

Mosaic origin of the eukaryotic kinetochore

Eelco C. Tromer^{a,b,c,1,2}, Jolien J. E. van Hooff^{a,b,1}, Geert J. P. L. Kops^{b,d,3}, and Berend Snel^{a,2,3}

^aTheoretical Biology and Bioinformatics, Biology, Science Faculty, Utrecht University, 3584 CH Utrecht, The Netherlands; ^bOncode Institute, Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, 3584 CT Utrecht, The Netherlands; ^cDepartment of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom; and ^dUniversity Medical Centre Utrecht, 3584 CX Utrecht, The Netherlands

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved April 30, 2019 (received for review December 24, 2018)

The emergence of eukaryotes from ancient prokaryotic lineages embodied a remarkable increase in cellular complexity. While prokaryotes operate simple systems to connect DNA to the segregation machinery during cell division, eukaryotes use a highly complex protein assembly known as the kinetochore. Although conceptually similar, prokaryotic segregation systems and the eukaryotic kinetochore are not homologous. Here we investigate the origins of the kinetochore before the last eukaryotic common ancestor (LECA) using phylogenetic trees, sensitive profile-versus-profile homology detection, and structural comparisons of its protein components. We show that LECA's kinetochore proteins share deep evolutionary histories with proteins involved in a few prokaryotic systems and a multitude of eukaryotic processes, including ubiquitination, transcription, and flagellar and vesicular transport systems. We find that gene duplications played a major role in shaping the kinetochore; more than half of LECA's kinetochore proteins have other kinetochore proteins as closest homologs. Some of these have no detectable homology to any other eukaryotic protein, suggesting that they arose as kinetochore-specific folds before LECA. We propose that the primordial kinetochore evolved from proteins involved in various (pre)eukaryotic systems as well as evolutionarily novel folds, after which a subset duplicated to give rise to the complex kinetochore of LECA.

kinetochore | mitosis | LECA | eukaryogenesis | gene duplication

During cell division, eukaryotes divide their duplicated chromosomes over both daughter cells by means of a microtubule-based apparatus called the spindle. Central to this process are kinetochores, large multiprotein structures that are built on centromeric DNA and connect chromosomes to microtubules. Although species vary hugely in how they exactly coordinate and execute chromosome segregation (1–4), all eukaryotes use a microtubule-based spindle, and thus the last eukaryotic common ancestor (LECA) likely featured one as well (Fig. 1A). Consequently, LECA's chromosomes probably contained a centromere and assembled a kinetochore. The centromeric DNA sequences of current-day eukaryotes are strikingly different across species and in fact are too diverse to allow reconstruction of LECA's centromeric sequences (5). In contrast, their conserved kinetochore components (6–9) did allow for the inference of LECA's kinetochore (10).

The LECA kinetochore was not directly derived from a prokaryote, because prokaryotes link their DNA to the segregation machinery via protein assemblies that are not homologous to the eukaryotic kinetochore (11–13) (Fig. 1A). Thus, like many other uniquely eukaryotic cellular systems, the LECA kinetochore must have originated after the first eukaryotic common ancestor (FECA) diverged from prokaryotes. Between FECA and LECA, the pre-eukaryotic lineage evolved from relatively simple and small prokaryotic cells to complex, organelle-bearing cells organized in a fundamentally different manner, a process referred to as “eukaryogenesis.” Uncovering the evolutionary events underlying eukaryogenesis is a major scientific endeavor (14) undertaken by investigating specific eukaryotic systems (15). Studies of, for example, the spliceosome, the intracellular membrane system, and the nuclear pore have revealed that repurposed prokaryotic genes played a role in their origin, as did evolutionarily novel, eukaryote-specific genes and gene duplications, albeit at varying degrees and in different ways (16–18).

In this study, we addressed the question of how the kinetochore originated. Leveraging the power of detailed phylogenetic analyses, improved sensitive sequence searches, and new structural insights, we traced the evolutionary origins of the 52 proteins that we now assign to the LECA kinetochore. Based on our findings, we propose that the LECA kinetochore was of mosaic origin; it contained proteins that shared ancestry with proteins involved in various core eukaryotic processes, as well as potentially novel proteins. After recruitment to a primordial (pre-LECA) kinetochore, many of these proteins duplicated, accounting for a 60% increase in kinetochore extent and thereby for the complex LECA kinetochore.

Results

LECA's Kinetochore. To study how the LECA kinetochore originated, we first needed to determine what proteins constituted it. While we reconstructed the LECA kinetochore previously (10), here we extend our analyses with Nkp1, Nkp2, and Csm1 (19) (*SI Appendix, Text*). For each protein present in human and yeast kinetochores, we asked (*i*) whether it was likely encoded in the genome of LECA, based on its distribution across the eukaryotic tree of life, and (*ii*) whether it likely operated in the LECA kinetochore, based on functional information. Following these criteria, we now propose that the LECA kinetochore consisted of at least 52 proteins (Fig. 1B and *SI Appendix, Table S2*), including the constitutive centromere-associated network (CCAN). Of note, based on various lines of evidence, we infer that the KKT/KKIP

Significance

Eukaryotes, which include diverse species like animals, fungi, and plants, have cells that are fundamentally more complex than prokaryotic cells, such as bacteria. However, eukaryotes did evolve from prokaryotes, so they must have acquired this cellular complexity after they diverged from prokaryotes. A key cellular feature unique to eukaryotes is the kinetochore, a large, multiprotein structure that plays an essential role in cell division. Here we shed light on the origination of the kinetochore by studying the evolution of its proteins. We find that the kinetochore has diverse evolutionary roots and that it expanded via gene duplications. We present a mode by which eukaryotic systems originated and illuminate the prokaryote-to-eukaryote transition.

Author contributions: E.C.T., J.J.E.v.H., G.J.P.L.K., and B.S. designed research; E.C.T. and J.J.E.v.H. performed research; E.C.T. and J.J.E.v.H. contributed new reagents/analytic tools; E.C.T., J.J.E.v.H., G.J.P.L.K., and B.S. analyzed data; and E.C.T., J.J.E.v.H., G.J.P.L.K., and B.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

See Commentary on page 12596.

¹E.C.T. and J.J.E.v.H. contributed equally to this work.

²To whom correspondence may be addressed. Email: ect2@cam.ac.uk or b.snel@uu.nl.

³G.J.P.L.K. and B.S. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821945116/-DCSupplemental.

Published online May 24, 2019.

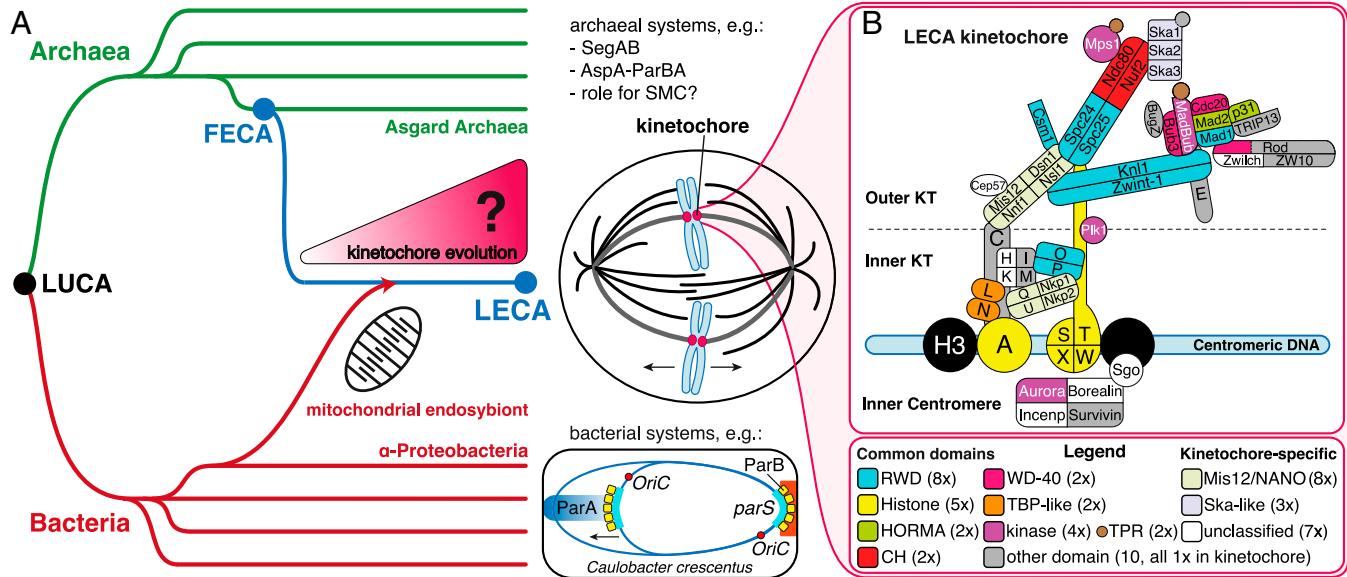


Fig. 1. The eukaryotic kinetochore and mitotic machinery originated between FECA and LECA. (A) How did the eukaryotic kinetochore originate and evolve between FECA and LECA? Eukaryotes (blue) are descended from Archaea (green) and likely are closely related to the Asgard archaeal superphylum (59). This Asgard-related lineage incorporated an Alphaproteobacterium via endosymbiosis; the latter gave rise to the eukaryotic mitochondrion. Archaea and Bacteria (red) do not separate their duplicated chromosome(s) via a mitotic spindle (11–13). For example, bacteria such as *Caulobacter crescentus* operate the *parABS* partitioning system, in which *parS* DNA sites are recognized by the protein ParB, stimulating ParA, which in turn pulls or pushes the chromosomes apart (12). Due to these differences, the mitotic spindle and the kinetochore probably originated between the FECA and the LECA. LUCA, the last universal common ancestor. (B) The kinetochore of LECA consisted of 52 proteins that contain domains found in other, nonkinetochore eukaryotic proteins as well (“common domains”) or that are unique to the kinetochore (“kinetochore-specific”). KT, kinetochore.

proteins of the analogous kinetochore system found in kinetoplastids (7, 8) likely were not part of the LECA kinetochore (*SI Appendix, Text*).

