010>.
43. F. Burki, A. J. Roger, M. W. Brown, and A. G. B. Simpson, "The New Tree of Eukaryotes," *Trends in Ecology & Evolution* 35 (2020): 43–55, <https://doi.org/10.1016/j.tree.2019.08.008>.
44. Y. Eglit, T. Shiratori, J. Jerlström-Hultqvist, et al., "Metora sporadica, a Protist With Incredible Cell Architecture, Is Related to Hemimastigophora," *Current Biology* 34 (2024): 451–459.e6, <https://doi.org/10.1016/j.cub.2023.12.032>.

45. M. E. Schön, V. V. Zlotogursky, R. P. Singh, et al., "Single Cell Genomics Reveals Plastid-Lacking Picozoa Are Close Relatives of Red Algae," *Nature Communications* 12 (2021): 6651, <https://doi.org/10.1038/s41467-021-26918-0>.
46. K. Williamson, L. Eme, H. Baños, et al., "A Robustly Rooted Tree of Eukaryotes Reveals Their Excavate Ancestry," *Nature* 640 (2025): 1–8, <https://doi.org/10.1038/s41586-025-08709-5>.
47. G. Torruella, L. J. Galindo, D. Moreira, and P. López-García, "Phylogenomics of Neglected Flagellated Protists Supports a Revised Eukaryotic Tree of Life," *Current Biology* 35 (2025): 198–207.e4, <https://doi.org/10.1016/j.cub.2024.10.075>.
48. I. J. Anderson and B. J. Loftus, "*Entamoeba histolytica*: Observations on Metabolism Based on the Genome Sequence," *Experimental Parasitology* 110 (2005): 173–177, <https://doi.org/10.1016/j.exppara.2005.03.010>.
49. L. Novák, Z. Zubáčová, A. Karnkowska, et al., "Arginine Deiminase Pathway Enzymes: Evolutionary History in Metamonads and Other Eukaryotes," *BMC Evolutionary Biology* 16 (2016): 197, <https://doi.org/10.1186/s12862-016-0771-4>.
50. M. Kawano-Kawada, K. Manabe, H. Ichimura, et al., "A PQ-Loop Protein Ypq2 Is Involved in the Exchange of Arginine and Histidine Across the Vacuolar Membrane of *Saccharomyces cerevisiae*," *Scientific Reports* 9 (2019): 15018, <https://doi.org/10.1038/s41598-019-51531-z>.
51. A. Spang, J. H. Saw, S. L. Jørgensen, et al., "Complex Archaea That Bridge the Gap Between Prokaryotes and Eukaryotes," *Nature* 521 (2015): 173–179, <https://doi.org/10.1038/nature14447>.
52. N. Dombrowski, A. P. Teske, and B. J. Baker, "Expansive Microbial Metabolic Versatility and Biodiversity in Dynamic Guaymas Basin Hydrothermal Sediments," *Nature Communications* 9 (2018): 4999, <https://doi.org/10.1038/s41467-018-07418-0>.
53. X. Dong, C. Greening, J. E. Rattray, et al., "Metabolic Potential of Uncultured Bacteria and Archaea Associated With Petroleum Seepage in Deep-Sea Sediments," *Nature Communications* 10 (2019): 1816, <https://doi.org/10.1038/s41467-019-09747-0>.
54. R. Liang, Z. Li, M. C. Y. Lau Vetter, et al., "Genomic Reconstruction of Fossil and Living Microorganisms in Ancient Siberian Permafrost," *Microbiome* 9 (2021): 110, <https://doi.org/10.1186/s40168-021-01057-2>.
55. H. L. Wong, F. I. MacLeod, R. A. White, P. T. Visscher, and B. P. Burns, "Microbial Dark Matter Filling the Niche in Hypersaline Microbial Mats," *Microbiome* 8 (2020): 135, <https://doi.org/10.1186/s40168-020-00910-0>.
56. R. Zhao and J. F. Biddle, "Helarchaeota and Co-Occurring Sulfate-Reducing Bacteria in Subseafloor Sediments From the Costa Rica Margin," *ISME Communications* 1 (2021): 25, <https://doi.org/10.1038/s43705-021-00027-x>.
57. J. Zhang, X. Feng, M. Li, et al., "Deep Origin of Eukaryotes Outside Heimdallarchaeia Within Asgardarchaeota," *Nature* 642 (2025): 1–9, <https://doi.org/10.1038/s41586-025-08955-7>.
58. L. T. Tran, C. Akıl, Y. Senju, and R. C. Robinson, "The Eukaryotic-Like Characteristics of Small GTPase, Roadblock and TRAPPC3 Proteins From Asgard Archaea," *Communications Biology* 7 (2024): 1–13, <https://doi.org/10.1038/s42003-024-05888-1>.
