

On the possibility of yet a third kinetochore system in the protist phylum Euglenozoa

Corinna Benz,¹ Maximilian W. D. Raas,^{2,3} Pragma Tripathi,¹ Drahomíra Faktorová,^{1,4} Eelco C. Tromer,⁵ Bungo Akiyoshi,⁶ Julius Lukeš^{1,4}

AUTHOR AFFILIATIONS See affiliation list on p. 23.

ABSTRACT Transmission of genetic material from one generation to the next is a fundamental feature of all living cells. In eukaryotes, a macromolecular complex called the kinetochore plays crucial roles during chromosome segregation by linking chromosomes to spindle microtubules. Little is known about this process in evolutionarily diverse protists. Within the supergroup Discoba, Euglenozoa forms a speciose group of unicellular flagellates—kinetoplastids, euglenids, and diplomemids. Kinetoplastids have an unconventional kinetochore system, while euglenids have subunits that are conserved among most eukaryotes. For diplomemids, a group of extremely diverse and abundant marine flagellates, it remains unclear what kind of kinetochores are present. Here, we employed deep homology detection protocols using profile-versus-profile Hidden Markov Model searches and AlphaFold-based structural comparisons to detect homologies that might have been previously missed. Interestingly, we still could not detect orthologs for most of the kinetoplastid or canonical kinetochore subunits with few exceptions including a putative centromere-specific histone H3 variant (cenH3/CENP-A), the spindle checkpoint protein Mad2, the chromosomal passenger complex members Aurora and INCENP, and broadly conserved proteins like CLK kinase and the meiotic synaptonemal complex proteins SYCP2/3 that also function at kinetoplastid kinetochores. We examined the localization of five candidate kinetochore-associated proteins in the model diplomemid, *Paradiplonema papillatum*. PpCENP-A shows discrete dots in the nucleus, implying that it is likely a kinetochore component. PpMad2, PpCLK^{KKT10/19}, PpSYCP2L1^{KKT17/18}, and PpINCENP reside in the nucleus, but no clear kinetochore localization was observed. Altogether, these results point to the possibility that diplomemids evolved a hitherto unknown type of kinetochore system.

IMPORTANCE A macromolecular assembly called the kinetochore is essential for the segregation of genetic material during eukaryotic cell division. Therefore, characterization of kinetochores across species is essential for understanding the mechanisms involved in this key process across the eukaryotic tree of life. In particular, little is known about kinetochores in divergent protists such as Euglenozoa, a group of unicellular flagellates that includes kinetoplastids, euglenids, and diplomemids, the latter being a highly diverse and abundant component of marine plankton. While kinetoplastids have an unconventional kinetochore system and euglenids have a canonical one similar to traditional model eukaryotes, preliminary searches detected neither unconventional nor canonical kinetochore components in diplomemids. Here, we employed state-of-the-art deep homology detection protocols but still could not detect orthologs for the bulk of kinetoplastid-specific nor canonical kinetochore proteins in diplomemids except for a putative centromere-specific histone H3 variant. Our results suggest that diplomemids evolved kinetochores that do not resemble previously known ones.

KEYWORDS *Paradiplonema*, cell division, kinetochore, cenH3/CENP-A, Diplonemea, Kinetoplastea

Editor L. David Sibley, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA

Address correspondence to Eelco C. Tromer, e.c.tromer@rug.nl, Bungo Akiyoshi, bungo.akiyoshi@ed.ac.uk, or Julius Lukeš, jula@paru.cas.cz.

Corinna Benz, Maximilian W. D. Raas, and Pragma Tripathi contributed equally to this article. Author order was determined alphabetically.

The authors declare no conflict of interest.

Received 3 October 2024

Accepted 9 October 2024

Published 30 October 2024

Copyright © 2024 Benz et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

A fundamental characteristic of living organisms is the ability to self-replicate. This ability requires accurate transmission of genetic information from one generation to the next during cell division (1, 2). In eukaryotes, a sophisticated macromolecular structure called the kinetochore (KT) assembles on centromeric chromatin and binds spindle microtubules to control chromosome movement during mitosis and meiosis (3). In addition, kinetochores are responsible for regulating a surveillance mechanism called the spindle assembly checkpoint (SAC), which delays cell cycle progression until all chromosomes have achieved proper kinetochore-microtubule attachment (4). The molecular mechanisms underlying these processes have been extensively investigated in traditional model eukaryotes, such as yeast, worm, fly, and human. Importantly, all of these organisms belong to the supergroup Opisthokonta, meaning that they are closely related in the timescale of eukaryotic evolution (Fig. 1A) (5). With the exceptions of plants, apicomplexans, and kinetoplastids, little is known about molecular underpinnings of mitotic mechanisms in members of other eukaryotic supergroups (6).

Studies in opisthokonts have shown that the kinetochore is a highly complex structure (more than 100 proteins in humans), which can be divided into two functional modules: the inner kinetochore that is built on centromeres and the outer kinetochore that binds microtubules and regulates the SAC (8, 9) (Fig. 1B). A key inner kinetochore component is CENP-A, a centromere-specific histone H3 variant that determines the position of kinetochores and initiates kinetochore assembly, while the outer kinetochore includes the Ndc80 complex that binds microtubules. The fact that CENP-A, its interaction partner CENP-C, the Ndc80 complex, and SAC proteins such as Mad2 and Cdc20 are widely conserved suggests that most eukaryotes use these proteins to perform basic kinetochore functions. However, phylogenetic profiling surveys across eukaryotes revealed remarkable plasticity of kinetochore composition (6, 10, 11). Strikingly, reconstructions based on these disparate presence/absence profiles of kinetochore proteins still suggest that the last eukaryotic common ancestor (LECA) possessed a highly complex ancestral kinetochore system on par with those of extant eukaryotes such as human and yeast (Fig. 1B) (7, 12). Such intuitively opposing patterns of ancestral complexity and contemporary plasticity imply that the kinetochores must have had a highly dynamic evolutionary history, despite their cardinal role in chromosome segregation.

Recent experimental studies started to probe kinetochore variations among eukaryotes (13). For example, several insects (e.g., butterflies and moths) and early diverging fungi (e.g., *Mucor circinelloides*) lack CENP-A, although these organisms have retained the Ndc80 complex (14–16). Furthermore, an increasing number of non-traditional model eukaryotes are being studied for their seemingly divergent kinetochore composition and function. An extensive bioinformatic survey could not detect most of the conventional kinetochore (cKT) in the free-living metamonad flagellate *Carpodidomonas membranifera* (17). On the other hand, an evolutionary cell biology approach in the apicomplexan parasites *Plasmodium* and *Toxoplasma* revealed the presence of an extremely divergent, yet conventional kinetochore makeup (18, 19).

The most extreme case of kinetochore divergence known to date is found in Kinetoplastea (Euglenozoa, Discoba) for which bioinformatic analysis failed to identify obvious orthologs of CENP-A, Ndc80, or any other canonical kinetochore proteins (20, 21). A localization-based screen in the model kinetoplastid parasite *Trypanosoma brucei* identified kinetoplastid kinetochore protein 1 (KKT1), and subsequent immunoprecipitation and mass spectrometry analyses identified many additional kinetochore proteins including KKT1–25 and KKIP1–12 (Fig. 1C) (22–26). Many of these proteins are conserved among kinetoplastids, including free-living bodonids and prokinetoplastids (Fig. 1A) (20, 27–30). However, clear homologs for most KKT/KKIPs are apparently absent in other eukaryotes, except for proteins with either generic domains for which pinpointing their evolutionary history is technically challenging (e.g. PH and FHA domains in KKT proteins and RRM domains in many KKIP proteins) or for proteins such as KKT16/17/18 and KKT10/19 that play key roles in meiotic chromosome synapsis and splicing in other

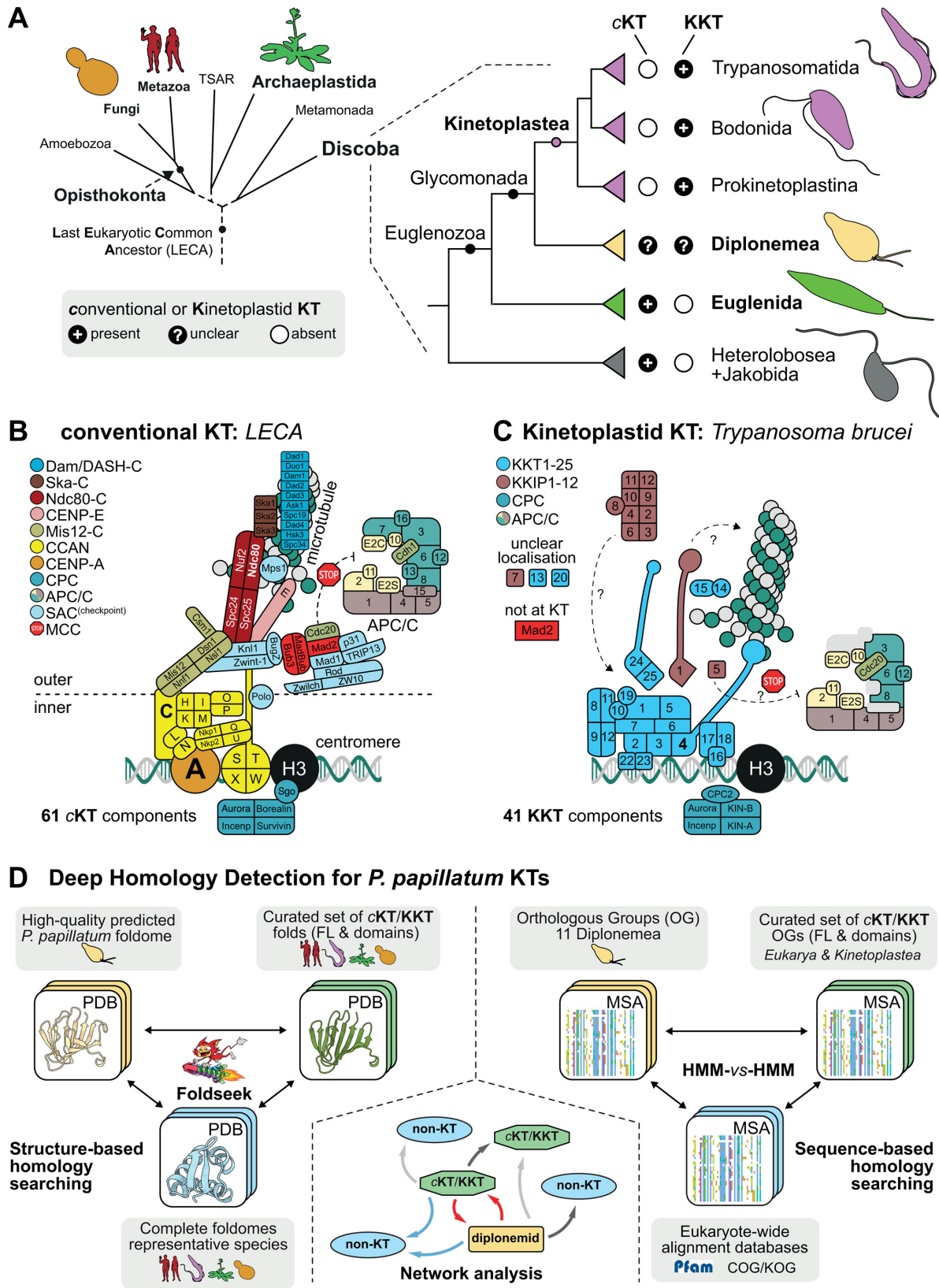


FIG 1 Kinetochore composition in Euglenozoa. (A) Cartoon of phyletic relationships of the superphylum Discoba with emphasis on the phylum Euglenozoa harboring kinetoplastids, euglenids, and diplomemids. (B) Cartoon of 61 components of the *conventional* KT (cKT), SAC, and the anaphase promoting complex/cyclosome (APC/C) as inferred to have been present in the ancestral eukaryotes (7). (C) Cartoon of 41 components and structure of the kinetoplastid (Continued on next page)

Fig 1 (Continued)

kinetochore (KKT) in *Trypanosoma brucei*, also including the SAC and APC/C components. (D) Overview of the twofold strategy taken toward highly sensitive identification of cKT/KKT orthologs in the *Paradiplonema papillatum* proteome. Schematic overviews of the reciprocal structure-based (left) and sequence-based homology searching strategies (right). In the center, an example of a network graph is shown that is produced based on these comprehensive searches, with green octagonal nodes indicating a cKT/KKT model, a rectangular yellow node, a *P. papillatum* accession or diplomemid orthologous group (OG), and in circular blue nodes, any non-KT protein from the complete foldomes of any of the representative species (*Homo sapiens*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Trypanosoma brucei*). Arrows (edges) between nodes represent homology links colored according to hierarchy in the search results based on E-value (red: best, blue: second-best, dark gray: third-best, and light gray: other hit).

eukaryotes, respectively. Although it has been proposed that the ancestor of kinetoplastids repurposed parts of the meiotic synapsis and homologous recombination machinery to assemble unique kinetochores (31), it remains unclear why and when kinetoplastids invented their unique kinetochore system.