Identifying Ancient Homologs of Kinetochore Proteins. To elucidate the ancient, pre-LECA homologs (either eukaryotic or prokaryotic) of LECA kinetochore proteins, we applied sensitive profile-versus-profile similarity searches (*Dataset S1*), followed by phylogenetic tree constructions (*SI Appendix, Fig. S1*), or, when available, published phylogenetic tree interpretations. If literature and/or structural comparisons provided additional information, we included these as an indication of a homologous relationship (*Dataset S2*). For each LECA kinetochore protein, we aimed to identify the protein that was its closest homolog before LECA (*SI Appendix, Table S1*). These proteins were classified as eukaryotic or prokaryotic, and as kinetochore or non-kinetochore (*SI Appendix, Data and Methods*).

Because different domains in a single protein may have had separate evolutionary histories before they joined, we searched primarily for homologs of LECA kinetochore domains. If from this analysis we deduced that multiple domains of a single LECA kinetochore protein share their evolutionary history, we report these as a single “domain” in *SI Appendix, Table S1*.

We inferred the closest homologs of kinetochore proteins on the domain level, using gene phylogenies for 17 of the 55 domains (31%), profile-versus-profile searches for 2 domains (3%), and structural information for 8 domains (15%). For 12 other domains (22%), we used a combination. For a total of 39 domains, we could identify the closest homolog. For eight (15%) of the remaining proteins, we found homologs but could not determine which one was closest, and for the other eight (15%), we could not find any ancient homologs (*SI Appendix, Table S1*).

Evolutionary Histories of Kinetochore Proteins. Here we discuss the evolutionary history of LECA kinetochore proteins grouped according to common domains. We highlight their affiliations with other

eukaryotic cellular processes, their prokaryotic homologs, and their ancient duplications within the kinetochore (*SI Appendix, Table S1*). **Kinetochore RWD.** The RING-WD40-DEAD (RWD) domains in kinetochore proteins are highly diverged and noncatalytic members of the E2 ubiquitin-like conjugase (UBC) family (20–22) (Fig. 2). For seven RWD kinetochore proteins, 3D structures have been resolved (Fig. 2C). These form heterodimers or homodimers with either a single RWD (Spc24-Spc25, Mad1-Mad1, and Csm1-Csm1) or a tandem (CenpO-CenpP and Knl1) RWD configuration. In contrast to previous efforts (20, 23), we uncovered significant sequence similarity between Zwint-1 and other (double) RWDs, suggesting that Zwint-1 and Knl1 form an RWD heterodimer similar to CenpO-CenpP (*SI Appendix, Text and Fig. S2*). Our phylogenetic analysis (*SI Appendix, Data and Methods* and Fig. S3) revealed that kinetochore RWDs and other RWDs are more closely related to one another (bootstrap: 96/100) than to eukaryotic and archaeal E2s (bootstrap: 77/100). A single Asgard sequence clustered at the base of canonical eukaryotic RWDs, suggesting that FECA may have already contained an RWD domain.

Strikingly, most kinetochore RWDs are each other’s closest homologs (*SI Appendix, Fig. S3*), as supported by our profile-versus-profile searches (*Dataset S1*) and structural alignments (Fig. 2C and *Dataset S2*). This indicates that kinetochore RWDs possibly arose from a single ancestral kinetochore RWD. This group may also include mediator subunits (Med14/15/17) and the E3 ubiquitin ligase FancL, signifying a shared evolutionary history of these systems with the kinetochore (Fig. 2D). We were not able to reliably reconstruct the exact order in which the kinetochore RWD proteins arose. We hypothesize that kinetochore RWDs and other RWDs (i.e., Gcn2, FancL, and Med14/15/17), resulted from an extensive radiation and neofunctionalization of an archaeal noncatalytic E2 UBC during eukaryogenesis (Fig. 2D).

Histones. The LECA kinetochore contained five histone proteins: CenpA and the CenpS-X-T-W tetramer (Fig. 3A). From FECA to LECA, an archaeal-derived histone-like protein (24, 25) duplicated

