Novel transglutaminase-like peptidase and C2 domains elucidate the structure, biogenesis and evolution of the ciliary compartment

Dapeng Zhang and L. Aravind*

National Center for Biotechnology Information; National Library of Medicine; National Institutes of Health; Bethesda, MD USA

Keywords: ciliogenesis, transglutaminase-like, membrane, tubulin-tyrosine ligase, C2, transition zone, Y-shaped linkers, evolution, origin of eukaryotes, ciliopathy

In addition to their role in motility, eukaryotic cilia serve as a distinct compartment for signal transduction and regulatory sequestration of biomolecules. Recent genetic and biochemical studies have revealed an extraordinary diversity of protein complexes involved in the biogenesis of cilia during each cell cycle. Mutations in components of these complexes are at the heart of human ciliopathies such as Nephronophthisis (NPHP), Meckel-Gruber syndrome (MKS), Bardet-Biedl syndrome (BBS) and Joubert syndrome (JBTS). Despite intense studies, proteins in some of these complexes, such as the NPHP1-4-8 and the MKS, remain poorly understood. Using a combination of computational analyses we studied these complexes to identify novel domains in them which might throw new light on their functions and evolutionary origins. First, we identified both catalytically active and inactive versions of transglutaminase-like (TGL) peptidase domains in key ciliary/centrosomal proteins CC2D2A/MKS6, CC2D2B, CEP76 and CCDC135. These ciliary TGL domains appear to have originated from prokaryotic TGL domains that act as peptidases, either in a prokaryotic protein degradation system with the MoxR AAA+ ATP"ｐase, the precursor of eukaryotic dyneins and midasins, or in a peptide-ligase system with an ATP-grasp enzyme comparable to tubulin-modifying TTL proteins. We suggest that active ciliary TGL proteins are part of a cilia-specific peptidase system that might remove tubulin modifications or cleave cilia-localized proteins, while the inactive versions are likely to bind peptides and mediate key interactions during ciliogenesis. Second, we observe a vast radiation of C2 domains, which are key membrane-localization modules, in multiple ciliary proteins, including those from the NPHP1-4-8 and the MKS complexes, such as CC2D2A/MKS6, RPGRIP1, RPGRIP1L, NPHP1, NPHP4, CC2D3, AH11/Jouberin and CEP76, most of which can be traced back to the last eukaryotic ancestor. Identification of these TGL and C2 domains aid in the proper reconstruction of the Y-shaped linkers, which are key structures in the transitional zone of cilia, by allowing precise prediction of the multiple membrane-contacting and protein-protein interaction sites in these structures. These findings help decipher key events in the evolutionary separation of the ciliary and nuclear compartments in course of the emergence of the eukaryotic cell.

Introduction

Eukaryotic cilia (or flagella) are fundamentally different in structural and mechanistic terms from the superficially similar motility organelles of the prokaryotic superkingdoms. Unlike prokaryotic flagella, which are extracellular organelles, these organelles are contiguous with the cytoplasm cell body and are supported by a distinct microtubular skeleton, the axoneme.1 In keeping with this, studies over the past 25 years have revealed that cilia are not just organelles of motility, but a distinct subcellular compartment, which has important additional roles as a locus for sensory signal transduction and for regulatory sequestering of proteins away from the cytoplasm of the cell body.2-3 Recent studies have indicated that several hundreds of proteins reside transiently or permanently in cilia.4-5 Beyond the microtubular axonemal core (typically adopting the “9 + 2” microtubule configuration) and motor proteins, the resident proteins form several distinct complexes, which are central to ciliary function and assembly.6-8 These include (1) the dynein-regulatory (nexin) complex, which connects the microtubule doublets of the cilium, and also links them to different dyneins;7 (2) the septin 2/7 complex, which forms a transport barrier at the base of the cilium;8,9 (3) the tubulin polyglutamylase complex,10 which covalently modifies the microtubules and regulates their interactions with the dyneins; (4) the Bardet-Biedl protein complex or BBsome;11 (5) the intraflagellar-transport complexes or IFT complexes;12,13 (6) the MKS complex14 and (7) the NPHP 1-4-8 complex.15 The last four complexes play distinct roles in the assembly and membrane association of the ciliary cytoskeleton and trafficking of membrane proteins into the ciliary compartment.15-17

While the structure and biogenesis of cilia show some tissue-specific and phyletic differences, certain common features can

*Correspondence to: L. Aravind; Email: aravind@ncbi.nlm.nih.gov
Submitted: 08/21/12; Accepted: 09/03/12
http://dx.doi.org/10.4161/cc.22068
be discerned across eukaryotes. In general, ciliogenesis follows a unique series of steps, which distinguish it from the dynamics of other subcellular compartments. In each cell cycle, the first step in ciliogenesis is the maturation of the basal body from the centriole or the cognate microtubule-organizing center. In the simplest cases, the basal body directly migrates to the proximity of the cell membrane to initiate ciliogenesis. In other cases, the basal body is first capped by a double-membrane sheath, the ciliary vesicle, which brings the basal body close to the cell membrane by fusing with it. The fusion might result in a local invagination, the ciliary pocket, with which the incipient cilium (equivalent to the transition zone of the mature cilium) associates and serves as the center for further trafficking of proteins in and out of the ciliary compartment. The specific targeting of axonemal and ciliary membrane proteins to this region then allows further growth of the cilium to a predetermined length. Genetic and cytological studies suggest that the early steps of ciliogenesis, including the association of the ciliary body with the vesicle, or its docking close to the cell membrane, are dependent on components of the MKS and NPHP 1-4-8 complexes. The subsequent steps involving preliminary growth and extension of the cilium beyond the transition zone are dependent on the BBSome and IFT complexes.