Kinetoplastids belong to the phylum Euglenozoa, which also includes euglenids and diplomemids (Fig. 1A) (32, 33). Phylogenetic studies indicate that kinetoplastids are more closely related to diplomemids than to euglenids (29, 34). Diplomemids are a group of heterotrophic flagellates that are not only highly abundant in marine environments (35–37), but also extremely diverse, with 18S rRNA-based estimates reaching ~70,000 species (38). Negligence of these omnipresent eukaryotes is reflected by the fact that the number of diplomemid species in culture for which morphological and sequence data is available still does not exceed a dozen (37, 39). Hence, despite an inevitable ecological importance of these highly abundant protists in the oceans, very little is known about their biology. Interestingly, unlike euglenids that have clear orthologs of canonical kinetochore proteins, only a putative CENP-A candidate has been identified in diplomemids (29). Hence, the kinetochore composition in diplomemids remains unclear.

To address this knowledge gap and to gain insights into the evolutionary origin of kinetoplastid kinetochores (KKTs), we set out to investigate the possible protein composition of kinetochores in diplomemids. We started by applying state-of-the-art sensitive homology detection protocols to explore the possibility of previously missed hidden homologies (19, 40). Indeed, in the malaria parasite *Plasmodium berghei*, a comparative analysis using both profile Hidden Markov Model (HMMs) and AlphaFold2 (AF2)-predicted protein structures uncovered the presence of many *bona fide* canonical kinetochore proteins that were previously refractory to discovery even by powerful tools such as iterative HMM searches (19). To achieve the same goal, we have used AlphaFold2 to predict the foldome for the model diplomemid *Paradiplonema papillatum* (formerly known as *Diplonema papillatum*).

Recently, *P. papillatum* became the first diplomemid (Diplonemea) for which methods of integration of extraneous DNA have been developed (41), thus turning it into a new model marine protist (42). Indeed, insertion of any DNA segment into a chromosome *via* homologous recombination can be achieved by using ~1.5 kb-long homology arms (43). The application of tagging to identify components of membrane trafficking and the mitochondrial ribosome confirmed experimental tractability of *P. papillatum* (44, 45). With a publicly available high-quality genome (46) and a set of transcriptomes (47), this technique enables functional studies in this group of evolutionarily and ecologically highly relevant protists. In this study, we explore the evolutionary history of euglenozoan kinetochores and examine the localization of five candidate kinetochore-associated proteins in *P. papillatum*.

RESULTS

Deep homology detection strategies confirm the general absence of known kinetochore proteins among Diplonemea

Previously, it was suggested that Diplonemea may lack several crucial components of the cKT and also of the KKT, leaving these flagellates seemingly without a known type of kinetochore structure (29). Here, we built on these prior analyses by performing a

comprehensive and highly sensitive deep homology search for components of the cKT and KKT, focusing on the predicted proteome from *P. papillatum*. This workflow allows us to scrutinize several mutually non-exclusive evolutionary scenarios for the diplonemid kinetochore, which is either (i) a highly divergent version of the cKT, (ii) a highly divergent version of the KKT, (iii) a mix of both cKT/KKT, or (iv) a unique kinetochore of distinct evolutionary origin.

To systematically search the proteome of *P. papillatum* for highly divergent orthologs of cKT and KKT components, we used a twofold strategy employing both sequence-based as well as predicted protein structure-based similarity searches (Fig. 1D; Fig. S1 and S2). In the former, we performed exhaustive multilevel profile-versus-profile HMM searches specifically exploring “grey-zone”/borderline similarities leveraging sensitive, manually curated models of both full-length proteins as well as domains and motifs for 71 cKT and 39 KKT-related components (Tables S3 and S4, at FigShare). For structural comparisons (Table S5, at FigShare), we predicted the protein structures of the *P. papillatum* proteome with AF2 using multiple sequence alignments including 10 Diplonemea species for which transcriptomic data are available (Tables S1 and S2, at FigShare). We followed a similar strategy taken in the profile-versus-profile HMM searches, in this case, using 272 manually curated cKT/KKT protein structure AF2 models from four representative species: *Homo sapiens*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and *Trypanosoma brucei* (Fig. 1D; Fig. S1).

Putative homologs obtained by these searches were analyzed in a network graph-based framework using Markov Clustering (MCL) (Fig. 2 and Materials and Methods), with the underlying reasoning that orthologous sequences and folds tend to form distinct similarity clusters in the MCL network that can be used as a proxy for their underlying phylogeny (40, 48). In parallel, we applied the bidirectional best/better hit principle to uncover strong homology associations (Tables S3C and S5C, at FigShare). Using these complementary methods, our analysis readily revealed *bona fide* orthologs of the spindle checkpoint (p31^{COMET}, Cdc20, Trip13, Mad2), chromosomal passenger complex (INCENP), inner kinetochore histone-like proteins CENP-S and CENP-X, as well as the previously reported putative CENP-A ortholog (Fig. 2). This approach also allowed the identification of multiple close paralogs, and in some cases the closest paralog, to cKT/KKT proteins where multiple duplications of a particular gene family occurred among Euglenozoa (e.g., CLK kinase KKT10/19 and the phosphatase KKIP7) (29). Lastly, we recovered two orthologs of the SYCP2-like proteins KKT17 and KKT18, corroborating previous findings (31).

By contrast, we were unable to identify additional cKT/KKT orthologs, including the principal microtubule-binding components of the cKT (NDC80/NUF2) and the KKT (KKT4) (Fig. 3), as well as other hallmark features of the inner (CENP-C) and outer cKT (MIS12) (Fig. S3). Although we found homologs for these proteins by virtue of their conserved domains, e.g., the CENP-C cupin domain and the NDC80/NUF2 CH domain, these are more similar to closely related known outgroup proteins (Fig. 3A; Fig. S3A). Identification of CCDC93 as an outgroup for NDC80 is in line with the LECA kinetochore reconstruction (7, 49). For other manual explorations and checks, we refer to the specific notes on each cKT/KKT component (Tables S3 to S5, at FigShare). Finally, we found a strong similarity link (best hit: DIPPA_32591) for KKT4, the only known microtubule-binding protein of the kinetoplastid kinetochore. However, the BRCT domain in DIPPA_32591 is a bidirectional best hit with NUP92/MLP2, a nuclear pore component in *T. brucei* (50). Indeed, our phylogenetic analysis shows that KKT4 originated through duplication of NUP92/MLP2 in the ancestor of all kinetoplastids and that DIPPA_32591 belongs to the NUP92/MLP2 family rather than KKT4 (Fig. 3B; Fig. S4A).

Combined, our approaches robustly confirm the absence of a large number of components of both the cKT and the KKT in our diplonemid transcriptomic database and the high-quality genome of *P. papillatum* (46).

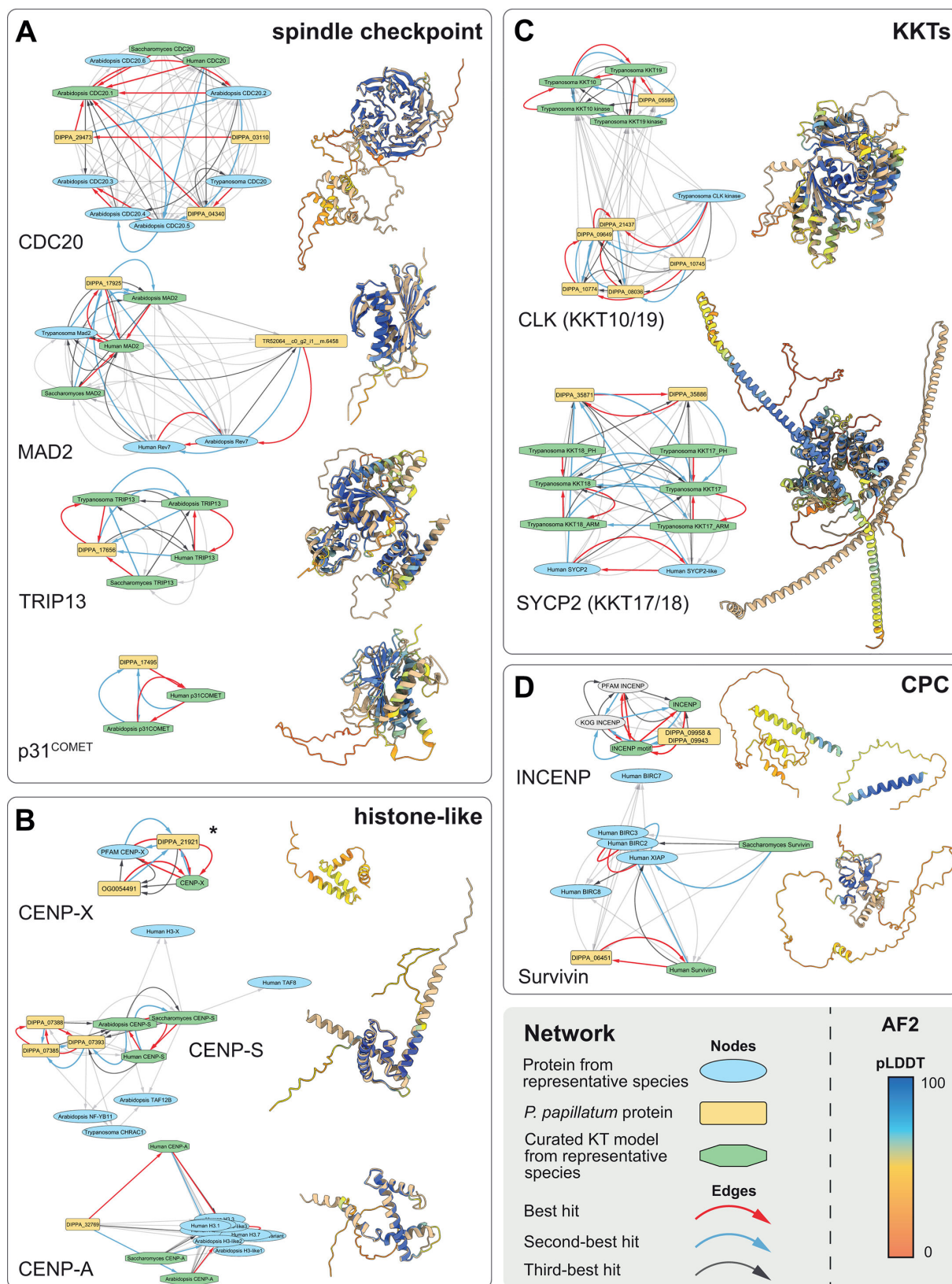


FIG 2 Detection of bona fide kinetochore proteins in *P. papillatum*. Network graphs showing similarity links (E-value) as inferred by Foldseek or HHsearch, displayed as edges between homologous folds depicted as nodes (left) and structural alignments as produced by Foldseek where available (right). Network graphs represent an individual MCL. Structural alignments show the predicted structure of the relevant *P. papillatum* protein placed in the cluster to the AF2 (Continued on next page)

Fig 2 (Continued)

structure that was best hit in the Foldseek against the foldomes of representative species, colored in tan. Predicted protein structures are colored by pLDDT score following the standard AF2 palette. Panels represent different functional/structural aspects of the cKT and KKT, as indicated. *MCL clusters obtained from a network based on HHsearch homology inferences, no structural alignment is available for these models as the folds do not contain a clear structural domain. The best-predicted (highest pLDDT score) AF2 model of the *P. papillatum* accession is shown on the right.