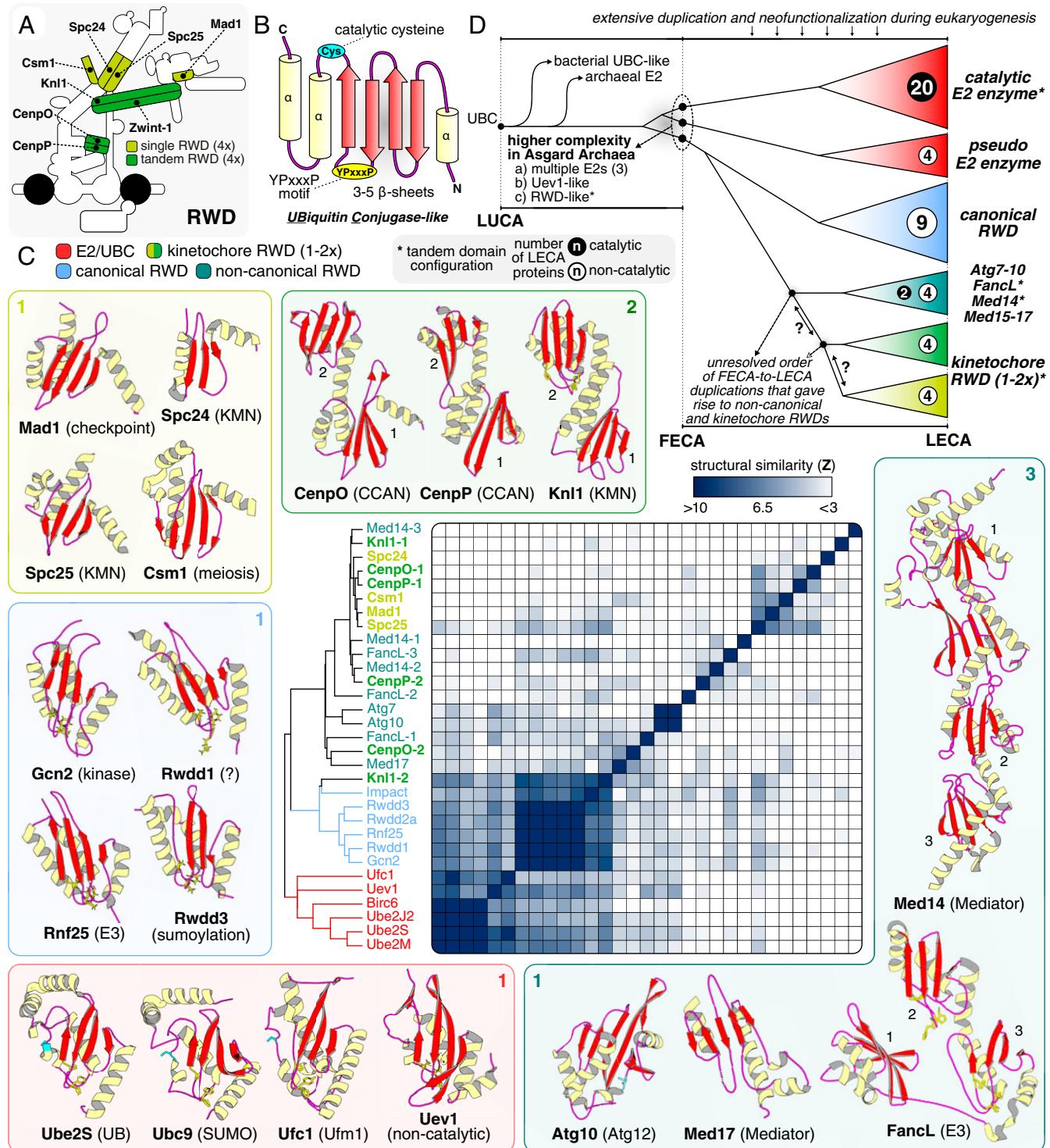


Fig. 2. Kinetochore RWDs are an expanded class of noncatalytic E2 UBCs. (A) Overview of the position of eight kinetochore proteins with a single (light green) or a tandem (dark green) RWD configuration. (B) RWD domains are part of the UBC superfamily. The secondary structure of the UBC superfamily is characterized by a “ β -meander” of three to five β -sheets enclosed by α -helices at both termini, a YPxxxP motif, and a catalytic cysteine residue (lost in RWDs). (C) The UBC superfamily can be subdivided into four classes: (i) E2 UBCs (E2), including noncatalytic pseudo E2s (e.g., Uev1); (ii) canonical RWDs; (iii) kinetochore RWDs; and (iv) atypical RWD/UBC-like (e.g., Fancl, Med14-17). Per class, the structure of various members is depicted to show the overall structural and topological similarity, and a known molecular function is indicated between brackets. When present, YPxxxP (yellow) and the catalytic cysteine (cyan) are represented in a sticks configuration. The average linkage clustering of structural similarity scores of single UBC domains (z -scores) demonstrates the close similarity amongst E2s and canonical RWD domains. Kinetochore RWDs and noncanonical domains are more divergent and cluster together. (D) Cartoon of the evolutionary reconstruction of the UBC superfamily based on phylogenetic analyses (SI Appendix, Figs. S1E and S3) and structural comparisons (Dataset S2). Extensive duplication and neofunctionalization of an archaeal E2 UBC gave rise to a large complexity of catalytic and noncatalytic E2/RWD proteins in LECA (see numbers per class). Possibly a part of this eukaryotic complexity was already present in FECA, since Asgard Archaea contain multiple E2 conjugases, an Uev-like homolog, and an RWD-like domain (SI Appendix, Fig. S3). Kinetochore RWDs might have a monophyletic origin, although a structural affiliation with other divergent proteins signify a more complex evolutionary scenario (see question marks).

many times, giving rise to proteins involved in all aspects of eukaryotic chromatin complexity (Fig. 3C). CenpA is a centromere-specific histone H3 variant and resulted from a pre-LECA dupli-

cation (10, 25). CenpS-X-T-W arose by two duplications: CenpS-T (bootstrap: 99/100) and CenpX-W (bootstrap: 77/100), indicating a coduplication of the two subunits of an ancestral heterodimer

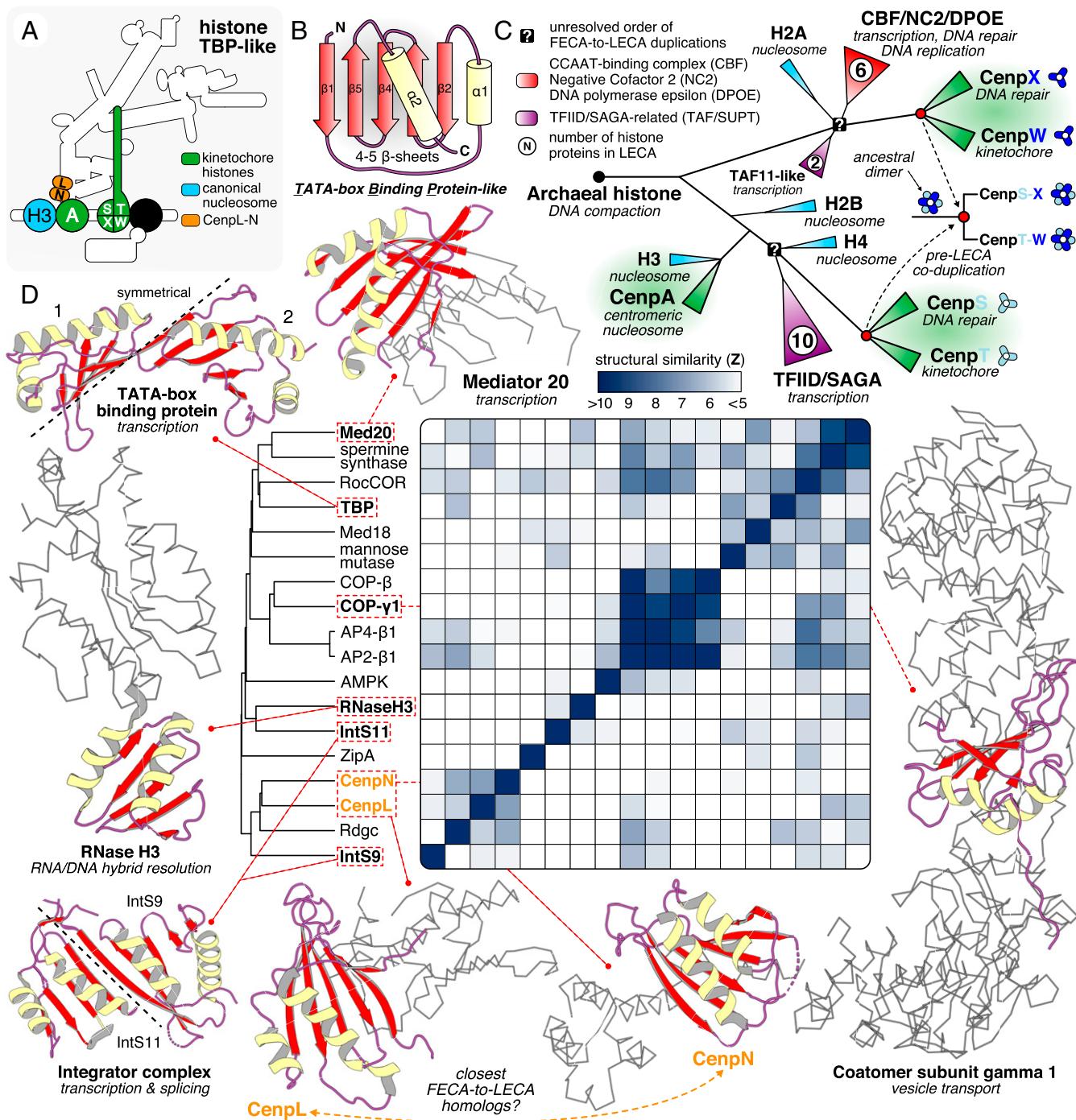


Fig. 3. A common origin of kinetochore histones and TBP-like proteins with complexes involved in DNA repair and transcription. (A) Overview of the position of CenpA and CenpS-X-T-W (histones, green) and CenpL-N (TBP-like, orange) in the kinetochore. (B) The TBP-like fold is a set of curved β -strands that form an interaction surface for substrates (RNA/DNA, amino acid motifs) and potential dimer interfaces. (C) A cartoon of the evolutionary reconstruction of kinetochore-related histone proteins CenpA and CenpS-X-T-W (based on *SI Appendix, Fig. S1*). A histone of archaeal descent duplicated and subfunctionalized many times, giving rise to a large diversity of histone proteins in eukaryotes, including those involved in the kinetochore, chromatin structure (nucleosome), transcription (TAF/SUPT/NC2/CBF), and DNA repair (DPOE). CenpA is the closest homolog of the nucleosomal histone H3. CenpS-T and CenpX-W are likely each other's closest paralogs, signifying a coduplication of an ancient dimer to form the tetramer CenpS-X-T-W. The CenpS-X dimer also plays a role in the Fanconi anemia pathway (DNA repair). (D) Yellow (helices) and red (sheets) show the location of a TBP-like domain in a subset of available TBP-like protein structures. The gray ribbon representation indicates the nonhomologous parts of the proteins; their cellular function is indicated between brackets. CenpL and CenpN contain a TBP-like fold. Average linkage clustering of similarity scores (Z -scores) indicates that CenpN and CenpL could be each other's closest homologs.