59. T. Hatano, S. Palani, D. Papatziomou, et al., "Asgard Archaea Shed Light on the Evolutionary Origins of the Eukaryotic Ubiquitin-ESCRT Machinery," *Nature Communications* 13 (2022): 3398, <https://doi.org/10.1038/s41467-022-30656-2>.
60. J. Serfontein, R. E. R. Nisbet, C. J. Howe, and P. J. de Vries, "Evolution of the TSC1/TSC2-TOR Signaling Pathway," *Science Signaling* 3 (2010): ra49, <https://doi.org/10.1126/scisignal.2000803>.
61. T. J. P. van Dam, F. J. T. Zwartkruis, J. L. Bos, and B. Snel, "Evolution of the TOR Pathway," *Journal of Molecular Evolution* 73 (2011): 209–220, <https://doi.org/10.1007/s00239-011-9469-9>.
62. D. J. Richter, C. Berney, J. F. H. Strasser, et al., "EukProt: A Database of Genome-Scale Predicted Proteins Across the Diversity of Eukaryotes," *Peer Community Journal* 2 (2022): e56, <https://doi.org/10.24072/pcjournal.173>.
63. J. Blaz, L. J. Galindo, A. A. Heiss, et al., "One High Quality Genome and Two Transcriptome Datasets for New Species of Mantamonas, a Deep-Branching Eukaryote Clade," *Scientific Data* 10 (2023): 603, <https://doi.org/10.1038/s41597-023-02488-2>.
64. K. Záhonová, R. S. Low, C. J. Warren, et al., "Evolutionary Analysis of Cellular Reduction and Anaerobicity in the Hyper-Prevalent Gut Microbe Blastocystis," *Current Biology* 33 (2023): 2449–2464.e8, <https://doi.org/10.1016/j.cub.2023.05.025>.
65. M. Valach, S. Moreira, C. Petitjean, et al., "Recent Expansion of Metabolic Versatility in *Diplonema papillatum*, the Model Species of a Highly Speciose Group of Marine Eukaryotes," *BMC Biology* 21 (2023): 99, <https://doi.org/10.1186/s12915-023-01563-9>.
66. D. V. Tikhonenkov, K. V. Mikhailov, R. M. R. Gawryluk, et al., "Microbial Predators Form a New Supergroup of Eukaryotes," *Nature* 612 (2022): 714–719, <https://doi.org/10.1038/s41586-022-05511-5>.
67. K. Katoh and D. M. Standley, "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability," *Molecular Biology and Evolution* 30 (2013): 772–780, <https://doi.org/10.1093/molbev/mst010>.
68. S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón, "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses," *Bioinformatics* 25 (2009): 1972–1973, <https://doi.org/10.1093/bioinformatics/btp348>.
69. B. Q. Minh, H. A. Schmidt, O. Chernomor, et al., "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era," *Molecular Biology and Evolution* 37 (2020): 1530–1534, <https://doi.org/10.1093/molbev/msaa015>.
70. A. Criscuolo and S. Gribaldo, "BMGE (Block Mapping and Gathering With Entropy): A New Software for Selection of Phylogenetic Informative Regions From Multiple Sequence Alignments," *BMC Evolutionary Biology* 10 (2010): 210, <https://doi.org/10.1186/1471-2148-10-210>.
71. G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, "WebLogo: A Sequence Logo Generator," *Genome Research* 14 (2004): 1188–1190, <https://doi.org/10.1101/gr.849004>.
72. A. Waterhouse, M. Bertoni, S. Bienert, et al., "SWISS-MODEL: Homology Modelling of Protein Structures and Complexes," *Nucleic Acids Research* 46 (2018): W296–W303, <https://doi.org/10.1093/nar/gky427>.
73. J. Jumper, R. Evans, A. Pritzel, et al., "Highly Accurate Protein Structure Prediction With AlphaFold," *Nature* 596 (2021): 583–589, <https://doi.org/10.1038/s41586-021-03819-2>.
74. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold: Making Protein Folding Accessible to All," *Nature Methods* 19 (2022): 679–682, <https://doi.org/10.1038/s41592-022-01488-1>.
75. C. L. M. Gilchrist, M. Mirdita, and M. Steinegger, *Multiple Protein Structure Alignment at Scale With FoldMason* (bioRxiv, 2024), <https://doi.org/10.1101/2024.08.01.606130>.
76. M.-Y. Su, S. A. Fromm, R. Zoncu, and J. H. Hurley, "Structure of the C9orf72 Arf GAP Complex Haploinsufficient in ALS and FTD," *Nature* 585 (2020): 251–255, <https://doi.org/10.1038/s41586-020-2633-x>.
77. Y. Zhang, "TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score," *Nucleic Acids Research* 33 (2005): 2302–2309, <https://doi.org/10.1093/nar/gki524>.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.