The presence of cilia (flagella) at the base of all major eukaryotic lineages, and also their presence in most “excavate” lineages, which are likely to be the earliest-branching eukaryotic clades, indicates that they are a shared derived character (synapomorphy) of the entire eukaryotic clade. Thus, the early evolution of cilia as a novel multifunctional organelle distinct from the motility systems observed in the prokaryotic superkingdoms is central to the question of eukaryotic origins. Answering this question is complicated by the emergence of a practically “fully formed” cilium in the last eukaryotic common ancestor (LECA), with apparently no intermediates or precursors. Some aspects of the early evolution of cilia are also related to the more general question of the emergence of subcellular compartments during the origin of eukaryotes. Maturation of the ciliary compartment and transport of proteins into it are dependent on a number of small GTPases of the extended Ras-like clade, such as Arl6, which functions with the BBSome, Arl13B, Arf4, Ran and Rab8. As suggested by previous studies, the explosive radiation of these small GTPases in eukaryotes from precursors acquired from prokaryotes appears to have been part of not just the emergence of the ciliary compartment, but also other subcellular membrane-bound structures, such as the nucleus and the Golgi-vesicular complex. Indeed, origin of the nuclear and the ciliary compartments might be closely linked, as suggested by their shared trafficking of GTPase, Ran and the nucleoporins, which also restrict protein transport into the cilium. Computational studies on the sequence relationships of component proteins have provided key leads regarding the origin of the microtubular cytoskeleton and the dynein and kinesin motors. However, it should be kept in mind that the question of the origin of cilia is related to, but not the same as, explaining the provenance of tubulin or the motor proteins. Central to explaining the origin of cilia are scenarios that can account for the microtubular skeleton and the motor proteins coming together as an assemblage with the membrane-associated protein complexes. In particular, elucidating the origin of the key players in early ciliogenesis (e.g., MKS and NPHP 1-4-8 complexes) and their links to the core ciliary components are likely to be of considerable value in explaining the early evolution of cilia. Our analysis of the B9 proteins, key components of the MKS complex, with identification of a distinct version of the C2 domain, have clarified certain aspects of the early evolution of cilia. This and other protein sequence and structure analysis studies have clarified the evolutionary origins of certain components of the cilium. They have also provided new leads to better understand both the cell biology of ciliogenesis and pathologies that are central to a broad class of human diseases known as ciliopathies. Given the efficacy of these computational methods in dissecting the structures and functions of these proteins, we resorted to an in depth sequence analysis of key players in ciliogenesis to detect novel domains and predicted functional linkages to other ciliary components. As a consequence, we were able to reconstruct certain key events in the evolution of these complexes, predict new functions and components and provide a possible explanation for how the cytoskeleton-membrane interactions arose during the origin of eukaryotic cilia.

Results and Discussion

Sequence-structure analysis of components of the MKS and NPHP complexes. Given the central role of the MKS and NPHP 1-4-8 complexes in early ciliogenesis, we attempted to establish the affinities and provenance of key components of these complexes. Interestingly, analysis of sequences of proteins belonging to these complexes using the SEG program, with parameters adjusted to detect globular domains, revealed that several of them contained globular regions that did not map to previously characterized domains. Hence, the relationships and structures of these distinct globular regions are vital to develop a better understanding of both the evolution and functions of these ciliary proteins. Table 1 displays a summary of primary components implicated in ciliogenesis that we analyzed in this study along with the known globular domains and those newly detected by us.

Detection of a conserved transglutaminase-like domain in the ciliary/centriolar proteins CC2D2A/MKS6, CC2D2B, CEP76 and CCDC135. The CC2D2A/MKS6 is a large protein in the MKS complex in which, previously, only a C2 domain (gi: 197209974, residues 1,040–1,200) in human CC2D2A; see below for details) could be detected, along with a N-terminal coiled-coil region (Fig. 1A). Further, we noticed that its C-terminal region (residues 1,320–1,430) is found in a paralogous protein CC2D2B (gi: 229577352, residues 100–220; human CC2D2B), which, however, lacks the C2 domain. A search of the non-redundant database with the PSI-BLAST program using this region as seed recovered homologous regions in CEP76, which is a centriole-associated protein, and CCDC135/FAP50/Lost Boys, which is a conserved ciliary protein tightly associated with the outer microtubule doublets of the...
This suggested that the common globular domain shared by CC2D2A, CC2D2B, CEP76 and CCDC135 is a version of the TGL domain (Fig. 1A). The TGL domain adopts the papain-like peptidase fold, whose active site is typically comprised of a catalytic triad formed by a cysteine from an N-terminal helix, and a histidine and an acidic residue from successive strands of the core β-barrel.43 Catalytically active versions are peptidases, peptide-N-glycanases, transglutaminases or deamidases, while

| Table 1. Summary of the primary components of ciliary complexes with the domain architectures and function predictions |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Protein complex | Gene name/complex | Gene name/complex | GI ID (human) | Domain architecture and mutation/deletion mapping | Functions |
| NPHP 1–4–8      | NPHP1/MKS6       | NPHP1/MKS6       | 189491774     | SH3+ NPHP1–C2 +α-helical domain | Membrane localization |
|                 | NPHP1/MKS6       | NPHP1/MKS6       | 23510323      | NPHP4C–C2 +NPHP4C–C2+G3 | Membrane localization |
|                 | RPGIRIP1L/MKS5   | NPHP8/MKS5       | 18442834      | CC+ RPGIRIP1L–C2+PKC–C2+RPGIRIP1–C2 | Membrane localization+protein interaction |
|                 | RPGIRIP1         | NPHP1/MKS6       | 112734867     | CC+ RPGIRIP1L–C2+PKC–C2+RPGIRIP1–C2 | Membrane localization+protein interaction |
|                 | MKS1/MKS1/BBS13  | MKS1/MKS1/BBS13  | 89242137      | B9–C2 | Membrane localization |
|                 | B9D1/MKS9        | B9D1/MKS9        | 7661536       | B9–C2 | Membrane localization |
|                 | B9D2/MKS10       | B9D2/MKS10       | 226371646     | B9–C2 | Membrane localization |
|                 | TCTN1/JBTS13     | TCTN1/JBTS13     | 91208022      | DUF1619 | Membrane localization |
|                 | TCTN2/MKS8       | TCTN2/MKS8       | 74731861      | DUF1619 | Membrane localization |
|                 | TCTN3/MKS8       | TCTN3/MKS8       | 91208025      | DUF1619 | Membrane localization |
|                 | CC2D2A/MKS6      | CC2D2A/MKS6      | 197209974     | CC+CC2D2AN–C2 +CC2D2AC–C2 +TGL | Membrane localization+protein interaction |
| Meckelin/TMEM67 | NPHP11/MKS3      | NPHP11/MKS3      | 317373389     | TM | Membrane localization+protein interaction |
|                 | TMEM216/MKS2     | TMEM216/MKS2     | 115387120     | TM | Membrane localization+protein interaction |
|                 | TMEM237/JBTS14   | TMEM237/JBTS14   | 113205077     | TM | Membrane localization+protein interaction |
| Novel members of MKS and NPHP complexes | AHII1/JOUBERIN/JBTS3 | JbTS3 | 31542701 | AHII1–C2+WD40+SH3 | Membrane localization |
|                 | C2CD3            | C2CD3           | 148596944     | C2CD3N–C2+PKC–C2*5 | Membrane localization |
|                 | CEP76            | CEP76           | 21314728      | CEP76–C2+TGL | Membrane localization+peptidase activity |
|                 | CCDC135/FAP50    | CCDC135/FAP50    | 223941912     | TGL+CC | Peptidase activity |
| NPHP 5–6        | CEP290/NPHP6     | CEP290/NPHP6     | 116241294     | CC (Coiled coil) | Membrane localization |
|                 | IQCB1/NPHP5      | IQCB1/NPHP5     | 3123054       | TPRs+IQ_repeat | Protein interaction |

The novel protein domains deleted or disrupted by ciliopathy mutations in humans are underlined.

axoneme (e ≈ 10^−5 in iterations 1–2). Further iterations of this search recovered a region in several prokaryotic proteins (e.g., Bacillus subtilis protein YebA and the Mycobacterium smegmatis protein, gi: 118473669; 10^−7–10^−15 in iterations 3–4). Similar results were obtained in searches with the JACKHMMER program. The region of significant similarity shared by these prokaryotic proteins and the eukaryotic ciliary/centriolar proteins corresponded to the Transglutaminase-like (TGL) domain.
catalytic inactive versions (e.g., Rad4 and Cyk3) function as peptide-binding domains.44