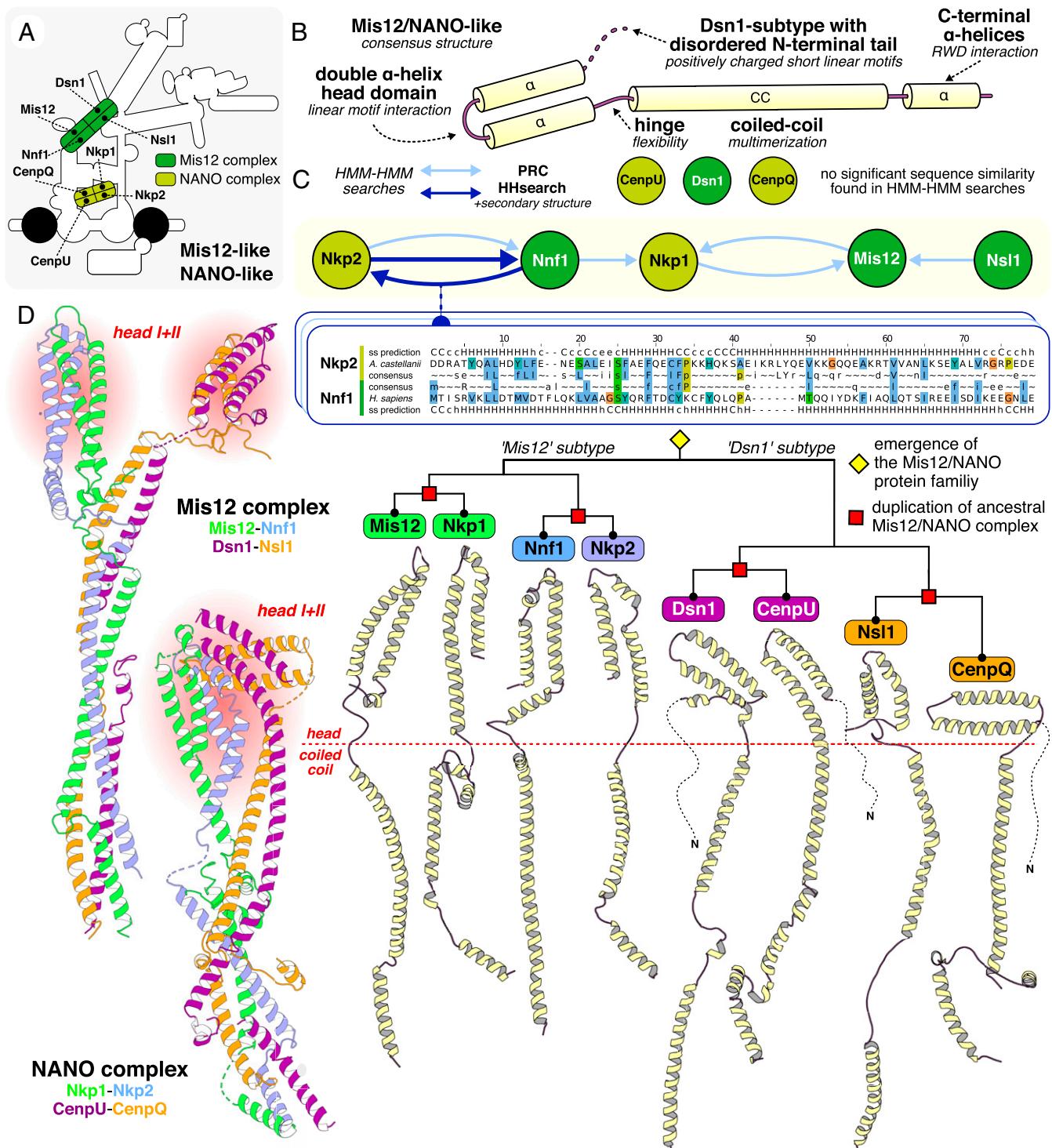


Fig. 4. The Mis12 and NANO complex have a common ancestry. (A) Overview of the position of the Mis12 complex and NANO tetramers in the LECA kinetochore. (B) Cartoon of the consensus topology of all eight Mis12/NANO subunits, illustrating disordered and globular regions. (C) Profile-versus-profile hits with HHsearch (dark blue) and PRC (light blue) indicate that Mis12, Nnf1, Nsp1, Nkp1 and Nkp2 are homologous (SI Appendix, Text and Dataset S1). No sequence similarity between CenpU, Dsn1, and CenpQ with any of the other Mis12/NANO subunits was detected. (D) The subunits of the Mis12 and NANO display a high degree of similarity with respect to the (i) size and orientation of the head domains, (ii) length of the coiled coils, and (iii) presence of disordered N-terminal tails. Based on these three criteria, we defined a 'Mis12' and 'Dsn1' subtype. We propose that the Mis12 and NANO complex are the result of an ancient whole complex duplication, which was preceded by two rounds of Mis12/Dsn1 subtype duplication. Distances in the tree do not reflect measured distances but indicate a higher degree of sequence and structural variation for the Dsn1 type compared with the Mis12 type.

(*SI Appendix, Text* and Fig. S1). We found CenpS-T to be phylogenetically affiliated to H2B-H3-H4-TFIID-SAGA-related histones, while CenpX-W clustered with H2A-CBF-NC2-DPOE-Taf11-related histones. These affiliations in combination with a primary role for CenpS-X in the Fanconi anemia pathway (26, 27) signify that the evolutionary history of the CenpS-X-T-W tetramer is highly interconnected with the origin of the eukaryotic transcription and DNA repair machinery.

TBP-like. CenpN and CenpL harbor a fold similar to the DNA-binding domain of the TATA box-binding protein (TBP) (28–30) (Fig. 3). Although we did not observe any significant sequence similarity for CenpL and CenpN (**Dataset S1**), we found previously reported structural similarity with proteins that function in nucleotide metabolism (e.g., spermine synthase), in transcription (TBP, integrator, and mediator) and in vesicle transport (coatomers and adaptors) (31) (Fig. 3D). TBP and structurally related enzymes (e.g., RNase HIII) (31) were found in Archaea (32), suggesting that eukaryotes acquired these proteins via vertical descent (Fig. 1A). The average linkage (hierarchical) clustering of the structural similarity scores of CenpL, CenpN, and other TBP-like proteins indicates that CenpN and CenpL were most similar (z -score = 7.3), although differences among scores were small (Fig. 3D and **Dataset S2**). Since CenpL and CenpN form a heterodimer (30), we propose that they are closest homologs, and that other TBP-like proteins are more distantly related.

Mis12/NANO. Through profile-versus-profile searches, we discovered a previously hidden homology: Nkp1 and Nkp2 were found to be highly similar to Mis12 and Nnf1 (Fig. 4C). These potential homologies were confirmed by a recent paper on the yeast CCAN structure (33), which also reported striking similarities between the other subunits of the Mis12 complex (Dsn1 and Nsl1) and the Nkp1-Ame1^{CenpU}-Nkp2-Okp1^{CenpQ} tetramer, which we term the NANO complex. Structural similarity scores did not indicate any clear closest homologs (**Dataset S2**); however, we propose a shared ancestry of the Mis12-Nnf1 and

Nkp1-Nkp2 dimers that differs from that of the Dsn1-Nsl1 and CenpQ-CenpU dimers, based on (i) the positions of the subunits within their complexes, (ii) the size and position of their head domains and coiled coils, and (iii) the presence/absence of a long N-terminal disordered tail. We hypothesize that the Mis12 and NANO complexes originated by a series of duplications of an ancestral multimer-forming protein, giving rise to a heteromeric complex, followed by a (co)duplication of all its subunits (Fig. 4D). We did not detect any homologs of Mis12/NANO-like proteins outside of the kinetochore.

HORMA-Trip13. Eukaryotic HORMA domain proteins operate in the kinetochore (Mad2, p31^{comet}), autophagy (Atg13–101), DNA repair (Rev7), and meiosis (HORMAD). The HORMA proteins p31^{comet} and HORMAD are structurally modified by Trip13, an AAA+ ATPase. Bacterial genomes also encode HORMA proteins, and, interestingly, these co-occur in one operon with an AAA+ ATPase that resembles Trip13 (34). In addition, we found the HORMA-Trip13-like operon in a few archaeal species belonging to the Haloarchaea class (Fig. 5, *SI Appendix*, Fig. S5, and **Dataset S5**). The eukaryotic HORMA proteins are monophyletic, indicating FECA-to-LECA duplications (*SI Appendix*, Fig. S1F). Eukaryotic Trip13 sequences are most closely related to the prokaryotic Trip13-like sequences, and thus we designate the latter evolutionarily as Trip13 (*SI Appendix*, Fig. S1G). Based on our phylogenetic analysis, we propose that the pre-eukaryotic lineage derived the HORA-Trip13 operon via horizontal transfer from Bacteria. Because in bacteria HORMA-Trip13 is part of operons involved in nucleotide signaling (34), it might initially have fulfilled such a role in the pre-eukaryotic lineage. Subsequently, HORMA duplicated and neofunctionalized, repurposing HORMA-Trip13 for, for example, DNA repair, meiosis, and the kinetochore.