Parallel expansions of kinetochore-related genes in Euglenozoa

In some of our clustering analyses, it proved difficult to make a confident assessment of putative orthologs solely through inspection of the network, particularly in gene families with a large amount of recent duplications. Furthermore, only the top 10 hits were considered in the network analysis, possibly missing highly divergent orthologs. We therefore manually explored both sequence and structure-based search outputs to investigate whether our network analysis missed cKT/KKT orthologs.

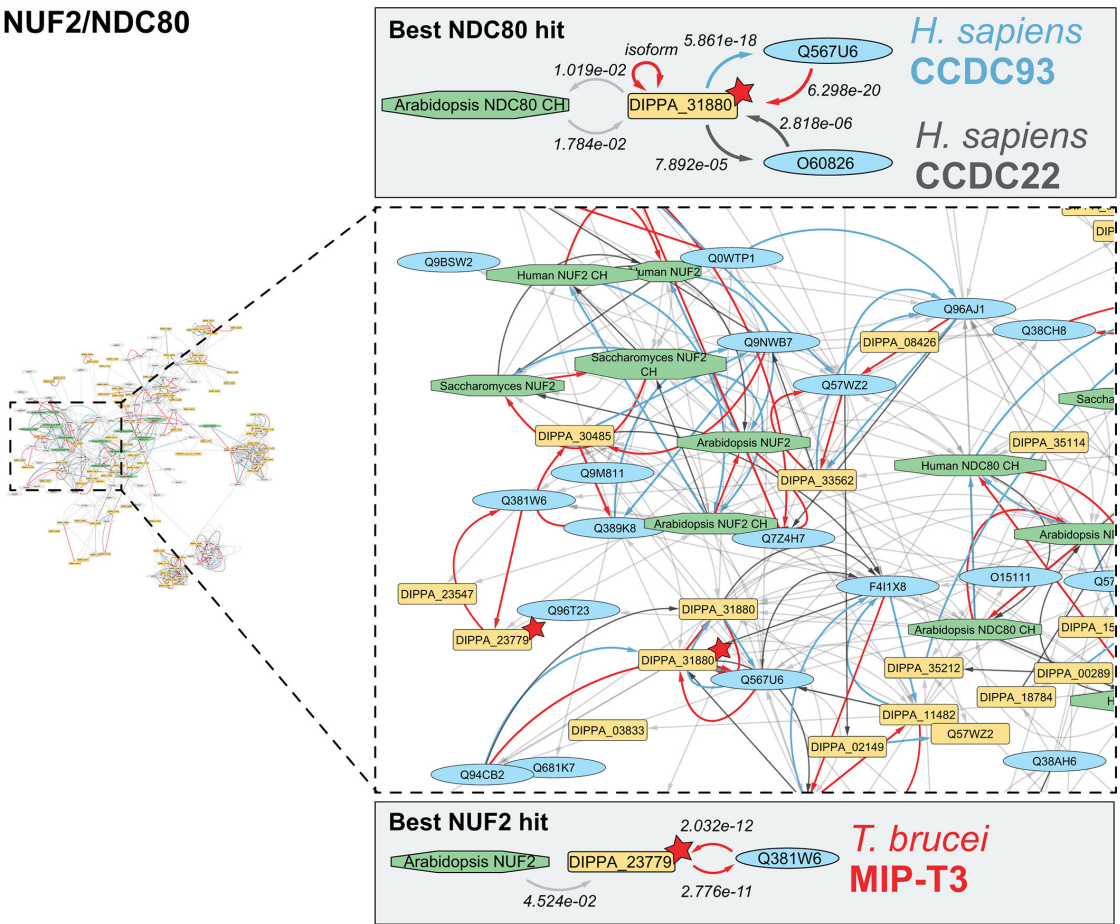
Through careful inspection of the data, we identified a *bona fide* ortholog of the cKT-associated kinesin CENP-E through a bidirectional best hit (DIPPA_27648), and the closely related KKT-associated kinesin KIN-A (DIPPA_28866 and DIPPA_16905), which appear to be part of one related family of kinesins (Fig. S4B; Table S3C, at FigShare). In addition, we found three BUB3/RAE1-like sequences (DIPPA_18051, DIPPA_04497, and DIPPA_14534), for which subsequent phylogenetic analysis showed that DIPPA_18051 and DIPPA_04497 are *bona fide* RAE1 orthologs, and that DIPPA_14534 branches off prior to the split of BUB3/RAE1, similar to the BUB3-like KKT15 protein (Fig. S5A). Based on this phylogeny, we speculate that DIPPA_14534 is most likely a BUB3 ortholog considering the presence of clear RAE1 orthologs in diplomids and it being similarly placed outside (but closer to BUB3/RAE1) of the RAE1/BUB3 group similarly to the BUB3-like protein KKT15 (51). For the phosphatase KPIP7, we found many paralogs in Euglenozoa which arose through a complex history of duplications, with DIPPA_23494 as the 1-to-1 KPIP7 ortholog (Fig. S5B). Lastly, we generated phylogenetic trees for the Aurora-Polo kinases families, which were previously found to be each other's closest paralogs in the LECA genome (52). We added the KKT2 and KKT3 kinases because these proteins also have a kinase and polo-box domain for which we found strong similarity with Polo kinases in our network analysis, similar to previous findings (23) (Fig. S6; Table S3, at FigShare). Diplonemid Aurora kinases appear to cluster close to those of kinetoplastids, suggesting a shared history of duplications, with a total of three Aurora kinases: Aurora-1 (Aurora B-like), Aurora-2, and Aurora-3 (Fig. S6). Remarkably, the kinase domains of KKT2 and KKT3 cluster at the base of the Polo kinase group closer to those of PLK4. Given that no PLK4 ortholog is present in Kinetoplastea, this might suggest that KKT2/3 are descended from PLK4. However, the higher similarity of the polo boxes of KKT2/3 with those of PLK1 and not PLK4 may well point to KKT2/3 being highly divergent Polo kinases not belonging to the PLK4 clade, and that the current position reflects extreme divergence causing long-branch attraction. We also found multiple additional parallel duplications among Polo kinases giving rise to distinct numbers of paralogs among various eukaryotic groups, including separate duplications in euglenids, diplomids, and kinetoplastids (Fig. S6).

Phylogenetic distributions of kinetochore proteins suggest distinct origins for the kinetochores in kinetoplastids, diplomids, and euglenids

We next profiled the presences and absences of cKT and KKT orthologs in an extended set of species (Table S1, at FigShare) belonging to the supergroup Discoba to infer the evolutionary timing of the losses, gains, and transitions that occurred among euglenozoans, and compare these to the kinetochore compositions of *H. sapiens*, *A. thaliana*, and *S. cerevisiae* (Fig. 4; Table S6, at FigShare).

We found that the majority of the cKT is conserved in a basal group of discobans, the Jakobida (53), including the complete CCAN, MIS12, NDC80, DAM1, and SKA complexes. The presence of the CCAN in other basal lineages among various eukaryotic supergroups (7, 54) further bolsters the notion that the cKT is an ancient structure that was already

A NUF2/NDC80



B KKT4

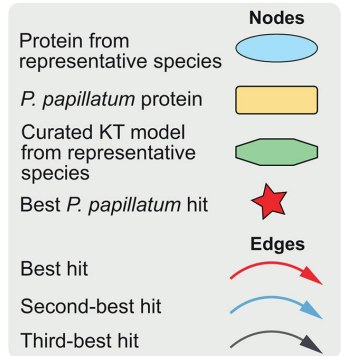
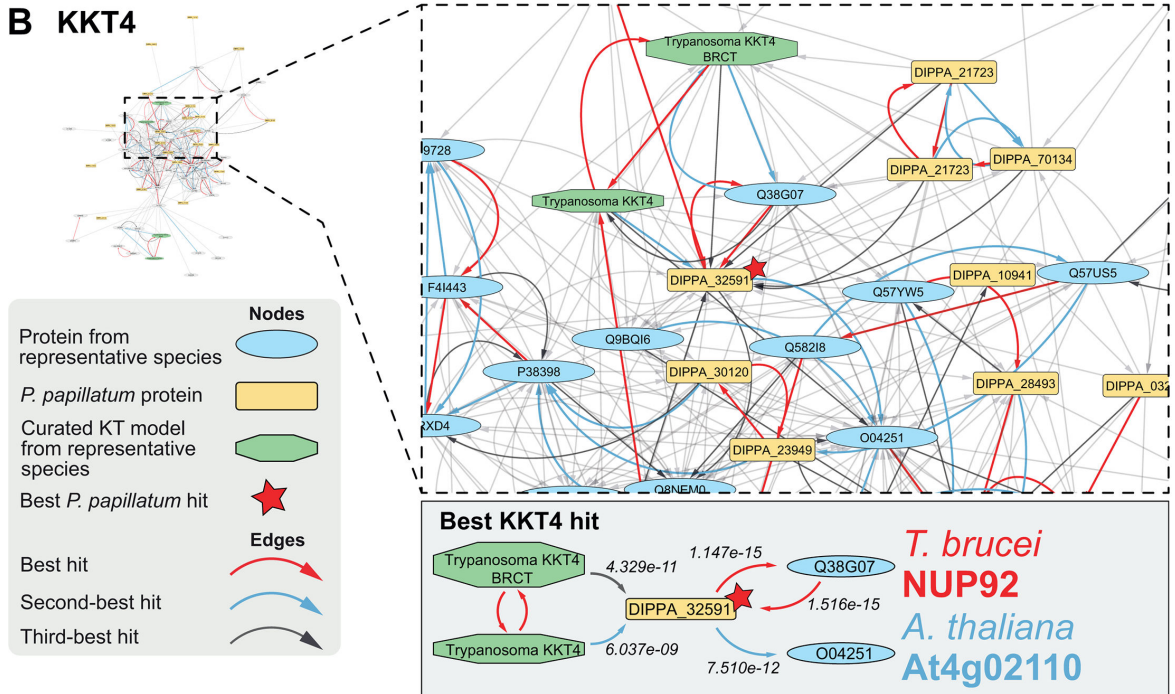


FIG 3 Closest homologs to NDC80/NUF2 and KKT4 are not 1-to-1 orthologs but outparalogs CCDC93/Cluap1 and NUP92, respectively. Network graphs showing homology hits as inferred by Foldseek, displayed as edges between homologous folds depicted as nodes. (A) All nodes connected to NUF2 and NDC80 nodes (Continued on next page)

Fig 3 (Continued)

through either incoming or outgoing edges, and all nodes connected to these nodes with either directionality to the edge (i.e., all nodes until one degree of separation to a NUF2 and/or NDC80 node). Nodes are positioned according to the E-value represented by connecting edges, where shorter edges equate to lower E-value. On the left, a zoom-in of the network around the node of the *P. papillatum* accession that is the closest homolog by one of the NUF2/NDC80 query models (marked by a red star) and its closely related hits. Schematic representations of the path connecting NDC80 (top) and NUF2 (bottom) query nodes to their respective top *P. papillatum* accession hit and its subsequent best hits are shown. (B) Similar to panel A, but with nodes connected to KKT4 through maximally one degree of separation.

present in the LECA (7). The presence of both DAM1 and SKA complexes in Jakobida is particularly striking, as this dual presence was recently inferred to have been the case in the LECA (55). All in all, such a complex kinetochore system in a basal lineage of a supergroup highlights the notion that Discoba is one of the candidate lineages to be close(st) to the root of the eukaryotic tree of life (56). Our phylogenetic profiling shows a sequential loss of cKT components along the Discoba phylogeny, with first the bulk of the CCAN and the DAM1 complex being lost in Euglenozoa and Heterolobosea, followed by the losses of CENP-C and MIS12 complex in Euglenozoa, and then finally the outer kinetochore (KMN and SKA complexes), components of SAC [e.g., Bub(R)1] and recurrent losses of CENP-A in Diplonemea and Kinetoplastea. This pattern may be indicative of a stepwise replacement of the cKT by a novel kinetochore-like system in Discoba.