An examination of the multiple sequence alignment of the core of the TGL domains from prokaryotic TGL proteins and the ciliary/centriolar proteins indicated that the former strongly conserve the catalytic triad, implying that there are active enzymes (Fig. 1A). However, the eukaryotic proteins present a more complex picture due to their drastic divergence relative to the TGL domain of their prokaryotic homologs. In order to understand their evolutionary history, we used the prokaryotic homologs as a root in a phylogenetic analysis of the TGL domains of the eukaryotic ciliary/centriolar proteins. This revealed three distinct clades of the TGL domains in eukaryotes, respectively prototyped by CCDC135, CEP76 and CC2D2A/MKS6 (Fig. 1B). The monophyly of these three distinct eukaryotic branches is supported by several sequence features that they share to the exclusion of all other TGL domains (Fig. 1A). They also share a unique C-terminal α+β extension to the core TGL domain that is likely to stack with the former to comprise the complete folded unit. Indeed, such extensions have been previously observed in other subfamilies of TGLs, such as the Rad4-PNGase subgroup.43,44 Of the three eukaryotic clades, CEP76 and CC2D2A/MKS6 form a strongly supported higher-order clade (Fig. 1B), which is characterized by a compact TGL domain. In contrast, CCDC135 has a large insert after the first strand of the core TGL domain (Fig. 1A), which is a common location for large inserts in several members of the TGL superfamily.44 Examination of the phylogenetic patterns indicates that each of the three eukaryotic clades has representatives from all major ciliated eukaryotic lineages, including the “excavate” taxa. Among the excavate taxa, Giardia and Trichomonas tend to form the basal-most branches, followed by kinetoplastids and heteroloboseans. This pattern is consistent with the previously inferred phylogeny of eukaryotic taxa25,45 and suggests that the three distinct clades of TGL domains had radiated prior to the LECA. Examination of the active site residues (Fig. 1A) reveals that all representatives of the CC2D2A/MKS6 clade are inactive. In the CEP76 clade, most representatives are inactive, though many of the animal sequences appear to possess intact active site residues. In the CCDC135 clade, one subset contains the usual active site cysteine, suggesting that they function as thiol peptidases. The other subset (e.g., in Drosophila; Fig. 1A) displays a serine in place of the catalytic cysteine, similar to what is observed in the TGL domain of the eukaryotic Menin proteins.46,47 Hence, these serine-containing TGL domains could be potentially active as serine peptidases, comparable to the previously noted case of the plasmodial SERA peptidase,48 which displays a cathepsin-like version of the papain-like fold with an active site serine. Thus, the eukaryotic ciliary/centriolar TGLs are predicted to include both catalytically active (mainly from the CCDC135 clade) and inactive versions (mostly from the CEP76 and CC2D2A/MKS6 clades).

Functional and evolutionary implications of ciliary TGL domains: Inferences from operonic associations of the bacterial homologs. The prokaryotic homologs recovered in searches with the eukaryotic ciliary/centriolar TGL domains provide functional clues regarding their eukaryotic counterparts. We used contextual information gleaned from domain architectures and conserved gene neighborhoods (predicted operons) to

Table 1. Summary of the primary components of ciliary complexes with the domain architectures and function predictions (continued)
Figure 1A and B. (A) Multiple sequence alignment of the core region of TGL domains from bacterial and eukaryotic ciliary proteins. The catalytic triad residues (C/S, H, E/D) are labeled with number sign (#) and highlighted in red background. Secondary structures are shown with α-helices in pink and β-strands in light blue. Domain architectures of each ciliary TGL domains are shown below the alignment. Proteins are denoted by their gene name, species abbreviations and GI (GenBank Index) numbers separated by underscores. For species abbreviations, refer to the Materials and Methods section. (B) Divergent evolution of ciliary TGL domains prior to LECA. Capital L in the circle indicates the presence of the domain in the LECA. The tree was reconstructed using an approximately maximum-likelihood method implemented in the FastTree 2.1 program under default parameters. Bootstrap values are shown at each node.
better characterize the functional association of the homologous prokaryotic TGL proteins. YebA-like TGL proteins are found in most members of the Gram-positive clades of bacteria, the lentisphaera- verrucomicrobia-planctomycetes clade, more sporadically in all other major bacterial lineages and in euryarchaea. This phyletic pattern is generally suggestive of an early origin in bacteria, followed by widespread dispersion through lateral transfer between different bacterial lineages, and also probably into the euryarchaea, early in their evolution. We found that the YebA-like proteins typically contain eight TM regions with the TGL domain occurring in the predicted cytoplasmic loop between transmembrane (TM) 7 and TM 8 (not shown). This suggests that the peptidase activity of the YebA protein operates in close proximity to the membrane. Further analysis of gene neighborhoods (Fig. 1C) revealed a strict association of the YebA-like TGL genes with two other genes, respectively coding for an AAA+ ATPase of the MoxR family and a von Willebrand factor A (vWA) protein in both archaea and bacteria. In some cases, the genes for the MoxR AAA+ ATPase and the vWA protein are fused into a single gene. Despite this three-gene operon being very common in prokaryotes, the functional significance of the encoded proteins has not been hitherto explained. Clues regarding their potential function emerge from comparisons with other AAA+ ATPases. First, on at least six independent occasions, different versions of the AAA+ domain from across this superfamily have been functionally combined with different types of peptidase domains to constitute ATP-dependent protein degradation systems such as the proteasome, HslUV, ClpXP, FtsH, Lon and YifB. Based on this, we propose that this version of the MoxR AAA+ ATPase functions as an ATP-dependent protein unwinding engine that facilitates protein cleavage by YebA-like TGL domains. Second, the vWA domain was earlier shown to have strong functional linkages with multiple AAA+ domains belonging to the clade that unites dynine-midasin, MoxR, YifB and the chelatases. While based on the precedence of the chelatases, the vWA domain was proposed to be required for metal insertion in substrates; the new evidence from midasin and the archaeal phage tail chaperones suggests that it is likely to function as metal-dependent substrate-binding co-chaperones for this clade of AAA+ domains. Thus, we infer that the conserved MoxR-vWA-YebA operon is likely to encode a membrane-proximal protein degradation system (Fig. 1C), in which the vWA domain protein recruits substrates, while the MoxR AAA+ ATPase unwinds them, and the YebA-like TGL domain cleaves the unwound polypeptide. Another group of homologous prokaryotic TGL domains recovered in the searches with the ciliary TGL domains are those which we had previously described as being combined with genes encoding predicted peptide-ligases of the ATP-grasp fold and also a conserved protein termed the α-E domain (Fig. 1C). Here, too, the TGL domains are predicted to function as peptidases that reverse the peptide tags ligated by the ATP-grasp ligase (comparable to the peptide tags such polyglutamate, polylglycine and tyrosine ligated by the tubulin-modifying TTL family of ATP-grasp ligases).
These observations have notable implications for both the function and evolution of the eukaryotic ciliary TGL proteins. Analysis of protein-protein interaction networks using the FUNCOP program points to an interaction between the MKS and the dynein complexes with a direct edge between the latter and CC2D2A/MKS6. This suggests that at least in part the functional connection between the AAA+ ATPase and the TGL domain protein is paralleled in both prokaryotes and eukaryotes. Importantly, the MoxR AAA+ domain is the closest prokaryotic homolog of the six AAA+ domains in dynein and midasin and is specifically united to them by several sequence features. MoxR appears to have given rise to a protein with six tandem AAA+ ATPase domains early in eukaryotic evolution, which further diverged to dynein and midasin. In the course of this event, there appears to have been several changes in the original "prokaryotic" configuration of interacting partners. Midasin retained the vWA domain and functioned primarily as a chaperone for the removal of ribosomal assembly factors and the assembled ribosome from the nucleolus and nucleus (Fig. 1C). On the other hand, dynein appears to have lost the vWA domain while retaining its interaction with TGL domain proteins (Fig. 1C). The loss of the vWA co-chaperone probably facilitated its transition from a protein unfolding engine to a motor driving movement and trafficking. Indeed, the diversification of midasin (with a predominantly nuclear function) and dynein (with an ancient ciliary function), is likely to correspond to the formation of distinct nuclear and ciliary compartments in the eukaryotic progenitor. The functional shift in dynein relative to the ancestral MoxR ATPase probably allowed the TGL domain proteins to acquire certain independent functions, which is probably illustrated by the inactive versions (e.g., CC2D2A/MKS6) that are predicted to merely function as peptide-binding domains. The versions of the TGL found in the prokaryotic operons with the ATP-grasp peptide ligases also provide clues for the catalytically active eukaryotic ciliary TGL domains (e.g., CCDC135). Interestingly, in certain eukaryotes, such as stramenopiles, this TGL domain is fused to a related polyglutamylase-like ATP-grasp domain (Fig. 1A). This leads to the interesting possibility that the peptide tags on tubulin, such as polyglutamate and the C-terminal tyrosine, are reversed by the peptidase action of the active TGL domains, whereas the inactive versions merely bind them (Fig. 1C).