NN-Calponin Homology. Calponin homology (CH) domain proteins operate in many different processes, including binding of actin and F-actin and in various cellular signaling pathways (35). In the

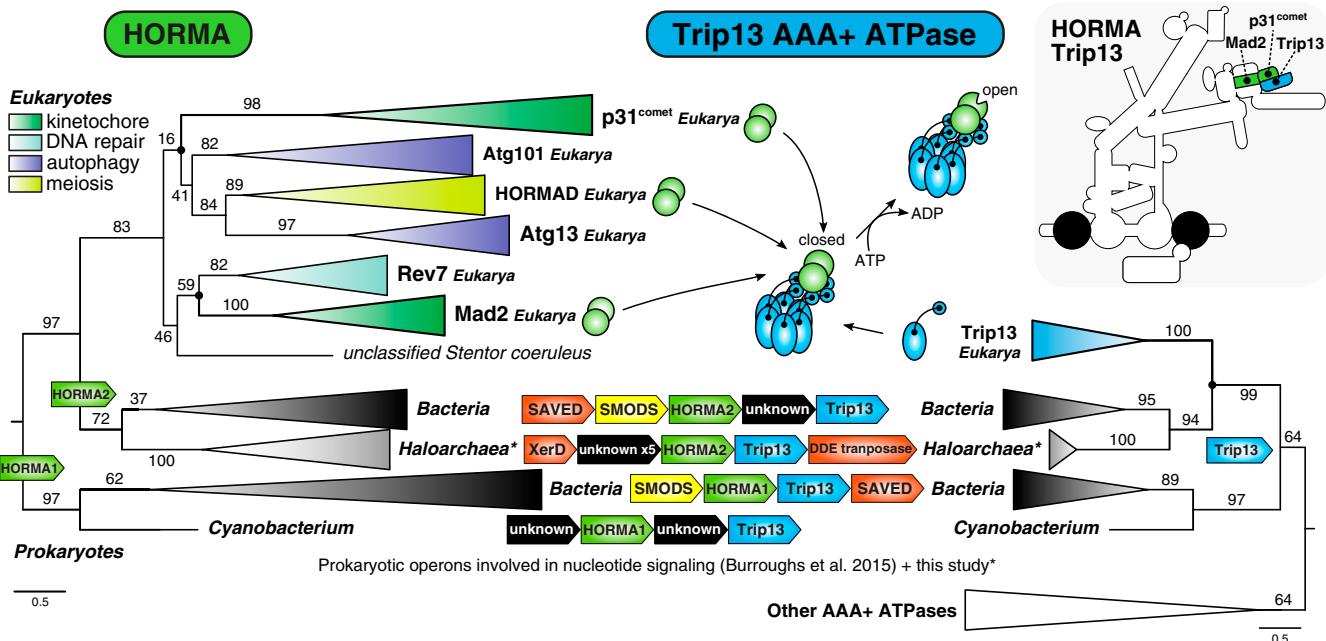


Fig. 5. The HORMA-Trip13 module is of prokaryotic origin. Shown are phylogenetic trees of HORMA domain proteins and AAA+ ATPases. In eukaryotes, HORMAD, Mad2, and p31^{comet} are structurally modified by a Trip13 hexamer (Upper, right side). In prokaryotes, HORMA (types 1 and 2) and Trip13 are present in a single operon, strongly suggesting that they also interact in these species, and thus that this interaction is ancient. The phylogenetic trees indeed suggest that the eukaryotic HORMA domain and Trip13 were derived from prokaryotes. In addition, the prokaryotic operons include proteins involved in nucleotide signaling [yellow, nucleotide transferase (SMODS); red, transposase-related; black, unknown] (34). The uncollapsed trees are shown in *SI Appendix*, Figs. S1 F and G. Asterisks indicate the species for which we discovered a HORMA-Trip13 operon (annotation in **Dataset S5** and *SI Appendix*, Fig. S5). *SI Appendix*, Figs. S1 F and G. Asterisks indicate the species for which we discovered a HORMA-Trip13 operon (annotation in **Dataset S5** and *SI Appendix*, Fig. S5).

kinetochore, Ndc80 and Nuf2 are the predominant microtubule-binding proteins. The ancestral function of the CH domain, which to our knowledge has not been found in prokaryotes, is not known. Ndc80 and Nuf2 have been reported to be part of a highly divergent subfamily of CH proteins (NN-CH) (36), which includes proteins involved in intraflagellar transport, ciliogenesis, the centrosome, vesicle-trafficking, and RNA transport (37–40). This NN-CH subfamily may be specialized toward binding microtubules, implying that the kinetochore function reflects the ancestral function (36).

Kinases and TPR. In a detailed eukaryotic kinase phylogeny, the kinetochore kinases Polo (Plk) and Aurora were closely related (*SI Appendix*, Fig. S1D). The closest relative of Plk is Plk4, probably signaling an ancestral function for Plk in centrosome/basal body function, since Plk is also still found at the centrosome. Aurora diverged from a duplication before the Plk-Plk4 divergence, suggesting that Plk and Aurora independently gained kinetochore functions after duplication. Alternatively, the Plk-Aurora ancestor operated in both the centrosome and the kinetochore, and Plk4 lost its kinetochore function. The polo box arose N-terminal to the ancestral Plk kinase domain after Aurora split off. The closest relative of Mps1 was Tlk (bootstrap: 36/100). The closest homolog of MadBub is an uncharacterized group of kinases. Interestingly, in contrast to their kinase domains, the TPR domains of Mps1 and MadBub are most closely related, as determined by profile-versus-profile searches (*Dataset S1*). This implies that the Mps1 and MadBub TPR domains joined with a kinase domain independently, as we reported previously (41). TPR domains have been found in many prokaryotes, and their presence in the prokaryotic ancestors of eukaryotes has been suggested but not confirmed (42).

Coats and Tethers. Zw10 homologs are involved in vesicle transport (43–45). Their closest homolog is Cog5, which is involved in intra-Golgi transport (*SI Appendix*, Fig. S14). Zw10 participates in two complexes: RZZ (Rod-Zwilch-Zw10), localized to the kinetochore, and the NRZ (Nag-Rint1-Zw10), involved in Golgi-to-ER transport. Of note, Rod is most closely related to Nag (*SI Appendix*, Fig. S1H), suggesting that their ancestor interacted with Zw10 before it duplicated to give rise to Rod and Nag. Whether this ancestral complex was involved in vesicle transport, in the kinetochore, or in both is unclear.

WD40. The relatives of the WD40 kinetochore proteins are highly diverse, and their repetitive nature has made it difficult to resolve their (deep) evolutionary origins. Cdc20, a WD40 repeat protein, is most closely related to Cdh1 (*SI Appendix*, Fig. S1B), which, like Cdc20, coactivates the anaphase-promoting complex/cyclosome (APC/C) (46). Bub3's closest homolog is Rae1 (*SI Appendix*, Fig. S1C), a protein involved in nuclear mRNA export (47). For both Cdc20 and Bub3, we cannot suggest nor exclude the possibility that their ancestors were part of the kinetochore network. While WD40 repeats are clearly present in current-day prokaryotes (48), these prokaryotes may have received these repeats recently from eukaryotes via horizontal gene transfer, and thus whether WD40 domains were already present in the prokaryotic ancestors of eukaryotes is unclear.

Unique Domains in the Kinetochore? In addition to the Mis12/NANO-like proteins, various other domains, such as Ska, Zwilch, Incenp, Borealin, Shugoshin, Cep57, CepnH, and CenpK, seem to be unique to the kinetochore (*SI Appendix*, Table S1). We cannot find any nonkinetochore eukaryotic or prokaryotic homologs. Possibly these domains are truly novel, in which case they originated between FECA and LECA and have roles only in the kinetochore. Alternatively, they may in fact have homologs that we were not able to detect due to extensive sequence divergence. Such divergence may have enforced proteins to adopt a completely novel fold and function. In that case, although strictly speaking these folds would not be novel, they would represent an evolutionary innovation unique to the kinetochore.

Mosaic Origin of the LECA Kinetochore. Most LECA kinetochore proteins consisted of domains found in other eukaryotic proteins (37/55; 67%), while the others had no detectable homology outside of the kinetochore (18/55; 33%) (*SI Appendix*, Table S1). Among the proteins with common domains, only one (Trip13) was directly derived from its prokaryotic ancestors. All others had eukaryotic homologs (paralogs) that were more closely related than prokaryotic homologs (if any). These paralogs are involved in an array of eukaryotic cellular processes. Altogether, the ancient homologs of kinetochore proteins indicate that the kinetochore is of a mosaic origin. Specific eukaryotic processes were prevalent among the evolutionary links (Fig. 6). Of the 14 closest nonkinetochore homologs that we identified, 7 were involved in chromatin and/or transcription regulation (Tlk1, H3, Rev7 Med14-15-17, and FancL), 2 played a role in Golgi and ER-related vesicle transport systems (Nag and Cog5), and 1 was associated with centriole biogenesis (Plk4). More distantly related homologs were involved in DNA repair and replication (FancI, Dpoe3-4, and the replication factors Cdt1, Cdc6, and Orc1), chromatin structure (nucleosomal histones), transcriptional regulation (e.g., TBP-like: Med18, Med20, TBP; histone: TAFs, CBF/NF, NC2), RNA splicing (Fam98, Syf1/Crooked neck-like, and Integrator subunits 9 and 11), vesicle transport (Kif1C, AP-2/4B, COPg1, AP-1G, COPb, Rab1A, Ccdc22, and Ccdc93), and intraflagellar transport (Cluap1, Ift54, and Ift81). Most LECA kinetochore proteins are part of families that have many members in eukaryotes, like UBC/RWD, kinases, and histones. Such families dramatically expanded between FECA and LECA and diversified into different eukaryotic cellular processes, including the kinetochore.