With respect to the KKT system, we found some new but in general lower number of orthologs of KKT/KKIP components in the early branching prokinetoplastids and bodonids compared to the crown trypanosomatids. Clearly, these KKT/KKIP components cannot be found outside of kinetoplastids, apart from the aforementioned CLK^{KKT10/19}, SYCP2^{KKT17/18}, and KKIP7 in most Discoba, as well as candidates for KKIP4, KKIP10, and KKT13 (Fig. 4; Table S6, at FigShare). Prokinetoplastina takes an interesting intermediate position, as they appear to have some cKT components, similarly to those found in diplomemids, namely p31^{comet} and CENP-A. Diplonemids additionally have a relatively complete list of the SAC and anaphase promoting complex/cyclosome (APC/C) components, with only Bub1, Cdh1, APC12, and APC16 missing. Furthermore, presence of a putative APC15 ortholog suggests that diplomemids have active SAC silencing (57). Remarkably, CENP-S/X were found in diplomemids, but not in any other euglenozoan. Such isolated presence of CENP-S/X may be explained by their role in DNA damage signaling through the Fanconi anemia pathway (58). Strikingly, however, our survey revealed the general absence of Fanconi anemia pathway components in diplomemids, suggesting that CENP-S/X might have yet another role to play.

Endogenous tagging of putative kinetochore-related proteins in *P. papillatum*

The *in silico* analyses were followed by experiments in which we focused on five proteins. A putative CENP-A homolog (DIPPA_32769), Mad2 (DIPPA_17925), INCENP (DIPPA_09943), CLK^{KKT10/19} (DIPPA_05595), and SYCP2L1^{KKT17/18} (DIPPA_35871) were C-terminally tagged with a protein A tag (PrA) using the plasmid pDP002 that contains neomycin as a selectable marker (43). In addition, we created a cell line in which CENP-A was C-terminally tagged with V5 using a modified pDP011 plasmid (44). Following electroporation and drug selection, we assayed the expression of tagged proteins by immunoblotting using polyclonal antibodies against either protein A or the V5 tag, and typically selected three random clones for further analysis. Immunoblots of a representative clone for each tagged protein are shown (Fig. 5). Specific bands of the expected molecular weight were detected for each cell line, confirming the proper integration of the tags.

The CENP-A candidate forms discrete nuclear foci in *P. papillatum*

CENP-A is a centromere-specific histone H3 variant, often characterized by the longer loop 1 region that separates the helices 1 and 2 in the histone fold domain (59). Among

Fig 4 (Continued)

(cartoon left). Kinetoplastea have kinetochores consisting of KKT/KKIP subunits. Euglenida have kinetochores with the conserved Ndc80 complex, CENP-A, the SKA complex, and most likely an active spindle checkpoint (Bub1/BubR1, Mad2, Mad1, Bub3). Diplonemea harbor a CENP-A-like protein, only minimal numbers of KKT/KKIPs (CLK^{KKT10/19}, SYCP3^{KKT16}, SYCP2^{KKT17/18}, KKIP7, and KKIP10) and cKT proteins (p31^{COMET}, CENP-S/X, Mad1, Mad2). Overall, Euglenozoa appear to have lost and replaced the cKT in multiple steps in the different clades—see before. Ancestrally, Discoba would have had a complex kinetochore system with a full CCAN (Jakobida), Mis12-C, Ndc80-C, DAM-C, and SKA-C, including an active spindle checkpoint and APC/C. Top left: cartoon of conserved cKT/KKT-related proteins present in *P. papillatum* as projected onto the LECA cKT composition; gray indicates absent, colored proteins indicate present. Top right: Coulson-style plots of cKT/KKT complexes—below: colored parts of the pie chart are present, light gray means these proteins are absent. Numbers indicate the number of paralogs present. Colors correspond to the cartoons of the cKT and KKT shown in Fig. 1. Per major taxonomic groups among Discoba, a representative species is highlighted, including the presence of a cartoon of its cellular outline. Bottom: presence/absence of three model eukaryotes for which kinetochores have been determined: *H. sapiens*, *S. cerevisiae*, and *A. thaliana*.

the histone H3-like proteins in *P. papillatum*, DIPPA_32769 has many features of CENP-A, including the longer loop 1 and replacement of glutamine (Gln, Q) that is conserved in histone H3 but often changed to another amino acid in CENP-A homologs (Fig. 6A). In fact, our reciprocal structure-based homology search firmly placed DIPPA_32769 in a *bona fide* CENP-A cluster (Fig. 2B). Fluorescence and confocal microscopy analyses revealed discrete foci throughout the nucleus for both CENP-A-PrA and CENP-A-V5 (Fig. 6B and C). To further characterize this protein, we performed affinity purification mass spectrometry on cells expressing CENP-A-V5, using wild-type *P. papillatum* cells as a control (Fig. S7). Although relatively high background and lower amount of peptides precluded clear designation of strong *Pp*CENP-A interactors, we found enrichment of a

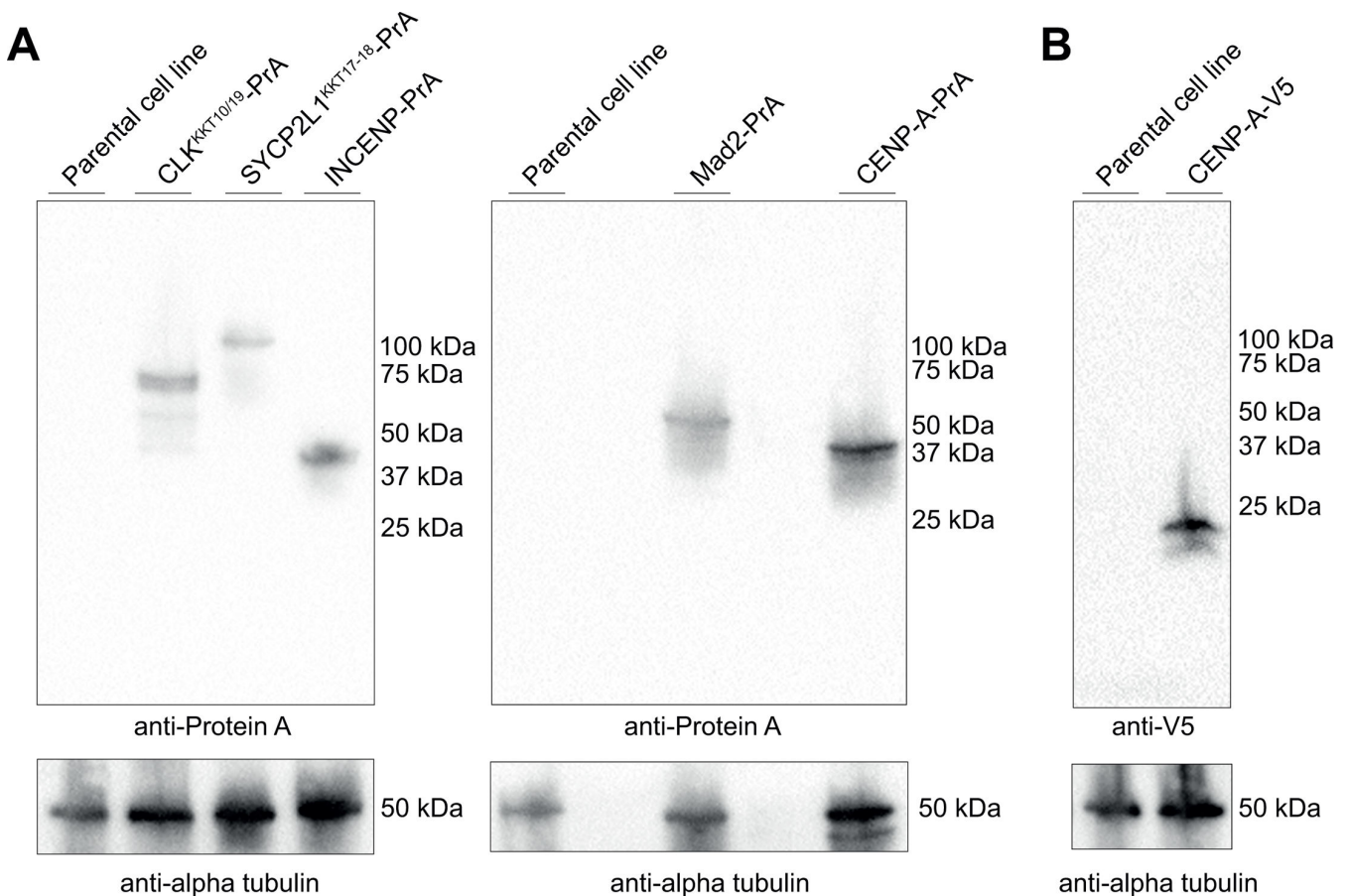
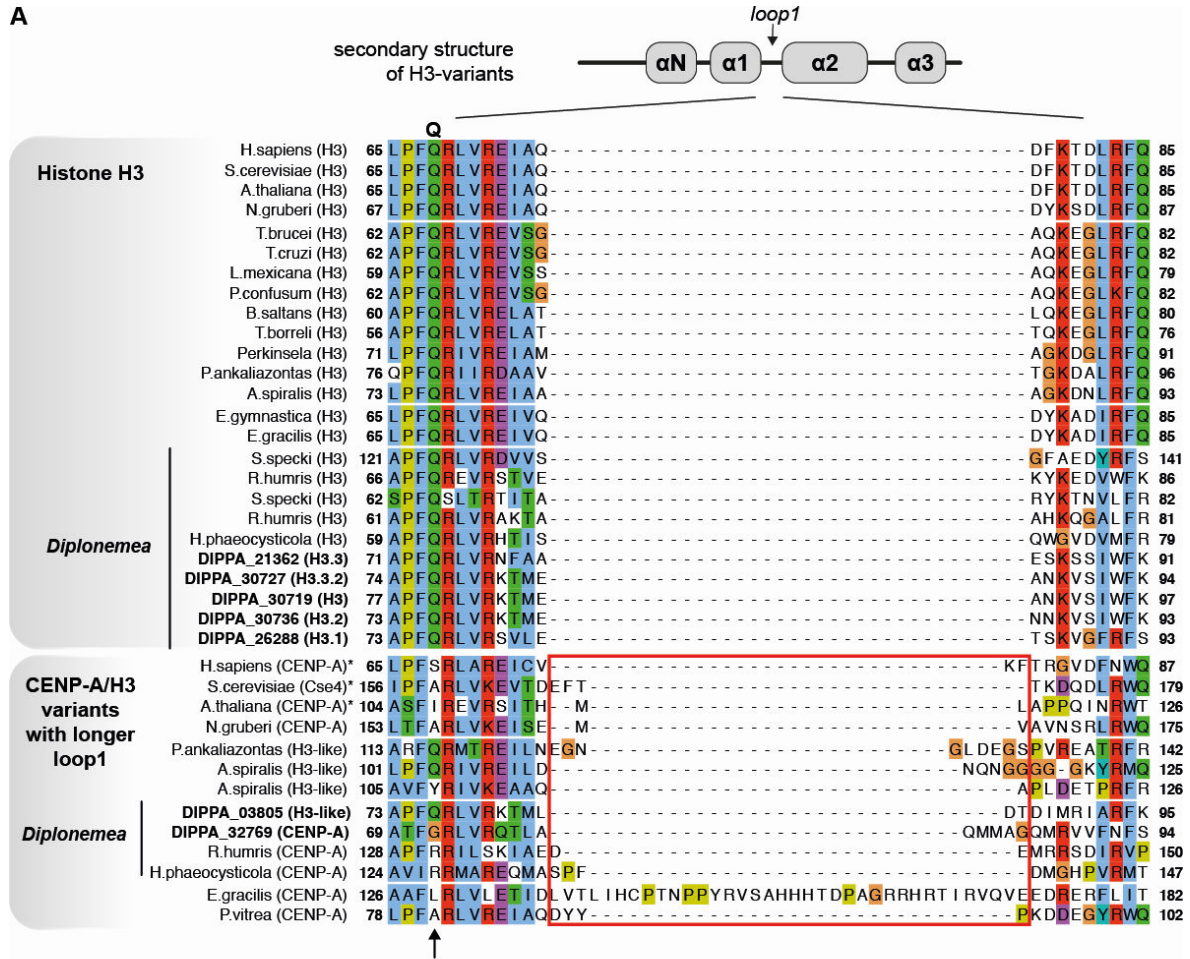
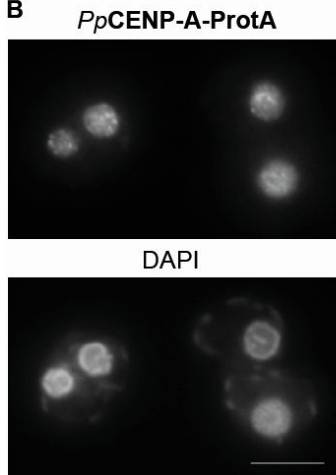