Identification of novel C2 domains in multiple components of the NPHP1-4-8 and MKS complexes. The functional parallel between the prokaryotic system, which combines MoxR with a membrane-linked TGL domain, and the eukaryotic dynein complex and ciliary TGLs suggested that membrane association of the axonemal and motor proteins might have been an early event in the origin of the cilia. Whereas the prokaryotic TGL protein associates with the membrane using TM helices, the CC2D2A/MKS6 is predicted to associate with membranes via a C2 domain. The presence of the C2 domain in CC2D2A/MKS6 belongs to a distinct clade of C2 domains in the N terminus and the central region (Fig. 2A). Similarly, we found a previously uncharacterized globular domain N terminal to the TGL domain in CEP76 (Fig. 2B). Profile-profile searches with this domain recovered C2 domains with significant scores (e.g., the myoferlin C2 domain, 2dmh; probability 97%; p = 10^-6), indicating the presence of yet another undetected C2 domain. Thus, taken together, we were able to identify four distinct varieties of the C2 domain (B9-C2, CC2D2AN-C2, CC2D2AC-C2, CEP76-C2) in the MKS complex and associated centriolar proteins.

Double mutants with disruptions of one component each, respectively from the MKS and NPHP1-4-8 complexes, display loss of the Y-shaped linkers in the ciliary transitional zone, which link the outer microtubule doublets to the ciliary membrane. This suggests that the membrane-anchoring function of the MKS complex acts in conjunction with the NPHP 1-4-8 complex. Interestingly, even RPGRIP1L/NPHP8/MKS5 shows a classical PKC-C2 domain in its central region, suggesting that direct membrane-association via C2 domains might be a feature that extends to the NPHP 1-4-8 complex. Given that there are several uncharacterized globular domains in components of the NPHP 1-4-8 complex, we analyzed them further with sequence profile, HMM and profile-profile comparison searches. As a result, we recovered two additional distinctive, divergent C2 domains flanking the central PKC-C2 domain in NPHP8 and its paralog RPGRIP1 (RPGRIP1N-C2 and RPGRIP1C-C2; Fig. 2C). By means of further searches, we found that NPHP1 had one novel C2 domain in the central region of the protein (NPHP1-C2; Fig. 3A), whereas NPHP4 had two previously undetected, divergent C2 domains in the N terminus and the central region of the protein (NPHP4N-C2 and NPHP4C-C2; Fig. 3B). Thus,
we were able to establish that all the three core components of the
of the NPHP 1-4-8 complex contain C2 domains belonging to a
total of five distinct varieties in addition to the classical PKC-C2
domain.

Furthermore, the C2CD3 protein, which was recovered in
mouse models as playing a role comparable to components of
MKS and NPHP 1-4-8 complexes in ciliogenesis, also contains
a tandem array of five classical PKC-C2 domains. We found that
C2CD3 contains an uncharacterized globular domain N terminal
to these C2 domains, which is yet another novel, divergent version
of the C2 domain (HHpred probability 90%, p = 10^{-5}) N terminal to the known WD40 β-propeller and SH3 domains
(Fig. 3D). In line with some recent cytological studies, the
presence of C2 domains in C2CD3 and AHI1/Jouberin, a feature
shared with components of the MKS and NPHP 1-4-8 complexes,
suggests that these two proteins are also likely to function as components or in close association with those complexes in cytoskeleton-membrane interactions. Another version of the C2
domain potentially involved in ciliogenesis is the AIDA-C2 that
we previously reported. Although no direct evidence is avail-
able, its domain architecture and phyletic pattern are consistent
with a ciliary or centriolar role.

Early radiation of C2 domains was central to the origin of
the protein complexes related to ciliogenesis. Identification of
parabasalids, diplomonads and the kinetoplastid-heterolobosean clade) (Fig. 4). This suggests that at least 10 C2 domain-containing proteins, with a total of at least 11 distinct C2 domains, were potentially present in the LECA, constituting ancestral versions of the NPHP and MKS complexes. We performed a phylogenetic analysis of the C2 domains, including all the new versions detected in this study. On account of the extreme divergence of the different versions of this domain, the overall topology of the tree should be viewed with circumspection (Fig. 4). Nevertheless, it was clear that each of the newly detected ciliary

![Figure 3](image_url). Multiple sequence alignment and domain architectures of novel C2 domains of ciliary proteins, such as NPHP1 (A), NPHP4 (B), C2CD3 (C) and AH11 (D). For species abbreviations, refer to the Materials and Methods section.

at least 12 distinct versions of C2 domains, in addition to the PKC-C2 in components of the key ciliogenesis complexes, MKS and NPHP 1-4-8, and other ciliary proteins predicted to functionally interact with them, suggests that a major radiation of the C2 domains was central to their emergence. Phyletic patterns of the novel C2 domains (Supplemental Material) revealed that three versions of the B9-C2 and one version each from NPHP1, NPHP4, C2CD3, CC2D2A, CEP76 and AH11/Jouberin, in addition to the PKC-C2 domains, can be found in basal eukaryotic lineages or the taxa with excavate morphology (namely...
versions formed well-supported branches that were distinct from the classical PKC-C2 domains and other ancient versions, such as NT-C2 (Fig. 4), which was previously implicated in anchoring the actin cytoskeleton.37 The monophyly of the distinct C2 clades found among the ciliogenesis components was also supported by the unique sequence signatures of each of the groups (Figs. 2 and 3). Together, these observations indicated that a notable part of the early radiation of the C2 domains, which resulted in the emergence of the key ciliogenesis complexes, probably happened between the time of the first eukaryotic common ancestor (FECA) and the LECA. Interestingly, this reconstruction also showed that the majority of the C2 domains, which can be potentially traced back to the LECA, are specifically associated with ciliogenesis. Hence, it raises the possibility that the original radiation the C2 domains in eukaryotes happened in the context of the membrane-association of the ciliary cytoskeleton.