In addition to their mosaic origins, many kinetochore proteins arose from intrakinetoche gene duplications. Of the 39 kinetochore domains with an identified closest homolog, 29/55 (53%) are most closely related to another kinetochore protein, indicating an important role for intrakinetoche duplications in its evolutionary origin (*SI Appendix*, Table S1). We inferred that the 55 domains resulted from 34 ancestral kinetochore units (“anc_KT” units), revealing that intrakinetoche gene duplications expanded the primordial kinetochore by a factor of ~1.6. We observed few domain fusions among LECA KT proteins—in fact, we found only three: in Mps1 and MadBub, whose TPR domains independently joined their kinase domains, and a fusion of a microtubule-binding winged helix and a Ska-like domain in Ska1 (*SI Appendix*, Table S1).

Discussion

Evolution of Eukaryotic Cellular Systems. We have shown that the kinetochore consists largely of paralogous proteins that either share deep evolutionary roots with various other eukaryotic cellular processes or are evolutionarily novel and specific to the kinetochore (Fig. 6). We here contextualize the evolutionary origin of the kinetochore by comparison with the origin of other eukaryotic cellular systems. In the origin of the kinetochore, gene duplications played a key role, which is in line with the observed elevated rate of gene duplications in eukaryogenesis (49). Duplications contributed to the expansions of, for example, the spliceosome (16), the intraflagellar transport complex (50), COPII (51), and the nuclear pore (18). However, the role of duplications in the origin of the kinetochore differs from their role in membrane-specifying complexes, in which paralogs are mainly shared between the different organelles rather than within them (52). In tethering complexes, duplications generate proteins both within and between complexes (43). When it comes to its proteins with prokaryotic roots, the kinetochore conserved certain prokaryotic biochemical functions (e.g., HORMA-Trip13 interaction, histone-DNA interaction by CenpA) but obviously no longer performs the ancestral cellular function. This evolutionary FECA-to-LECA path is in contrast to that of, for example, NADH:ubiquinone oxidoreductase

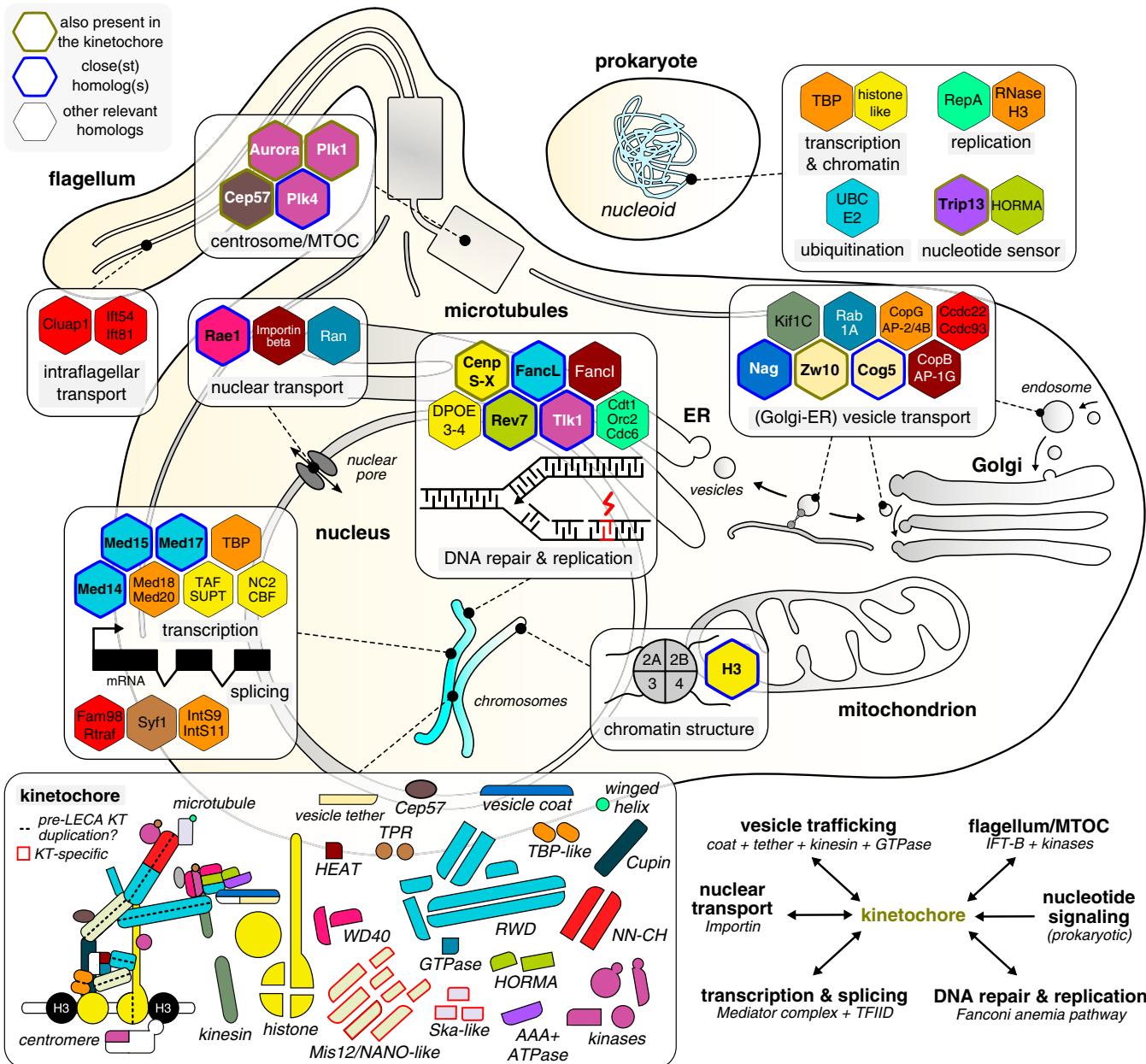


Fig. 6. Mosaic origin of the eukaryotic kinetochore. Overview of the eukaryotic and prokaryotic close(st) homologs of LECA kinetochore proteins, which play roles in a wide variety of cellular processes, signifying the mosaic origin of the eukaryotic kinetochore. Relevant eukaryotic and prokaryotic homologs (hexagons) of LECA kinetochore proteins are colored based on the presence of a common domain (Bottom Left: overview of kinetochore parts), and projected onto the location(s) in the eukaryotic cell at which they operate (SI Appendix, Table S1). The hexagons of homologs are lined with different colors indicate a LECA kinetochore protein with a nonkinetochore function (green), the closest homolog to a LECA kinetochore protein (blue), and other close homologs of LECA kinetochore proteins (black). In addition, distantly related homologs of TBP-like, histones, UBC/RWD, and HORMA domain-containing kinetochore proteins were already present in prokaryotes (Top Right). (Bottom Left) Overview of the different number and types of domains in the LECA kinetochore. The Mis12/NANO and Ska domains are kinetochore-specific and thus are not found in other systems. The dotted lines indicate a potential intrakinetochore duplication during eukaryogenesis leading to the formation of various heteromeric (sub)complexes within the kinetochore. (Bottom Right) summary of the evolutionary links between the kinetochore and selected prokaryotic/eukaryotic molecular systems.

(Complex I) (53), which was directly derived from the Alphaproteobacterium that became the mitochondrion (Fig. 1) and maintained its cellular role while expanding by incorporating additional proteins of different origins. The Golgi and ER also differ from the kinetochore, as their protein constituents have mainly archaeal roots (54). The nuclear pore, while resembling the kinetochore in having a mosaic origin, was assembled with a substantial number of proteins derived from prokaryotic ancestors (16, 18), as was the spliceosome (16, 18).

Intrakinetochore Duplication. The intrakinetochore duplications suggest an evolutionary trajectory by which the kinetochore partially expanded through homodimers that became heterodimers via gene duplication (55). A primordial kinetochore might have been composed of complexes consisting of multimers of single ancestral proteins (anc_KT in SI Appendix, Table S1). After these proteins duplicated, the resulting paralogs maintained the capacity to interact, resulting in a heteromer. For example, the Ndc80 complex might have consisted of a tetramer of two copies of an

ancient CH protein and two copies of an ancient RWD protein. According to this model, the proteins with shared domains within complexes should be most closely related to one another. This paradigm holds for the Ska, NN-CH, RWD, and the histone tetramer CenpS-X-T-W. We observed many paralogous proteins positioned along the inner-outer kinetochore axis (Fig. 6, dashed line). We speculate that not too long before LECA, the genes encoding the proteins and/or complexes along this axis duplicated in quick stepwise succession or in a single event (55–57), which would be consistent with the proposed syncytial nature of lineages that gave rise to LECA (58).