FIG 5 Endogenous tagging of putative mitotic proteins in *P. papillatum*. (A) Immunoblot analysis of *P. papillatum* cell lines expressing a PrA fusion of CENP-A, INCENP, Mad2, CLK^{KKT10/19}, and SYCP2L1^{KKT17/18}. Anti-alpha-tubulin was used as a loading control. (B) Western blot analysis of *P. papillatum* parental cell line and transgenic cell lines expressing CENP-A-V5.

A



B



C

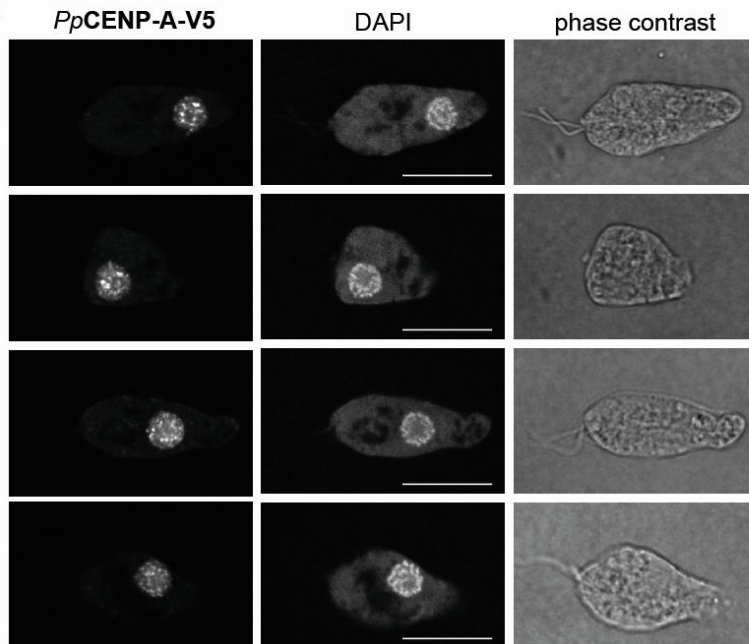


FIG 6 DIPPA_32769 is a putative CENP-A candidate in *P. papillatum* that forms distinct foci in the nucleus reminiscent of centromeric staining. (A) Multiple sequence alignment of the CENP-A candidate in *P. papillatum* with histone H3 and CENP-A from various eukaryotes. Top: secondary structure of histone H3-like proteins. Note that DIPPA_32769 has a longer loop 1 (highlighted in the red box) as well as replacement of Gln (arrow), which are characteristic (Continued on next page)

Fig 6 (Continued)

features of CENP-A. Note that DIPPA_03805 has a longer loop 1 but has Gln. CENP-A candidates are present in other diplomemids (*Rhynchopus humris* and *Hemistasia phaeocysticola*), and putatively in Prokinetoplastida (*Apiculatamorpho spiralis*). (B) Immunostaining of CENP-A-PrA visualized by standard fluorescence microscopy. (C) Immunostaining of CENP-A-V5 visualized using a confocal microscope confirming its distinct distribution pattern in the nucleus. Single slices are shown. Image files that have full z-sections are available in File S3, at Figshare. Scale bars: 10 μ m.

number of candidate interactors using a *P*-value cut-off of 0.01 (Fig. S7). Enriched proteins reside or have a function in the nucleolus, mitochondrion, and at the chromatin, such as nucleosomal histone proteins H2A (DIPPA_34783), H2B (DIPPA_34374), a protamine-like protein (DIPPA_34630) and an HMG-box protein (DIPPA_17448). Lastly, we found eight paralogous proteins that are Diplonemea-specific, and which are of apparent giant virus or prophage origin enriched in CENP-A-V5 pulldowns (Fig. S7 and S8A). Subsequent phylogenetic analysis of both phage, giant virus, and the diplomemid-specific paralogs reveals monophyletic clades for each of these groups with two giant virus sequences clustering more closely with the diplomemid sequence, one of which is *Theiavirus salishense* known to infect the kinetoplastid *Bodo saltans* (Fig. S8B). It is possible that similar giant viruses transferred this group of likely chromatin-related proteins to the ancestor of diplomemids, which may then have been co-opted in a diplomemid-specific kinetochore system.

Diplonemids have a putative spindle assembly checkpoint system

Mad2 is a component of the spindle checkpoint conserved in many eukaryotes (12). It interacts with Cdc20, an activator of the anaphase promoting complex, which has a Mad2-binding motif (MIM) (60). Mad2 is part of the HORMA domain family, which additionally includes p31^{COMET}, Hop1, and Rev7 (61). In *T. brucei* that lacks a canonical spindle checkpoint, the Mad2-like protein is associated with the basal body area and its Cdc20 lacks a Mad2-binding motif (21). Interestingly, our immunofluorescence assay revealed that Mad2 has nuclear localization in *P. papillatum* (Fig. 7A), implying that this protein may play a spindle checkpoint function. Consistent with this possibility, a Mad2-binding motif is present in *P. papillatum* Cdc20 (Fig. 7A), which is predicted by AF3 to interact with Mad2 (iPTM: 0.66). These findings suggest that a functional interaction is present between Mad2 and Cdc20 and that diplomemids likely have some form of a spindle checkpoint system.

The chromosomal passenger complex (CPC) in *P. papillatum* is unconventional

The CPC is essential for orchestrating both chromosome segregation and cytokinesis (63). In most eukaryotes, the CPC consists of Aurora B kinase, INCENP, Borealin, and Survivin. However, the CPC differs from this conventional form in *T. brucei*, where no orthologs of Borealin and Survivin have been identified. Besides Aurora B-like and a highly divergent INCENP-like protein CPC1, *T. brucei* has three novel CPC components known as CPC2, KIN-A, and KIN-B (Fig. 1C) (64, 65). Through our bioinformatics analyses, we identified four Aurora homologs (Fig. S6) and putative orthologs of INCENP (DIPPA_09943 and DIPPA_09958) and Survivin (DIPPA_06451) in *P. papillatum* (Fig. 2). Interestingly, these CPC components have atypical features in this flagellate. The C-terminus of INCENP is defined by an IN box, which is essential for Aurora B activation (66, 67). Although orthologs of INCENP in *Discoba* have a highly conserved IN box, the otherwise conserved Aurora activation site located downstream of the IN box is absent in Diplonemea, with DIPPA_09958 entirely lacking the IN box (Fig. 7B). It is unknown to which extent this truncation in Diplonemea influences its efficacy to activate Aurora, but we note that the *S. cerevisiae* ortholog appears similarly truncated (Fig. 7B) (66). *P. papillatum* is the only diplomemid with two INCENP orthologs, which likely arose from a recent duplication after which DIPPA_09958 lost the IN box.

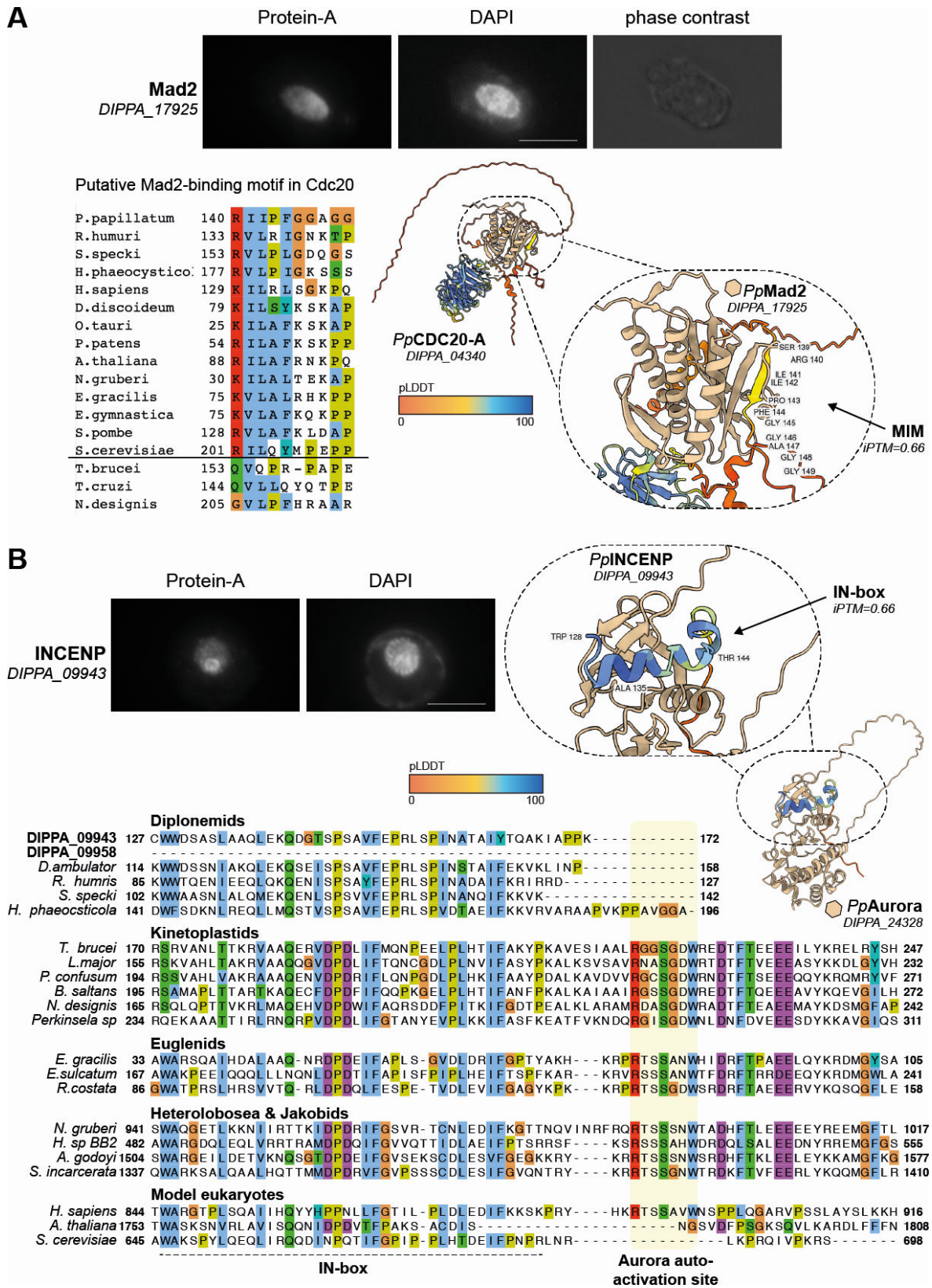


FIG 7 *P. papillatum* Mad2 and INCENP are nuclear proteins. (A) Mad2-PrA localizes in the nucleus. Scale bar: 10 μ m. Cdc20 in diplonemids (*P. papillatum*, *R. humris*, and *H. phaeocysticola*), but not in kinetoplastids (*T. brucei*, *Trypanosoma cruzi*, and *Neobodo designis*), has a putative Mad2-binding motif. AlphaFold3 prediction of *PpCdc20-A* (DIPPA_04340) interaction with *PpMad2* (DIPPA_17925) via the predicted MIM (62), with an iPTM score of 0.66. The predicted Mad2-Cdc20 interaction (Continued on next page)

Fig 7 (Continued)

indicates that a form of SAC-like signaling is active in *Diplonemea*. (B) *Pp*INCENP lacks the C-terminal cross-phosphorylation Threonine-Serine-Serine (TSS) site needed for proper activation for Aurora kinases in some model organisms, e.g., humans. Yeasts and *Arabidopsis* appeared to have lost this motif. AF3 prediction of INCENP interaction with Aurora paralogs pinpointed DIPPA_24328 as the most likely interaction partner (interface Predicted Template Modelling (iPTM) score:0.66). See Fig. S9 for AF3 interaction predictions between INCENP and Aurora kinases paralogs.