Functional implications of diverse C2 domain proteins in ciliogenesis. While the C2 domains are a strong predictor of membrane interactions,37 the versions that are potentially traceable to the LECA show considerable sequence diversification and also diversity of domain architectures (Figs. 2 and 3). Hence, there is likely to have been a functional diversification of the C2 domains and the proteins containing them even at the time of the LECA. Of these, the pre-LECA triplication of proteins with B9-C2 domains and strict maintenance of these three distinct representatives throughout eukaryotes was interpreted

Figure 4. Evolutionary relationship of the different C2 domain families and the comparison of their phyletic patterns, functions and features of domain architecture. The tree was reconstructed using an approximately maximum-likelihood method implemented in the FastTree 2.1 program under default parameters. Nodes supported with bootstrap values greater than 75% are shown.
as implying that the three proteins form a trimeric subcomplex within the MKS complex, whose subunit stoichiometry is strongly conserved. This complex might be compared with the other complexes in eukaryotes with multiple paralogous subunits forming symmetrical toroidal or spherical structures with fixed subunit stoichiometry, e.g., the CCT/TRiC-complex, which also forms part of the ciliary BBSome, the Rad9–1-1 complex in DNA repair and the core proteasome. Consistent with this, the B9-C2 proteins, as a rule, have simple domain architectures with no combinations with other domains. Hence, we predict that the B9-C2 is likely to form a torroidal structure at the heart of the MKS complex. In contrast, the C2 domains of the CC2D2A/MKS6 and CEP76 proteins occur in multidomain architectures combined with TGL domains. Hence, these proteins are more likely to function as adaptors, with the C2 domains binding the membrane and their catalytically inactive TGL domains (e.g., those in CC2D2A/MKS6 and CC2D2B) probably binding peptides from ciliary cytoplasm or axonemal proteins (Fig. 2A and B). The presence of coiled coil segments in these proteins is also suggestive of dimerization or interaction with other structural components with coiled coil segments (e.g., components of the IFT complex). A similar scenario is likely for the RPRGIP1/NPHP8, which combine the C2 domains with N-terminal coiled-coil segments (Fig. 2C). In light of their proposed role in forming scaffolds in the ciliary transition zone, we propose that they are likely to be important as membrane anchors for targets bound by the MKS complex for intra-ciliary transport or signaling via the coiled-coil regions in these proteins. AH11/Jouberin could also function similarly with the WD40 domains present C-terminal to the C2 domain playing a role in recruiting specific target proteins to the ciliary membrane (Fig. 3D).

Interestingly, we also found evidence for considerable plasticity in the domain architectures of these proteins beyond the conserved core, which includes the C2 domain. For example, in the case of AH11/Jouberin, the conserved core, which is traceable to the LECA, includes the C2 domain and a WD40 β-propeller. In subsequent eukaryotic evolution, SH3 domain seems to have been added at the C terminus (Fig. 3D). In the case of the NPHP4, the C-terminal region contains immunoglobulin (IG) domains in certain eukaryotes, and EF-hand domains in other eukaryotes (Fig. 3B). On the other hand, in NPHP1 the N-terminal region is variable, with certain eukaryotic clades showing N-terminal SH3 domains, while others show EF-hand domains (Fig. 3A). The presence of the EF-hand domains in multiple proteins in these complexes suggest that, at least in certain eukaryotes, they are likely to respond to the presence of Ca2+. This is consistent with the recruitment of two highly conserved paralogs of the Ca2+-binding EF-hand protein centrin associated with the contractile function of the microtubular skeleton right from the time of the LECA. However, examination of the sequence alignments suggests that, interestingly, none of the C2 domains traceable to the ciliary complexes in the LECA have Ca2+-binding ability. Thus, the Ca2+-binding role of the C2 domain (PKC-C2) is likely to have arisen independently of the ciliary function among C2 domains. The above-reported domain architectural variability of these ciliary proteins is consistent with the lineage-specific diversification of target proteins that are transported to the cilium by the action of these proteins in different eukaryotic lineages. For example, the hedgehog-signaling pathway proteins such as patched, smoothed, SuFu and Gli, which are localized to the cilium, are only found in the animal lineage. Hence, the domain architectures of the primary components of the MKS and NPHP1-4-8 complexes indicate that C2 and TGL domains constitute the ancient core, which might be further extended accretion of different domains in particular eukaryotic lineages.

Based on the above analysis and the available genetic evidence, we predict that the C2 and TGL domain-containing proteins can be visualized as constituting the Y-shaped linkers of the ciliary transitional zone, with the arms of the Y-shaped linkers formed primarily by the coiled coil segments and the heads contacting the membrane being mainly comprised of the C2 domains. The three B9-C2 domains potentially form a central torroidal core at these membrane-contacting sites. On the other hand, the active and inactive TGL domains might constitute the axoneme- and substrate-interaction interfaces of the Y-shaped linkers. Thus, the combination of the C2 and TGL domains and coiled coil segments in these proteins explains their role both as structural components and as gatekeepers of intra-ciliary transport.