Rapid Sequence Evolution of Kinetochore Components. The LECA kinetochore contains protein domains that are unique to the kinetochore and thus, by definition, unique to eukaryotes (33% of LECA kinetochore protein domains). New and more diverse genomes or elucidated protein structures may allow for the detection of their distant homologs in the future. Kinetochore proteins that share domains with other eukaryotic systems, such as the RWD, TBP-like, histone, and TPR domains, seem to be strongly diverged in the kinetochore. For example, the TPR domains of Mps1 and MadBub are more derived than those of the APC/C. This suggests that after these domains became involved in the kinetochore, their sequences evolved more rapidly and then continued to do so after LECA (10). Rapid evolution after LECA may be correlated with the widespread rapid divergence of centromere sequences. An evolutionary acceleration also may have occurred in the evolutionarily novel proteins in the LECA kinetochore, possibly explaining our failure to detect homology for some of these.

Possible Origins of the Kinetochore During Eukaryogenesis. Tracing the order in which proteins or domains became involved in the kinetochore relative to the origin of other eukaryotic features would be highly interesting. Possibly, an early, very basic kinetochore was composed simply of the centromere- and microtubule-binding proteins, similar to prokaryotic systems, while the CCAN (the “Cenp” proteins), which serves as their bridge, was added later. The relative timings of such contributions could potentially shed light on the evolution of eukaryotic chromosome segregation. Although little is known about the evolution of the eukaryotic segregation machinery, it must be associated with the evolution of linear chromosomes, the nucleus, and the eukaryotic cytoskeleton, including centrosomes.

Because the kinetochore shares ancestry with many other eukaryotic processes and cellular features and does not seem to have an explicit prokaryotic or eukaryote template (Fig. 6), we envision that it originated late during eukaryogenesis, for several reasons. First, the strong evolutionary link with flagellar transport systems (Fig. 6) may signify an early role for the flagellum in coordinating microtubule-based chromosome segregation, which is consistent with the function of the centriole as the microtubule-organizing center in most eukaryotes. Second, a large number of homologs related to vesicular transport components that function in the Golgi and ER point to membrane-based mechanisms of chromosome segregation in pre-LECA lineages, similar to those found among prokaryotes (Fig. 1A). Third, the prokaryotic roots of the HORMA proteins Mad2 and p31^{comet} and the AAA+ ATPase Trip13 suggest the (partial) incorporation of prokaryotic nucleotide sensing systems for setting up spindle checkpoint signaling. Finally, shared ancestries with complexes involved in transcription (Mediator and TFIID) and DNA replication/repair (Fanconi anemia pathway) suggest that kinetochores may be par-

tially descendant from systems involved in the control of transposons and/or repeated genomic regions, such as centromeres.

Because currently no eukaryotes or proto-eukaryotes are known that might segregate chromosomes in a pre-LECA manner, unravelling the series of events that gave rise to the spindle apparatus, the centromere, and the kinetochore remains difficult. The genomes of the currently known closest archaeal relatives of eukaryotes, the Asgard Archaea (59, 60) (Fig. 1A), clearly do not encode a eukaryote-like chromosome segregation system, but yet unidentified more closely related prokaryotes or proto-eukaryotes could do so. New (meta)genomic sequences have aided reconstruction of the evolution of the ubiquitin system (61) and the membrane trafficking system (54). Similarly, such newly identified species may enhance our understanding of the evolution of the eukaryotic kinetochore and chromosome segregation machinery.

Methods

Detailed descriptions of the methodology and data for this study are provided in *SI Appendix, Data and Methods*.

Profile-Versus-Profile Searches. Full-length and domain-specific hidden Markov model (HMM) profiles of kinetochore proteins were constructed using the hmmer package (version HMMER 3.1b1) (62), based on multiple sequence alignments [MSA; MAFFT, v.7.149b (63) “einsi” or “linsi”] of previously established orthologs (*SI Appendix, External Data: Hidden Markov Models*) (10, 19). Kinetochore profiles were searched against PANTHER11.1 profiles (64), using PRC (version 1.5.6) (65), and compiled domain profiles consisting of scop70 (March 1, 2016), pdb70 (September 14, 2016) and PfamA version 31.0, downloaded from the HH-suite depository (http://wwwuser.gwdg.de/~combiol/data/hhsuite/databases/hhsuite_dbs/), downloaded on July 15, 2017, using the secondary structure-guided HHsearch algorithm, version 2.0.15 (66). Raw data are provided in *Dataset S1*. The (bidirectional) best hits (E-value cutoff 1 or 10) of domain profile searches (HHsearch) were clustered and visualized using Cytoscape version 3.5.1 (67).

Phylogenetic Trees. Eukaryotic homologs were collected by searching with tailor-made and Pfam HMM profiles against our local proteome database (*SI Appendix, Table S3*) (10). For prokaryotic sequences, we performed online jackhmmer (<https://www.ebi.ac.uk/Tools/hmmer/>) (68) searches against the UniProt database. MSAs were inferred using MAFFT v.7.149b (63) and processed with trimAl (1.2rev59, various options) (69). For highly divergent protein families, we constructed a superalignment of trusted trimmed orthologous groups using the “merge” function of MAFFT (ginsi, unalignlevel 0.6). We scrutinized the resulting MSAs based on structure-based alignments (*SI Appendix, Data and Methods*). Trees were made using RAxML version 8.0.20 (automatic substitution model selection, GAMMA model of rate heterogeneity, rapid bootstrap analysis of 100 replicates) (70) and/or IQ-TREE version 1.6.3 [extended model selection, ultrafast bootstrap (1,000) and SH-like approximate likelihood ratio test] (71), and visualized and annotated using FigTree (72).

Structural Similarity. To identify homologs based on structural similarity with LECA kinetochore proteins, we searched both the literature and such databases as Pfam (<http://pfam.xfam.org>) (73), ECOD (<http://prodata.swmed.edu/ecod/>) (74), RCSB Protein Data Bank (<https://www.rcsb.org>) (75), and CATH (<http://www.cathdb.info>) (76). All-versus-all structural similarity z-scores (*Dataset S2*) were derived using the DALI webserver (77).

ACKNOWLEDGMENTS. We thank Leny van Wijk for providing the phylogenetic tree of eukaryotic kinases and helping to construct the eukaryotic proteome database, for which we also thank John van Dam. We also thank Stephen Hinshaw for sharing the .pdb file of the Ctf19/CCAN complex ahead of publication. We are indebted to the members of the G.J.P.L.K. and B.S. labs for helpful discussions on the research. Finally, we thank Bungo Akiyoshi for lively discussions on the origin of the kinetochore and the nature of LECA. This work was supported by the Netherlands Organisation for Scientific Research (NWO-Vici 016.160.638, to B.S.). E.C.T. is supported by a postdoctoral fellowship from the Herchel Smith Fund of the University of Cambridge.

1. M. Makarova, S. Olierenko, Mixing and matching nuclear envelope remodeling and spindle assembly strategies in the evolution of mitosis. *Curr. Opin. Cell Biol.* **41**, 43–50 (2016).
2. C. P. C. De Souza, S. A. Osmani, Mitosis, not just open or closed. *Eukaryot. Cell* **6**, 1521–1527 (2007).
3. H. Drechsler, A. D. McAinsh, Exotic mitotic mechanisms. *Open Biol.* **2**, 120140 (2012).
4. S. Sazer, M. Lynch, D. Needleman, Deciphering the evolutionary history of open and closed mitosis. *Curr. Biol.* **24**, R1099–R1103 (2014).
5. S. Henikoff, K. Ahmad, H. S. Malik, The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
6. I. A. Drinnenberg, S. Henikoff, H. S. Malik, Evolutionary turnover of kinetochore proteins: A ship of theseus? *Trends Cell Biol.* **26**, 498–510 (2016).