We then generated AF3 models (62) of INCENP^{DIPPA_09943} onto the four Aurora homologs in *P. papillatum* to investigate the binding of the IN box to the Aurora kinase domain and to delineate which Aurora homolog is its likely interaction partner. We find that the fold with the Aurora^{DIPPA_24328} (homologous to Aurora-3/AUK3 in *T. brucei*) (Figs. S6 and S9A) had both the highest iPTM score (0.66) as well as the highest pLDDT score in the IN box of DIPPA_09943 (Fig. 8A). Interestingly, INCENP^{DIPPA_09943} and Aurora^{DIPPA_24328} are predicted to interact in a similar manner as in other eukaryotes (Fig. S9).

Unexpectedly, we identified DIPPA_06451 as a BIR domain-containing protein that has a bidirectional best hit in our Foldseek analysis to the BIR domain found exclusively in the opisthokont Survivin orthologs. However, it has been shown recently that the BIR domain is not an ancestral feature of Survivin, as it is only found in Opisthokonta (68). Furthermore, we do not find homology of the N-terminus of DIPPA_06451 to the N-terminal coil that is the defining feature of Survivin orthologs across eukaryotes (68). DIPPA_06451 is therefore unlikely a true Survivin ortholog. This reiterates the importance of looking beyond a bidirectional best hit, which is often considered sufficient to classify orthologs.

Finally, we identified three putative KIN-A orthologs (DIPPA_17532, DIPPA_28866, DIPPA_16905; Fig. S4B) through a phylogenetic analysis of kinesin-related proteins. Our phylogeny places KIN-A and KIN-B as early branching clades nested within a larger clade of CENP-E orthologs. However, the KIN-B clade has a long branch, which is relatively poorly supported by bootstrap scores (54/100). By contrast, KIN-A is placed within the CENP-E group with a high support from bootstraps (96/100), providing a robust phylogenetic support for the notion that KIN-A might be a highly divergent CENP-E ortholog. It is important to note that this tree topology is unlikely due to under-sampling of KIN-A-related kinesins, as we have specifically included additional sequences from dedicated KIN-A and KIN-B HMM searches. Furthermore, a profile-versus-profile search shows that CENP-E is the most significant hit for KIN-A, and that KIN-A is among the top hits when the CENP-E profile is taken as a query (Table S4, at FigShare).

CLK^{KKT10/19} and SYCP2L1^{KKT17/18} localize in the nucleus

We next examined the localization of proteins that have similarity to kinetoplastid kinetochore proteins, but most likely perform other functions in *P. papillatum*, namely CLK^{KKT10/19} and SYCP2L1^{KKT17/18} (Fig. 8). Unlike CENP-A, we did not observe discrete dot signals for CLK^{KKT10/19} or SYCP2L1^{KKT17/18}, although both are localized in the nucleus. It therefore remains unclear whether they are kinetochore proteins in *P. papillatum*. It is noteworthy that expression of the SYCP2L1 protein was observed in our mitotically dividing cell culture despite the fact that SYCP2 is typically a meiosis-specific protein in other eukaryotes (69). Besides SYCP2L1, diplomids have another paralog (SYCP2L2; DIPPA_35886), similarly to KKT17/18 in kinetoplastids and SYCP2/SYCP2L in vertebrates (31). SYCP2-like proteins in part function during meiosis by specifically recruiting Hop1/HORMAD to the synaptonemal complex. Indeed, we find a predicted interaction motif (iPTM: 0.31), termed closure motif, in SYCP2L2 of *P. papillatum*, suggesting that this paralog might perform a function during meiosis (Fig. 8A; Fig. S10). DIPPA_25410 is a candidate SYCP3-like protein that may interact with the SYCP2L1-2 paralogs. Like SYCP2L1, expression of SYCP3 (DIPPA_25410) is found in the background of the CENP-A-V5 pulldown (Table S7, at FigShare).

The CLK kinase family proteins involved in splicing in other eukaryotes operate at kinetoplastid kinetochores as KKT10/19 (70). Since we observed that this gene family

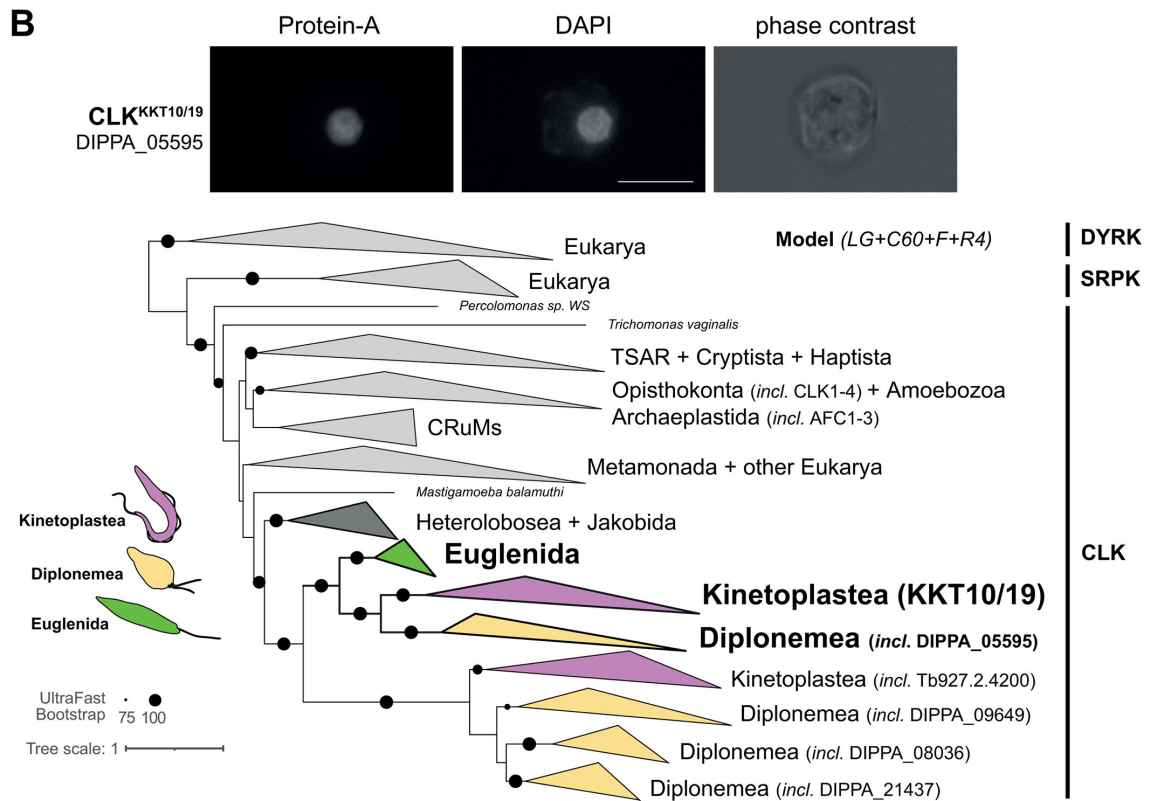
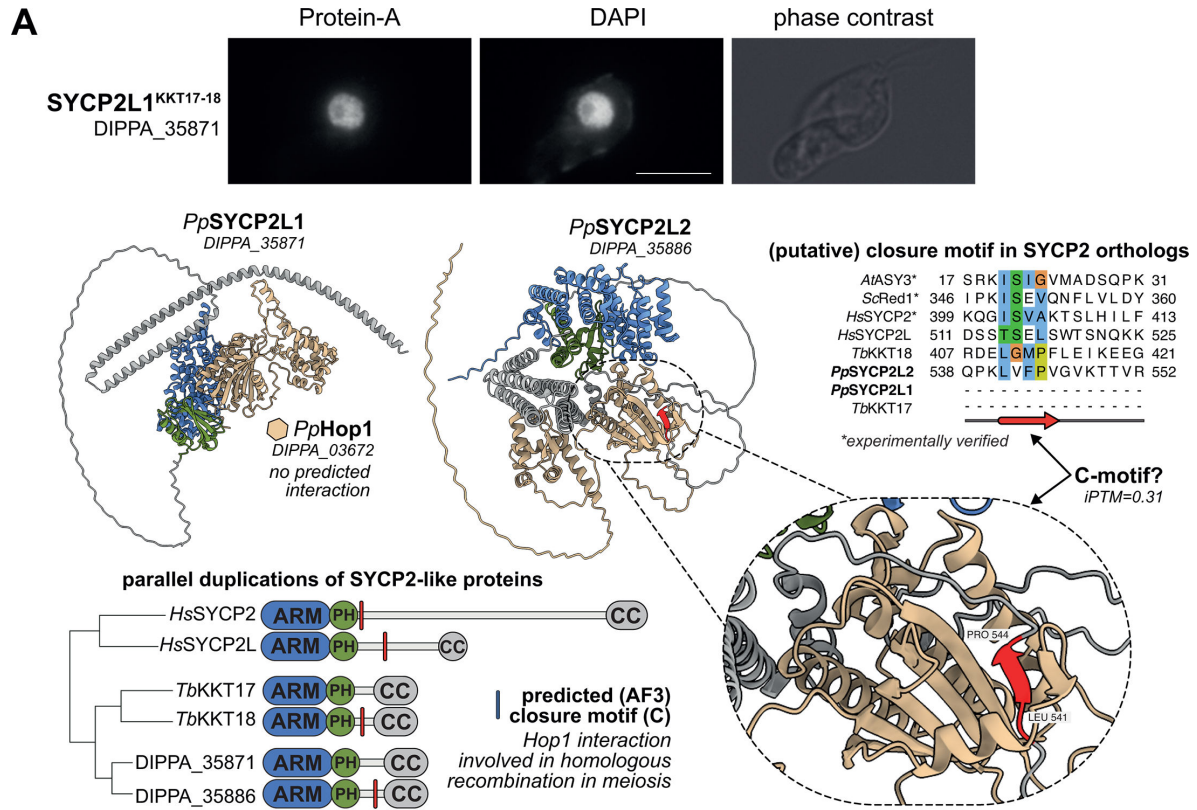


FIG 8 SYCP2L1^{KKT17/18} and CLK^{KKT10/19} are localized to the nucleus. (A) SYCP2L1^{KKT17/18} (DIPPA_35871)-PrA localizes in the nucleus. Scale bar: 10 μ m. Two paralogs of SYCP2 are present in *P. papillatum*: DIPPA_35871 (SYCP2L1) and DIPPA_35886 (SYCP2L2). Right: AF3-based predictions of putative closure motifs SYCP2 orthologs that interact with the HORMA protein Hop1, showing specifically the interactions predicted for SYCP2L1 and SYCP2L2. No closure motif was found in SYCP2L1, (Continued on next page)

Fig 8 (Continued)

hinting at a role for SYCP2L1 during mitosis and SYCP2L2 during meiosis (iPTM:0.31). Top right: alignment of verified and predicted closure motifs. (B) CLK^{KKT10/19} (DIPPA_05595)-PrA localizes in the nucleus. Scale bar: 10 μ m. Phylogenetic tree of the CLK kinase family including both KKT10/19 and the closest paralog in *P. papillatum* (DIPPA_05595).

underwent specific expansions in diplomonids, akin to humans and plants (71, 72), we wondered whether KKT10/19 were phylogenetically more associated to one particular CLK in these protists. Previous studies identified DIPPA_05595 as a KKT10/19 ortholog (29). Hence, we performed more extensive analyses that included additional eukaryotic lineages as well as available discoban lineages (Fig. 8B). We found that KKT10/19 and DIPPA_05595 follow the species tree and are part of a seemingly slower evolving clade of the CLKs in Euglenozoa, with a split at the base of Euglenozoa subsequently giving rise to three and one diplomonid and kinetoplastid CLK, respectively (Fig. 8B).