**Evolutionary implications for origin of eukaryotic cilia and general conclusions.** Several recent studies have aimed at tracing the provenance of key eukaryotic cellular components to the prokaryotic superkingdoms. For example, both actin and tubulin have been infrequently found in a small number of archaeal and bacterial lineages. Phylogenetic analysis has been used to argue that the archaea might have been the source of tubulin. Based on the presence of tubulin in thaumarchaea and related lineages, it has been argued that the eukaryotic microtubular skeleton was acquired from an archaeal progenitor of the eukaryotes that resembled a thaumarcheon. This possibility is consistent with the presence of other eukaryote-like features in thaumarchaea, such as the ubiquitin (Ub) system and actin-like proteins. However, it should be borne in mind that many of these eukaryote-like features are not found in the same thaumarcheon, are also rather infrequently found in prokaryotes and are encoded by potentially mobile operons (as demonstrated in the case of the Ub-system). Hence, while a thaumarcheon-like organism could have been the archaeal progenitor of eukaryotes, it is possible that this progenitor did not have all the eukaryote-like genes together in its genome right from the inception, but accreted them over time via lateral transfer. Indeed, such gradual accretion of numerous, diverse mobile operons is observed in certain bacteria with large genomes, such as colonial myxobacteria. This accretion then created conditions that brought together diverse systems, allowing their “mixing and matching” to generate novel systems such as the cilium. In particular, the origin of the common ancestor of dyneins and midasins from the mobile MoxR-like systems favors a scenario, where the eukaryotic progenitor acquired additional systems by lateral transfer either from other co-occurring archaea and bacteria or from the bacterial endosymbiont that gave rise to the mitochondrion. Our observation of the presence of a functional linkage between the dynactin complex and the ciliary TGL domains (Fig. 1C), which
is mirrored in the mobile prokaryotic MoxR-like systems with membrane-linked TGL domains, suggests that the acquisition of the MoxR-like precursor of dynein and the functionally linked TGL domains might have been a critical event for the emergence of the eukaryote-type cilia from the microtubular cytoskeleton. This is also comparable to the earlier observation on the mobile Mgl operons of prokaryotes, which encode a small GTPase of the eukaryote-type cilia from the microtubular cytoskeleton. TGL domains might have been a critical event for the emergence of the mobile prokaryotic MoxR-like systems, with the localization system dependent on the karyopherins and the nuclear and ciliary compartments. The major event in emergence of the eukaryotic ciliium.

In bacteria and certain archaea, the proteins of the FtsZ-tubulin superfamily form a polymeric ring that mediates cytokinesis or chromosome segregation, while in thumarchaea, there is no evidence for the FtsZ-tubulin superfamily proteins participating in cell division. This could have “freed” the FtsZ-tubulin proteins in the lineage leading, eukaryotes to adopt alternative functions. Our above analysis suggests that it was the coming together of the above-described components with the tubulin-like proteins that triggered the emergence of the eukaryotic ciliary system. Multiple lines of evidence point to this event being closely linked to the emergence of the nuclear compartment, which is the quintessential feature of eukaryotes. In functional terms, loss of the ancient chromosome segregation mechanisms dependent on nucleic acid pumps in eukaryotes was possibly the main factor that linked the emergence of the nucleus with the origin of cilia; both eukaryotic chromosome segregation and motility came to depend on the same contractile microtubular cytoskeleton. The major event in emergence of both the nuclear and ciliary compartments was the emergence of the localization system dependent on the karyopherins and the RAN GTPase. As noted above, this event is likely to have also been accompanied by the divergence of midasin and dynein from their common ancestor, with only dynein retaining the functional interactions with TGL domain proteins in proximity of the membrane, a key feature for the origin of the ciliun. Nevertheless, the question remains as to what differentiated the nuclear and ciliary compartments if they utilized the same localization and gating mechanisms. Our identification of novel C2 domains in the two key complexes related to the transition zone morphology, ciliogenesis and intra-ciliary transport aids in answering this question. While several nuclear membrane proteins have been identified, including those which can be confidently traced back to the LECA, none of them contain C2 domains. In contrast, the core of both the MKS and NPHP complexes, which are central to ciliogenesis, have multiple C2 domains that can be traced back to the LECA (Fig. 4). This suggests that the radiation of the C2 domains was central to differentiation of the nuclear and ciliary compartments, with these domains shaping unique membrane-associated structures that came to define ciliary function. Thus, in a sense, in the case of ciliogenesis, ontology, i.e., development of the cilia dependent on the MKS and NPHP complexes, follows evolution. Currently, C2 domains are not known outside of eukaryotes; hence, it is unclear if they were a eukaryotic innovation or emerged from a preexisting prokaryotic version. Nevertheless, it is clear that a major part of their early radiation specifically occurred in the context of the eukaryotic ciliary apparatus. Given that number of lines of evidence point toward the bacterial endosymbiont and genetic material contributed by it being critical for the origin of the nucleus in the ancestral eukaryote, it is likely that the main steps in the emergence of cilia also happened after the endosymbiotic event. Indeed, in support of this, the currently available evidence points toward a bacterial origin for the TTL ATP-grasp domains that are central to microtubular modifications in ciliary function.

In conclusion, the findings reported here offer certain key testable hypotheses regarding ciliogenesis and ciliary function. First, the identification of both inactive and potentially active versions of the TGL domain help identify an important determinant for protein-protein interactions in the ciliary transitional zone and predict a potential proteolytic processing activity in the ciliary compartment that could target microtubule modifications or transported proteins. Second, our discovery of multiple new C2 domains in ciliary components helps in a conceptual reconstruction of the Y-shaped linkers in the transitional zone, with multiple membrane-contacting sites that could help both in the association of ciliary cytoskeleton with the membrane and also membrane-linked intra-ciliary trafficking. Importantly, the identification of novel C2 domains in CEP76, AH11/Joubolin and C2CD3 proteins supports their function in close proximity to the core MKS and NPHP complexes. We hope that further tests of the predictions presented here would help in a better understanding of eukaryotic ciliogenesis and also clarify the biochemical basis for a number of human ciliopathies.

**Materials and Methods**

Iterative profile searches with the PSI-BLAST and JHACKHMMER programs were used to retrieve homologous sequences in the protein non-redundant (NR) database at National Center for Biotechnology Information (NCBI). For most searches, a cut-off e-value of 0.01 was used to assess significance. In each iteration, the newly detected sequences that had e-values lower than the cut-off were examined for being false positives. Similarity-based clustering was performed using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html) to remove the highly similar sequences. Multiple sequence alignments were built by Kalign and Muscle programs, followed by manual adjustments based on profile-profile alignment, secondary structure prediction and structural alignment. Consensus secondary structures were predicted using the JPred program. Protein remote homology relationship was detected by sequence-profile comparisons with the...
PSI-BLAST program and profile-profile comparisons with the HHpred program.\textsuperscript{83} Phylogenetic analysis was conducted using an approximately maximum-likelihood method implemented in the FastTree 2.1 program under default parameters.\textsuperscript{84} The tree was rendered using the MEGA Tree Explorer.\textsuperscript{85} For bacterial TGL genes, their gene neighborhoods were extracted and analyzed. The protein sequences of all neighbors were clustered using the BLASTCLUST program to identify related sequences in gene neighborhoods. Each cluster of homologous proteins was then assigned an annotation based on the domain architecture or shared conserved domain. A complete list of Genbank Gis for proteins investigated in this study are provided in the Supplemental Material.