7. B. Akiyoshi, K. Gull, Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156**, 1247–1258 (2014).
8. S. D'Archivio, B. Wickstead, Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J. Cell Biol.* **216**, 379–391 (2017).
9. I. A. Drinnenberg, B. Akiyoshi, Evolutionary lessons from species with unique kinetochores. *Prog. Mol. Subcell. Biol.* **56**, 111–138 (2017).
10. J. J. van Hooff, E. Tromer, L. M. van Wijk, B. Snel, G. J. Kops, Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep.* **18**, 1559–1571 (2017).
11. D. Barillà, Driving apart and segregating genomes in archaea. *Trends Microbiol.* **24**, 957–967 (2016).
12. A. Badrinarayanan, T. B. K. Le, M. T. Laub, Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.* **31**, 171–199 (2015).
13. A.-C. Lindås, R. Bernander, The cell cycle of archaea. *Nat. Rev. Microbiol.* **11**, 627–638 (2013).
14. J. B. Dacks *et al.*, The changing view of eukaryogenesis: Fossils, cells, lineages and how they all come together. *J. Cell Sci.* **129**, 3695–3703 (2016).
15. E. V. Koonin, The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* **11**, 209 (2010).
16. J. Vosseberg, B. Snel, Domestication of self-splicing introns during eukaryogenesis: The rise of the complex spliceosomal machinery. *Biol. Direct* **12**, 30 (2017).
17. M. C. Field, J. B. Dacks, First and last ancestors: Reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr. Opin. Cell Biol.* **21**, 4–13 (2009).
18. B. J. Mans, V. Anantharaman, L. Aravind, E. V. Koonin, Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* **3**, 1612–1637 (2004).
19. R. Plowman *et al.*, The molecular basis of monopolin recruitment to the kinetochore. *Chromosoma*, 10.1007/s00412-019-00700-0 (2019).
20. F. Schmitzberger, S. C. Harrison, RWD domain: A recurring module in kinetochore architecture shown by a Ctf19-Mcm21 complex structure. *EMBO Rep.* **13**, 216–222 (2012).
21. T. Doerks, R. R. Copley, J. Schultz, C. P. Ponting, P. Bork, Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* **12**, 47–56 (2002).
22. A. M. Burroughs, M. Jaffee, L. M. Iyer, L. Aravind, Anatomy of the E2 ligase fold: Implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J. Struct. Biol.* **162**, 205–218 (2008).
23. A. Petrovic *et al.*, Modular assembly of RWD domains on the Mis12 complex underlies outer kinetochore organization. *Mol. Cell* **53**, 591–605 (2014).
24. F. Mattioli *et al.*, Structure of histone-based chromatin in Archaea. *Science* **357**, 609–612 (2017).
25. H. S. Malik, S. Henikoff, Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10**, 882–891 (2003).
26. Q. Zhao *et al.*, The MHF complex senses branched DNA by binding a pair of crossover DNA duplexes. *Nat. Commun.* **5**, 2987 (2014).
27. Y. Tao *et al.*, The structure of the FANCM-MHF complex reveals physical features for functional assembly. *Nat. Commun.* **3**, 782 (2012).
28. S. Pentakota *et al.*, Decoding the centromeric nucleosome through CENP-N. *eLife* **6**, e33442 (2017).
29. S. Chittori *et al.*, Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. *Science* **359**, 339–343 (2018).
30. S. M. Hinshaw, S. C. Harrison, An Im13-Chl4 heterodimer links the core centromere to factors required for accurate chromosome segregation. *Cell Rep.* **5**, 29–36 (2013).
31. B. Brindefalk *et al.*, Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res.* **41**, 2832–2845 (2013).
32. M. J. E. Koster, B. Snel, H. T. M. Timmers, Genesis of chromatin and transcription dynamics in the origin of species. *Cell* **161**, 724–736 (2015).
33. S. M. Hinshaw, S. C. Harrison, The structure of the Ctf19/CCAN from budding yeast. *eLife* **8**, e44239 (2019).
34. A. M. Burroughs, D. Zhang, D. E. Schäffer, L. M. Iyer, L. Aravind, Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015).
35. M. Gimona, K. Djinovic-Carugo, W. J. Kranevitter, S. J. Winder, Functional plasticity of CH domains. *FEBS Lett.* **513**, 98–106 (2002).
36. K. B. Schou, J. S. Andersen, L. B. Pedersen, A divergent calponin homology (NN-CH) domain defines a novel family: Implications for evolution of ciliary IFT complex B proteins. *Bioinformatics* **30**, 899–902 (2014).
37. R. C. Pasek, N. F. Berbari, W. R. Lewis, R. A. Kesterson, B. K. Yoder, Mammalian Clusterin-associated protein 1 is an evolutionarily conserved protein required for ciliogenesis. *Cilia* **1**, 20 (2012).
38. A. Pérez-González *et al.*, hCLE/C14orf166 associates with DDX1-HSPC117-FAM98B in a novel transcription-dependent shuttling RNA-transporting complex. *PLoS One* **9**, e90957 (2014).
39. M. D. Healy *et al.*, Structural insights into the architecture and membrane interactions of the conserved COMMD proteins. *eLife* **7**, e35898 (2018).
40. A. L. Mallam, E. M. Marcotte, Systems-wide studies uncover commander, a multi-protein complex essential to human development. *Cell Syst.* **4**, 483–494 (2017).
41. W. Nijenhuis *et al.*, A TPR domain-containing N-terminal module of MPS1 is required for its kinetochore localization by Aurora B. *J. Cell Biol.* **201**, 217–231 (2013).
42. T. Schlegel, O. Mirus, A. von Haeseler, E. Schleiff, The tetratricopeptide repeats of receptors involved in protein translocation across membranes. *Mol. Biol. Evol.* **24**, 2763–2774 (2007).
43. V. L. Koumandou, J. B. Dacks, R. M. R. Coulson, M. C. Field, Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evol. Biol.* **7**, 29 (2007).
44. W. Hong, S. Lev, Tethering the assembly of SNARE complexes. *Trends Cell Biol.* **24**, 35–43 (2014).
45. S. Schroeter, S. Beckmann, H. D. Schmitt, Coat/tether interactions—Exception or rule? *Front. Cell Dev. Biol.* **4**, 44 (2016). Correction in: *Front. Cell Dev. Biol.* **4**, 90 (2016).
46. C. M. Pfleger, E. Lee, M. W. Kirschner, Substrate recognition by the Cdc20 and Cdh1 components of the anaphase-promoting complex. *Genes Dev.* **15**, 2396–2407 (2001).
47. R. Murphy, J. L. Watkins, S. R. Wente, GLE2, a *Saccharomyces cerevisiae* homologue of the *Schizosaccharomyces pombe* export factor RAE1, is required for nuclear pore complex structure and function. *Mol. Biol. Cell* **7**, 1921–1937 (1996).
48. X. J. Hu *et al.*, Prokaryotic and highly-repetitive WD40 proteins: A systematic study. *Sci. Rep.* **7**, 10585 (2017).
49. K. S. Makarova, Y. I. Wolf, S. L. Mekhedov, B. G. Mirkin, E. V. Koonin, Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626–4638 (2005).
50. T. J. P. van Dam *et al.*, Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6943–6948 (2013).
51. A. Schlacht, J. B. Dacks, Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. *Genome Biol. Evol.* **7**, 1098–1109 (2015).
52. F. D. Mast, L. D. Barlow, R. A. Rachubinski, J. B. Dacks, Evolutionary mechanisms for establishing eukaryotic cellular complexity. *Trends Cell Biol.* **24**, 435–442 (2014).
53. T. Gabaldón, D. Rainey, M. A. Huynen, Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–870 (2005).
54. C. M. Klinger, A. Spang, J. B. Dacks, T. J. G. Ettema, Tracing the archaeal origins of eukaryotic membrane-trafficking system building blocks. *Mol. Biol. Evol.* **33**, 1528–1541 (2016).
55. J. B. Pereira-Leal, E. D. Levy, C. Kamp, S. A. Teichmann, Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* **8**, R51 (2007).
56. J. B. Dacks, A. A. Peden, M. C. Field, Evolution of specificity in the eukaryotic endomembrane system. *Int. J. Biochem. Cell Biol.* **41**, 330–340 (2009).
57. J. B. Dacks, M. C. Field, Evolutionary origins and specialisation of membrane transport. *Curr. Opin. Cell Biol.* **53**, 70–76 (2018).
58. S. G. Garg, W. F. Martin, Mitochondria, the cell cycle, and the origin of sex via a syncytial eukaryote common ancestor. *Genome Biol. Evol.* **8**, 1950–1970 (2016).
59. K. Zaremba-Niedzwiedzka *et al.*, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
60. A. Spang *et al.*, Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
61. X. Grau-Bové, A. Sebé-Pedrós, I. Ruiz-Trillo, The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. *Mol. Biol. Evol.* **32**, 726–739 (2015).
62. S. R. Eddy Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
63. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
64. H. Mi *et al.*, PANTHER version 11: Expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
65. M. Madera, Profile comparer: A program for scoring and aligning profile hidden Markov models. *Bioinformatics* **24**, 2630–2631 (2008).
66. J. Söding, Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
67. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
68. R. D. Finn, J. Clements, S. R. Eddy, HHMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
69. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
70. A. Stamatakis RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
71. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
72. A. Rambaut, FigTree v1. 4. Molecular evolution, phylogenetics, and epidemiology (2012). <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 4 May 2019.
73. R. D. Finn *et al.*, Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
74. H. Cheng *et al.*, ECOD: An evolutionary classification of protein domains. *PLoS Comput. Biol.* **10**, e1003926 (2014).
75. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
76. N. L. Dawson *et al.*, CATH: An expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* **45**, D289–D295 (2017).
77. L. Holm, L. M. Laakso, Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).