Taken together, our deep homology searches did not reveal orthologs for most of the kinetoplastid or canonical kinetochore subunits in Diplonemea, suggesting that a functionally analogous structure must be responsible for facilitating chromosome segregation. This as yet elusive kinetochore structure possibly consists of a novel type of kinetochore proteins.

DISCUSSION

Presently, the only group of euglenozoans where the kinetochore has been experimentally studied in detail are the trypanosomatid flagellates, which include human pathogens such as *T. brucei*, *T. cruzi*, and *Leishmania* spp. (22, 24, 73, 74). Unique KKT proteins identified in *T. brucei* are conserved even in early diverging prokinetoplastids (29), meaning that the KKT-based system likely already emerged in the last common ancestor of Kinetoplastea. To gain insights into the origin of the unique kinetoplastid kinetochore, and to uncover a putative novel kinetochore system, we sought to examine and identify kinetochore proteins in diplomonids, the sister clade of kinetoplastids (29, 33, 34). To this end, we developed a highly sensitive pipeline tailored to identify even highly divergent orthologs of the cKT and KKT, by leveraging sequence- and structure-based deep homology detection methods, using, for instance, a near-complete predicted foldome of the model diplomonid *P. papillatum*, and combining these with extensive phylogenetic analysis. Our approach has led to the identification of several highly divergent cKT orthologs, including a p31^{COMET}, Mad1, and Bub3 ortholog in *P. papillatum*, establishing the validity of our approaches. Yet, even these state-of-the-art methods were unable to identify additional cKT/KKT orthologs in *P. papillatum* and other diplomonids. In particular, they apparently do not have any of the structural kinetochore components of these systems, making it highly unlikely that these closest relatives of kinetoplastids possess any type of the known kinetochore repertoire.

Sensitivity of bioinformatics approaches to detect highly divergent homologs remains a recurring theme. Our previous efforts to characterize the evolution of kinetochore compositions across eukaryotes using iterative HMM searches suggested extensive or complete loss of core subunits not only in the euglenozoan lineages, but also in Metamonada and Alveolata (11). Although for some species like the metamonad *Carpodimonas membranifera*, the widespread loss of ancestral kinetochore subunits appears to hold (17), recent work in the apicomplexan parasite *Plasmodium berghei* revealed that a combination of advanced biochemistry (BioID), profile-vs-profile HMM searches, and AF2-based structure prediction was paramount to uncover extremely divergent subunits, which in hindsight appeared to be conventional counterparts (19). Our approach adds a full structural search of the predicted foldome of the model diplomonid *P. papillatum* using AlphaFold2 and Foldseek and includes high-resolution orthologous group (OG) definitions for 11 diplomonid species to strengthen profile-vs-profile HMM searches. In addition, we used standard HMM searches without heuristic filters, which we recently found out to yield more *bona fide* orthologs, though at the cost of higher computational time. To our knowledge, we, here, performed the most

extensive set of computational analyses to disprove the presence of any protein with similarity to both canonical and kinetoplastid kinetochore subunits, or any eukaryotic protein for that matter. *A priori* we can never fully exclude missed detection of orthologs. As an example, our recent work revealed the likely orthology of KKT14-15 with Bub1-Bub3, which was previously missed using standard BLAST and HMM searches (51). In light of the ever increasing sensitivity of tools to detect highly divergent homologs, skepticism toward bold claims of protein/gene absence appear justified—and perhaps the null hypothesis for any absence should be a detection problem. We therefore strongly advocate to any who put forward claims of surprising losses to follow our approach and to go to great lengths to disprove the presence of highly divergent orthologs.

The only canonical structural kinetochore candidate identifiable in diplomemids is CENP-A, a variant histone that replaces the canonical histone H3 at centromeres. Based on the findings that the histone H3 variant in *P. papillatum* has many features of CENP-A and localizes to discrete foci in the nucleus, we deem it most likely that it is a true CENP-A homolog. Early studies using electron microscopy found numerous short condensed chromosomes in the interphase nucleus of *P. papillatum* and *Diplonema ambulator* (75, 76). More recently, using expansion microscopy of *P. papillatum* (46) and transmission electron microscopy of *Lacrimia vacuolata* (39), we were able to visualize sub-nuclear structures, namely the large centrally located nucleolus, numerous peripheral concentric-oriented and condensed chromosomes, and one or two small electron-dense spherical bodies. Consistent with a high number of chromosomes in diplomemid nuclei, confocal microscopy allowed us to discern many CENP-A-V5 foci. However, it is important to note that we do not have a kinetochore marker to prove that these signals represent true kinetochore localization and, notably, CENP-A is not the only divergent H3 variant that localizes into the nuclear foci. For example, there is an H3 variant in *T. brucei* which has been shown to localize to telomeric foci (77). Nonetheless, formation of discrete nuclear foci, together with its previously noted structural and sequence features (Fig. 6), suggests that DIPPA_32769 is likely a true CENP-A ortholog in *P. papillatum*, indicating presence of the cKT nucleosomal cornerstone.

We identified numerous key components of a putative spindle assembly checkpoint in *P. papillatum*, such as Mad2, Cdc20, and a Bub3 candidate. However, members of the central checkpoint protein family Bub(R)1/Mad3 were not detected. It is intriguing that *P. papillatum* has a Mad2-like protein that localizes in the nucleus, unlike *T. brucei* Mad2 that localizes near the basal body (21). Together with the finding that *P. papillatum* Cdc20 has a conserved Mad2-binding motif, it is likely that Mad2 interacts with Cdc20 to regulate APC/C activity. Interestingly, diplomemids and the early branching prokinetoplastids encode p31^{COMET}, which in other eukaryotes is involved in the regulation of the open and closed forms of Mad2 and other HORMA domain proteins (78). While it was thought that Mad1, a kinetochore recruitment factor for Mad2, would be absent from kinetoplastids, our iterative HMM searches in combination with Foldseek of AF2-predicted folds revealed the presence of an RWD domain reminiscent of Mad1 in the *T. brucei* protein LAP71, which was recently described to be a nuclear pore protein (79). Candidate orthologs in diplomemids were also found, including the *P. papillatum* protein DIPPA_30775, although a putative MIM could not be discerned. It is possible that these Mad1-like proteins have other functions at the nuclear pore complex in kinetoplastids and diplomemids. In any case, it will be important to examine whether *P. papillatum* has a functional spindle assembly checkpoint and delays cell cycle progression in response to treatment with microtubule poisons.

In addition to the absence of the bulk of the cKT, we also did not find critical components of the KKT. Instead, we only found orthologs for proteins that are broadly conserved among eukaryotes with non-kinetochore functions (e.g., CLK and SYCP2) (29, 31, 80). Hence, these proteins most likely also have non-kinetochore functions in kinetoplastids and diplomemids. In traditional model eukaryotes such as humans and *Drosophila melanogaster*, CLK kinases do not localize at kinetochores but are instead

involved in the regulation of splicing (81). Similarly, KKT16/17/18 have a combination of Armadillo, PH, and coiled-coil domains, which are found in the broadly conserved synaptonemal complex proteins SYCP2 and SYCP3 (31). The facts that euglenids have homologs of the canonical kinetochore proteins and that additional orthologs of the kinetoplastid-specific KKT proteins have not been identified suggest that euglenids most likely have canonical kinetochores and that their CLK^{KKT10/19} and SYCP2/3^{KKT16/17/18} proteins have non-kinetochore functions. Although these proteins localize in the nucleus of *P. papillatum*, further studies are needed to reveal their functions.

The cell lines prepared in this study will serve as an invaluable resource in further dissection of the seemingly unique *P. papillatum* kinetochore whose composition remains unknown. Although the resolution of the performed pull-down experiments with CENP-A-V5 was generally low, our experiments hint at a putative interaction with eight proteins that all bear a homologous coiled-coil domain that is of prophage origin. These proteins should be characterized in the future, as they might play a role in the diplonemid kinetochore.

Diplonemids are among the most abundant and diverse protists in the oceans (35–38), yet we know very little about their basic cell biology (44–47). Here, we have shown that these heterotrophic flagellates must segregate their chromosomes in a process involving an as yet elusive kinetochore-like structure, a core apparatus in the eukaryotic cell cycle. Moreover, as a sister clade to the (mostly) parasitic kinetoplastids, which are endowed with a number of unique features (82), diplonemid cell biology may shed light on their evolution as well as increase our understanding of one of the most abundant yet deeply understudied group of marine eukaryotes.

MATERIALS AND METHODS

Diplonemid transcriptomic data set

For this study, a set of predicted proteomes of 11 diplonemid species was used: 1 genome (*P. papillatum*) (46) and 10 transcriptomes (29, 46, 83–85) (Table S1, at FigShare). For *de novo* assembly of RNA-Seq data sets, Trinity v.2.2.0 software was used with default parameters (<https://github.com/trinityrnaseq/trinityrnaseq>) (86). TransDecoder v.5.5.0 (Haas B.J., <https://github.com/TransDecoder/TransDecoder>) was used to generate proteomes using “Universal” genetic code. Proteins were then clustered with CD-HIT v.4.8.1 at 99% identity (<https://github.com/weizhongli/cdhit/>) (87) to reduce redundancy and proteome complexity. *Paradiplonema papillatum* identifiers were translated to the following: “Paradiplonema_papillatum_XXXXXX,” where x is a six-digit number (Table S2, at FigShare).

AlphaFold2 foldome for *P. papillatum*

The structures of all predicted protein sequences from the recently sequenced *P. papillatum* genome were predicted with AF2 (46, 88). It has been shown that supplying multiple sequence alignments (MSAs) from closely related species can improve the confidence and likely accuracy of AF2 models, especially for sequences from divergent and underrepresented taxa (89). MSAs were generated using the MMseqs package (90), where each *P. papillatum* sequence was clustered and aligned to sequences in a database comprised of the EukProt v.3 data set (91), the Discoba-specific data set by reference (89), and our own set of predicted diplonemid proteomes (see above). The AF2 algorithm was run on these MSAs using the ColabFold implementation (92). An upper limit of 3,000 amino acids per *P. papillatum* amino acid sequence was imposed for computational tractability, resulting in a total of 41,990 predicted structures. The AF2 predictions with the highest average pLDDT per protein (rank 1) and all other related files to the AF2 predictions can be accessed through FigShare.

Deep homology detection protocols

A three-pronged highly sensitive homology detection was used consisting of (i) a unidirectional profile HMM search of known kinetochore components against the *P. papillatum* proteome, (ii) an extensive profile-vs-profile HMM search, and (iii) an extensive three-dimensional protein structure-based search. A graphic overview of the pipeline for the Diplonemea data set generation and the HHsearch/AF2 strategy can be found in Fig. S1 and S2. Raw files pertaining to these searches can be accessed through FigShare.