Species abbreviations: Aaeg, \textit{Aedes aegypti}; Aano, \textit{Auroroccus anophegefferens}; Addar, \textit{Anopheles darlingi}; Agam, \textit{Anopheles gambiae}; Alai, \textit{Albulgalabachii}; Amel, \textit{Apsi mellifera}; Aapi, \textit{Acythosiphon pisum}; Aque, \textit{Aphrimedon queenslandica}; Asuu, \textit{Ascaris suum}; Bany, \textit{Batillus amyloglucosienes}; Bden, \textit{Batrachochytrium dendrobatidis}; Bmal, \textit{Brugia malayi}; Bmar, \textit{Bmanellana marisrubri}; Bmar, \textit{Blastoporella marina}; Ceae, \textit{Caenorhabditis elegans}; Cinta, \textit{Ciona intestinalis}; Cphy, \textit{Chalmydiotherma phytotermens}; Cqui, \textit{Culex quinquefasciatus}; Crei, \textit{Chlamydomonas reinhardtii}; Csin, \textit{Clorochinis sinensis}; Cvar, \textit{Chlorella variabilis}; Dmel, \textit{Drosophila melanogaster}; Drer, \textit{Danio rerio}; Ehar, \textit{Ehaholobigen binarinise}; Eoli, \textit{Emiticia oligotrigona}; Esil, \textit{Ectocarps siliculosus}; Gbem, \textit{Geobacter bermoidiensis}; Glam, \textit{Gladia lambiら}; Hmag, \textit{Hydra magannipilla}; Hoch, \textit{Halantium ochraceum}; Hsal, \textit{Harpaphe salator}; Hsap, \textit{Homo sapiens}; Imul, \textit{Ichthyoplospira multifilis}; Libra, \textit{Leishmania braziliensis}; Ldon, \textit{Leishmania donovani}; Linf, \textit{Leishmania infantum}; Lloa, \textit{Loa loa}; Lmaj, \textit{Leishmania major}; Mbre, \textit{Monastiga brevicollis}; Mocc, \textit{Metaseiulus occidentalis}; Mps, \textit{Micromonas pusilla}; Mrox, \textit{Megabaculum rotundata}; Msp, \textit{Micromonas sp}; Ncan, \textit{Neospora caninum}; Ngru, \textit{Nagleria gruberi}; Nvec, \textit{Nematostella vectensis}; Odio, \textit{Oikopleura dioica}; Pinf, \textit{Phytophthora infestans}; Pmar, \textit{Perkinsus maricus}; Ppap, \textit{Plecostylus pacificus}; Ppat, \textit{Psychotria potentilam}; Psjo, \textit{Psychotria sojae}; Ppet, \textit{Paramacium tetraurelia}; Rvory, \textit{Rhizobius oryzae}; Shel, \textit{Schla;kia heliotrichiresced}; Skow, \textit{Saccoglossus kowalevskii}; Sman, \textit{Schisto soma mansonii}; Smoe, \textit{Selaginella moellendorfii}; Spiu, \textit{Strongylocentrotus purpuratus}; Ssp, \textit{Strongypogoa sp}; Tadu, \textit{Triophoxis adhaerens}; Tbru, \textit{Trypanosoma brucei}; Tcs, \textit{Trichobium castaneum}; Tcon, \textit{Trypanosoma congoense}; Tcru, \textit{Trypanosoma cruzi}; Tpse, \textit{Thalassiosira pseudonana}; Tsp1, \textit{Trichinella spiralis}; Ttbe, \textit{Tetrahynma thermophila}; Tvag, \textit{Trichomonas vaginalis}; Vcar, \textit{Volvox carteri}; Vmar, \textit{Verrucosipora maris}; Vpar, \textit{Variorovax paradoxus}.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Supplemental Material
Supplemental material may be downloaded here: www.landesbioscience.com/journals/cc/article/22068/
25. Hampel V, Hug L, Leigh JW, Dacks JB, Lang BE, Simpson AG, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc Natl Acad Sci USA 2004; 101:16959-64; PMID:15192375; http://dx.doi.org/10.1073/pnas.0407888101.

26. Hodges ME, Schumann N, Wickstead B, Langdale JA, Gull K. Reconstructing the evolutionary history of the centrosome from protein components. J Cell Sci 2010; 123:1407-13; PMID:20368874; http://dx.doi.org/10.1242/jcs.066473.

27. Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Bertencourt-Dias M. Evolution: Tracking the origins of centrioles, cilia, and flagella. J Cell Biol 2011; 194:165-75; PMID:21788366; http://dx.doi.org/10.1083/jcb.201011132.

28. Jin H, White SR, Shiota T, Schulz A, Aigou G, Giogi SP, et al. The conserved Bardet-Biedl syndrome protein complex assembles a coat that traffics membrane proteins to cilia. Cell 2010; 141:1288-19; PMID:20603001; http://dx.doi.org/10.1016/j.cell.2010.05.015.

29. Jékely G. Small GTPases and the evolution of the eukaryotic cytoskeleton. Bioessays 2003; 25:1129-38; PMID:14579253.

30. Mans B, Anantharaman V, Aravind L, Koonin EV. Comparative genomics, evolution and origins of the novel C2 domain superfamily and nuclear pore complex. Cell Cycle 2004; 3:1612-37; PMID:15611647; http://dx.doi.org/10.4161/cc.3.12.1316.

31. Jékely G, Arendt D. Evolution of intrflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. Bioessays 2006; 28:191-8; PMID:16439390.

32. Kee HL, Duhinger JB, Blaus IT, Liu CJ, Margolis B, Verhey KJ. A size-exclusion permeability barrier underlying amino acid transport involves Glu and Glu in both the centrosomal and nuclear pore complexes. Proc Natl Acad Sci USA 2009; 106:3859-64; PMID:19237557; http://dx.doi.org/10.1073/pnas.0811564106.

33. Zhang D, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol 2008; 38:1-31; PMID:17949255; http://dx.doi.org/10.1016/j.ijpara.2007.08.018.

34. Aravind V, Koonin EV. Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold. Hum Mol Genet 2001; 10:1627-30; PMID:11867565; http://dx.doi.org/10.1093/hmg/10.16.1627.

35. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol 2009; 39:1263-75; PMID:19211911; http://dx.doi.org/10.1016/j.ijpara.2009.04.027.

36. Snider J, Houry WA. MoxR AAA+ (ATPases: a novel family of molecular chaperones? J Struct Biol 2006; 156:72-83; PMID:16919475; http://dx.doi.org/10.1006/jsbi.2005.583.

37. Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases related to P-loop GTPases and related ATPases. J Mol Biol 2002; 317:41-72; PMID:11916378; http://dx.doi.org/10.1006/jmbi.2001.5378.

38. Yariv MJ, Schmitt T, Tjärnberg A, Guala D, Anantharaman V, Zhang D, De Souza RF, et al. Comparative polymerase systems: comprehensive and detailed examination of the ultraconserved N-terminal domain. J Comput Biol 2003; 75:979-87; PMID:12804900; http://dx.doi.org/10.1089/107092603515493620.

39. Arns HH, Doherry D, van Beersum SE, Parisi MA, Lettenboer SJ, Gorden NT, et al. Mutations in the gene encoding the basal body protein RPGRIP1L, a nephrocystin-4 interactor, cause Joubert syndrome. Nat Genet 2007; 39:88-92; PMID:17558407; http://dx.doi.org/10.1038/ng.1082.

40. Araniz PA, Bennett CL, Eckert ML, Dobyns WB, Gleeson JG, Shaw DW, et al. The NPHP1 gene deletion associated with juvenile nephropathia is present in a subset of individuals with Joubert syndrome. Am J Hum Genet 2004; 75:82-91; PMID:15138899; http://dx.doi.org/10.1086/421846.

41. Wikc AK, Wide C, Biagi T, Rosprad EO, Bjorkes E, Lindblad-Toh K, et al. A deletion in nephropathia 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund. Genome Res 2008; 18:1415-21; PMID:18687788; http://dx.doi.org/10.1101/gr.074302.107.

42. Hoover AN, Wynkoop A, Zeng H, Jia N, Niwander LA, Liu A. C2c3d is required for cilium formation and Hedgehog signaling in mice. Development 2008; 135:4049-58; PMID:19004860; http://dx.doi.org/10.1242/dev.029835.

43. Dixon-Salazar TA, Silhavy JL, Marsh SE, Louis CM, Scott LC, Gururaj A, et al. Mutations in the AHI1 gene, encoding jouberin, cause Joubert syndrome with retinal polymegathism. Am J Hum Genet 2004; 75:79-87; PMID:15467982; http://dx.doi.org/10.1086/425018.

44. Anantharaman V, Aravind LM. Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation. Annu Rev Microbiol 2001; 65:45-73; PMID:11072439; http://dx.doi.org/10.1146/annurev.micro.61.080706.093309.

45. See O, Bay LM, Schulz NP, Beek JS, Zagh Q, Slusarski DC, et al. RB5/RB11 and RB5/112 form a complex with CCT/TRiC family chaperonins and mediate BBsome assembly. Proc Natl Acad Sci USA 2010; 107:1488-93; PMID:20800638; http://dx.doi.org/10.1073/pnas.0910286107.

46. Doré AS, Kulkenny ML, Riezlerneck JZ, Pearl LH. Crystal structure of the rad9-rad1-hus1 DNA damage checkpoint complex—implications for clamp loading and regulation. Mol Cell 2009; 34:73-85; PMID:19446481; http://dx.doi.org/10.1016/j.molcel.2009.04.027.

47. Smith DM, Benaroudj N, Goldberg A. Proteasomes and their associated ATPases: a destructive combination. J Struct Biol 2006; 156:72-83; PMID:16194735; http://dx.doi.org/10.1016/j.jmb.2006.04.012.

48. Coene KL, Mans DA, Boldt K, Gloeckner CJ, van Reeswick J, Bolat E, et al. The ciliopathy-associated protein homologs RPGRIP1 and RPGRIP1L are linked to cilium integrity through interaction with Nek6 serine/threonine kinase. Hum Mol Genet 2011; 20:3952-60; PMID:21685204; http://dx.doi.org/10.1093/hmg/ddr280.

49. Mahajan B, Sivapandian A, Gerald NJ, Majam V, Zheng H, Wickramarachchi T, et al. Centrin, cell cycle regulation proteins in human malaria parasite Plasmodium falciparum. J Biol Chem 2008; 283:31871-83; PMID:18693242; http://dx.doi.org/10.1074/jbc.M108002200.

50. Aravind V, Anantharaman V, Zhang D, De Souza RF. Iyer LM. Gene flow and biological conflicts in the origin and evolution of eukaryotes. Frontiers in Cellular and Infection Microbiology 2012; 2.

51. Zhang D, De Souza RF, Anantharaman V, Iyer LM. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processes, mechanisms of action, immunity and ecology using comparative genomics. Biol Direct 2012; 7:18; PMID:22731697.
68. Nounoura T, Takaki Y, Kakura J, Nishi S, Sagahara J, Kazama H, et al. Insights into the evolution of Archaea and euukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res 2011; 39:3204-23; PMID:21169198; http://dx.doi.org/10.1093/nar/gkr228.

69. Burroughs AM, Iyer LM, Aravind L. Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system. Mol Biosyst 2011; 7:2261-77; PMID:21547297; http://dx.doi.org/10.1039/c1mb05061c.

70. Koonin EV, Aravind L. Dynein light chains of the Roadblock/LIC7 group belong to an ancient protein superfamily implicated in NTPase regulation. Curr Biol 2000; 10:R774-6; PMID:11084347.

71. Mierzeckie M, Koerner C, Vetter IR, Keilberg D, Hot E, Leonardy S, et al. Structural analysis of the Ras-like G protein MglA and its cognate GAP MglB and implications for bacterial polarity. EMBO J 2011; 30:4185-97; PMID:21847100; http://dx.doi.org/10.1038/emboj.2011.291.

72. Zhang Y, Franco M, Ducret A, Mignot T. A bacterial Ras-like small GTP-binding protein and its cognate GAP establish a dynamic spatial polarity axis to control directed motility. FLSx Biol 2010; 8:e1000430; PMID:20652023; http://dx.doi.org/10.1038/flsx.2010.430.

73. Oliva MA, Martin-Galiano AJ, Sakaguchi Y, Andreu JM. Tubulin homolog TubZ in a phage-encoded partition system. Proc Natl Acad Sci USA 2012; 109:7711-6; PMID:22538818; http://dx.doi.org/10.1073/pnas.1121546109.

74. Busiek KK, Margolin W. Split decision: a thamarchaeon encoding both FtsZ and Cdv cell division proteins chooses Cdv for cytokinesis. Mol Microbiol 2011; 82:535-8; PMID:21985799; http://dx.doi.org/10.1111/j.1365-2958.2011.07833.x.

75. Li N, Xu W, Kumarawatani M, Schumacher MA. Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. Proc Natl Acad Sci USA 2010; 107:11763-8; PMID:20534443; http://dx.doi.org/10.1073/pnas.1003817107.

76. Pyle EA, Lindas AC, Martens-Habbena W, de la Torre JR, Stahl DA, Bernander R. Cdv-based cell division and cell cycle organization in the thaumarchaeon Nitrosopumilus maritimus. Mol Microbiol 2011; 82:555-66; PMID:21923770; http://dx.doi.org/10.1111/j.1365-2958.2011.07834.x.

77. Burroughs AM, Iyer LM, Aravind L. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. Genome Dyn 2007; 3:48-65; PMID:18733784; http://dx.doi.org/10.1016/j.combinatorics.2011.07834.x.

78. Aertschul SE, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-402; PMID:92846094; http://dx.doi.org/10.1093/nar/25.17.3389.

79. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 2010; 11:431; PMID:20718988; http://dx.doi.org/10.1186/1471-2105-11-431.

80. Laszlo D, Sonnhammer EL. Kalgin--an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics 2005; 6:298; PMID:16343337; http://dx.doi.org/10.1186/1471-2105-6-298.

81. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32:1792-7; PMID:15034147; http://dx.doi.org/10.1093/nar/gkh340.

82. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics 1998; 14:892-3; PMID:9927721; http://dx.doi.org/10.1093/bioinformatics/14.10.892.

83. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics 2005; 21:951-60; PMID:15531603; http://dx.doi.org/10.1093/bioinformatics/bti125.

84. Rice MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009; 26:1641-50; PMID:19377059; http://dx.doi.org/10.1093/molbev/msn092.