Unidirectional Hidden Markov Model searches without heuristic filters

Curated multiple sequence alignments of both full-length and/or domains and motifs for 206 canonical KT orthologs/outparalogs (7), including highly divergent apicomplexan kinetochores (19), as well as 36 kinetoplastid-specific kinetochore (interacting) (KKT/KKIP) proteins, were converted into profile HMMs using the *hmmbuild* from HMMer package (v.3.3, custom settings) (93). To distinguish from previous unidirectional profile HMM approaches (29), searches were performed with *hmmsearch* without any of the standard heuristic filters (option `-max`) that are commonly in place to maintain reasonable computational tractability in larger sequence databases. Surprisingly, we have found that turning these heuristic filters off regularly reveals high scoring “hits” of highly divergent homologs.

Profile-versus-profile HMM

OrthoFinder (v.2.5.5, `-S blast`; otherwise default settings) was used to generate OGs from the predicted proteomes in our diplomemid set (see above) (94). OGs were aligned with MAFFT (v.7.520, `E-INS-i`) (95). Using the HHSuite package (v3, standard settings), a consensus sequence was derived for each alignment (.a3m format) in order to generate secondary structure decorated profile HMMs (.hmm format), which were collated in a custom HHsearch database (96). The same procedure was followed for our manually curated set of known kinetochore proteins (see above) and queried using the hhsearch tool against the diplomemid database. First, our manually curated set of KT/KKT/KKIP/AKiT alignments was queried against the diplomemid OG database. The top 10 hits for each of these initial searches were taken and were subsequently queried against the KT/KKT database, the diplomemid database, as well as against the COG/KOG and PFAM precomputed hh-suite database (https://wwwuser.gwdguser.de/~compbiol/data/hhsuite/databases/hhsuite_dbs/, last accessed 01-04-2024) to provide a comprehensive and high-quality background set of profiles. This search step is considered the “forward search.” In turn, the top 10 hits for each of these forward searches was again queried against the same set of databases in a step considered the “reverse search.” Finally, the top 10 hits from all forward and reverse searches were parsed and compiled into a data set that describes the relations of the queries and top 10 targets of each performed search. This data set was analyzed using Cytoscape software v.3.10.2 (97). MCL was performed on the data set to identify clusters of nodes, which typically are most strongly interconnected among each other and show similar levels of connectivity to other regions in the network, which is indicative of a closely related phylogenetic affiliation relative to other nodes in the network. MCL was performed with `-log(E-value)` taken as input array for edge weights and an inflation parameter of $I = 2$. The MCL network can be found in File S1, at FigShare.

3D protein structure comparison

Manually curated AF2 models (downloaded from uniprot, last accessed 25 March 2024) of known KT/KKT/KKIP proteins from four representative species, *Arabidopsis thaliana*, *Homo sapiens*, *Saccharomyces cerevisiae*, and *Trypanosoma brucei* (88), were searched against our proteome-wide predicted structures for *P. papillatum* using Foldseek (98).

Both full-length and domain-specific structures were queried. First, an initial search was performed against only the *P. papillatum* AF2 foldome, from which the top 10 hits were taken as query for a “forward” search against the complete database, comprising of the curated cKT/KKT/KKIP folds, the *P. papillatum* AF2 foldome, and the complete foldomes of the representative species. Finally, the top 10 hits for each “forward” query were taken and were queried again against the complete database in a “reverse” search step. The top 10 hits for all forward and reverse searches were compiled into a data set which was analyzed with the Cytoscape software, primarily through the MCL algorithm implementation. MCL was performed with $-\log(\text{E-value})$ taken as input array for edge weights and an inflation parameter of $l = 7$. The MCL network can be found in File S2, at FigShare.

AlphaFold3 protein-protein interaction modeling

To structurally model protein-protein interactions, AlphaFold3 was used through its online implementation at alphafoldserver.com (62), last accessed 13 July 2024.

Phylogenetic analyses

To delineate orthologs from homologs in select cases, phylogenetic analyses were performed for several orthologous groups as follows. First, similarity searches were performed using a combination of *hmmsearch* and *jackhmmer* from the HMMER package (v.3.1b2). These searches were conducted against our in-house database (Table S1, at FigShare) and a eukaryote-wide selection of proteomes as used by De Potter et al. (99). The obtained homologs were then aligned with MAFFT E-INS-i or L-INS-i (95). Trimming of the MSA was performed using trimAl. Sites were removed with the criterion $\leq 10\%$ or $\leq 30\%$ occupancy, and finally, poorly aligned sequences were manually discarded. A phylogenetic tree was then inferred using IQ-TREE (100) with the best-fitting substitution model as chosen by ModelFinder (101), which in all cases converged on LG + C60+F + R4. In all cases, the top 500 from our searches were included in the phylogenetic analyses. For BUB3, we specifically added a known outgroup for RAE1/BUB3: WDR74 (51). A set of WDR74 ortholog sequences was downloaded from the eggNOG database (KOG3881; v.5). These WDR74 sequences were combined with manually curated MSAs of BUB3, RAE1, and KKT15. Then, the Diplonemea sequences that were assigned to the same OrthoFinder orthologous group as our candidate BUB3 sequence were also added to this set of sequences. For Aurora/Polo/KKT2/3, we included previously established Aurora and Polo clusters (7), including the LECA kinase OG PLK, Aurora, and PLK4, PLK and PLK4 sharing the most recent duplication, and Aurora being an outgroup.

Endogenous C-terminal tagging of CENP-A, Mad2, INCENP, CLK^{KKT10/19}, and SYCP2L1^{KKT17/18}

Parts of the open reading frame and 3′ untranslated region of *P. papillatum* genomic DNA were amplified using specific primers containing sequences overlapping with the protein A neomycin cassette of the plasmid pDP002 in order to endogenously tag CENP-A, CLK^{KKT10/19}, SYCP2L1^{KKT17/18}, INCENP, and Mad2 with a C-terminal protein A tag. The pDP002 plasmid was used as a template for amplification of the protein A tag and the downstream Neo^R marker. To ligate all three fragments together, Phusion polymerase (NEB) was applied in a nested PCR using these fragments as a template. Finally, 3 μg –5 μg of the nested PCR product was electroporated into *P. papillatum* as described elsewhere (43). Similarly, the 3xV5 and hygromycin^R cassette was amplified from a slightly modified pDP011 plasmid (44), named pDP011-A (for primers, see Table S8 at FigShare). The PCR product was A-tailed and cloned into pCR 2.1-TOPO (ThermoFisher) or alternatively gel-purified, ethanol-precipitated, and used directly for transfections. Restriction enzyme digestion and sequencing were used to confirm the correct structure of the generated plasmid (if applicable). To release the tagging segment, 10 μg of the final plasmid was

cut with EcoRI (NEB), ethanol-precipitated, resuspended in 10 μL of water, and used to transfect *P. papillatum* as described elsewhere (43).

Electroporation of *P. papillatum*

Amaxa Nucleofector II was used to transform a total of 5×10^7 cells, as previously described (41, 42). Clones were selected in 24-well plates at 27°C using varied doses of G418 for protein A tagging (25 $\mu\text{g}/\text{mL}$ –80 $\mu\text{g}/\text{mL}$) and of hygromycin for V5 tagging (100 $\mu\text{g}/\text{mL}$ –225 $\mu\text{g}/\text{mL}$). Successful transfectants appeared after 2 weeks of selection. Before being tested by western blot, each clone was expanded to a volume of 10 mL and cultured for up to 2 to 3 weeks.

Immunofluorescence

A 5 to 10 mL of logarithmically growing cell culture was centrifuged at $1,000 \times g$ for 5 min. Cells were fixed in 4% paraformaldehyde in seawater at room temperature for 30 min. The fixative was washed out of the cells with seawater and 1 \times PBS (phosphate buffered saline) (1:1). The cells were rinsed once more in 1 \times PBS before being spotted on a gelatin-coated slide, and permeabilized for 20 min for antibody labeling in 100% ice-cold methanol. Throughout the procedure, the slides were stored in a humidified chamber, washed in PBS after 20 min, and blocked with 5% milk in PBS-T (PBS with 0.05% Tween) for 45 min. After removing the blocking solution, the cells were again washed in 1 \times PBS. The primary antibody (anti-protein A; Sigma, 1:2,000) was applied on slides covered with parafilm and incubated overnight at 4°C. After removing the primary antibody, the slides were washed three times with 1 \times PBS. The secondary antibody (Alexa Fluor 555 goat-anti-rabbit; Invitrogen, 1:1,000) was added and the slides were incubated for 1 hour in the dark at room temperature, covered with parafilm. Finally, the slides were washed in 1 \times PBS and covered with 4',6-diamidino-2-phenylindole (DAPI) containing the antifade reagent ProlongGold (Life Technologies). Images for protein A fusion proteins were obtained using camera Olympus DP73 (Axioplan 2 imaging). CENP-A-V5 images were acquired on a FV3000 Olympus confocal microscope with an HCPL apochromatic 100 \times /NA 1.40 oil immersion objective. Excitation was performed with a 405 nm diode laser (50 mW) in the case of DAPI and in the case of Red channel which is Alexa Fluor 555 (543 nm, 555 nm–615 nm); emissions were detected using hybrid detectors (HyD). Z-stacks were acquired with a step size of 100 nm (without averaging). The pixel size and dwell time were between 52 and 97 nm and 400 ns, respectively. The size of the pinhole was adjusted based on the signal strength but was typically around 0.4 Airy unit to improve resolution.

Western blotting

Protein samples from 5×10^5 cells were prepared by pelleting, resuspended in 25 μL of 2 \times SDS sample buffer, separated on 4%–20% Mini-protein TGX stain-free gels (Bio-Rad), and subsequently transferred to a polyvinylidene fluoride (PVDF) membrane. After blocking with 5% milk in PBS-T for at least 30 min at room temperature, the membrane was incubated with an anti-protein A (Sigma, P3775; used at 1:10,000 dilution) or anti-V5 antibody (Merck, V8137; used at 1:1,000 dilution) in 5% milk in PBS-T overnight at 4°C. After three washes in PBS-T, the membrane was incubated with anti-rabbit-horseradish peroxidase (HRP) (Merck; 1:1,000) and incubated at room temperature for 1 hour. The membrane was then washed three times in PBS-T and the signal was developed using Clarity Western ECL Substrate (Bio-Rad). The mouse anti-alpha-tubulin antibody (Merck, T9026; 1:10,000) was used as a loading control.

Immunoprecipitation

Approximately 5×10^8 cells expressing V5-tagged CENP-A proteins, as well as the wild-type control cells were grown in Diplonema growth media with the appropriate

selection antibiotic. Cells were harvested at 1,000 g for 10 min, lysed using ice cold IPP150 buffer (10 mM Tris-HCl, pH 6.8, 150 mM NaCl, 0.1% Igepal, CA-630) supplemented with 1× complete EDTA-free protease inhibitors (Sigma-Aldrich, 11873580001) and 5× passed through a 30 gauge needle. The cell lysate was cleared twice by centrifugation (12,000 g, 10 min, 4°C). Fifty microliters of V5-Trap magnetic particles (M-270) (Chromotek, v5td-100) was added to the cleared cell lysate and rotated at 4°C for 2 to 3 hours. The beads were washed three times with the IPP150 buffer containing detergent and then twice without the detergent. Bound proteins with beads were processed for immunoblotting and mass spectrometry.

Mass spectrometry analysis

For mass spectrometry sample prep, on-bead digestion protocol was used. Briefly, washed beads were resuspended in sodium deoxycholate [SDC, final concentration 1% (wt/vol) in 100 mM triethylammonium bicarbonate], reduced with 5 mM TCEP [tris(2-carboxyethyl)phosphine], alkylated with 10 mM MMTS (S-methyl methanethiosulfonate), and digested sequentially with Lys-C and trypsin. SDC was removed by extraction with ethyl acetate saturated with water (102). Samples were desalted on Empore C18 columns, dried in a SpeedVac, and dissolved in 0.1% trifluoroacetic acid (TFA) + 2% acetonitrile. Desalted peptide digests were separated on a 50 cm C18 column using 60-min elution gradient and analyzed in a Data Dependent Acquisition (DDA) mode on an Orbitrap Exploris 480 (Thermo Fisher Scientific) mass spectrometer equipped with the FAIMS unit operated at –40 and –60 V compensation voltages (CVs).

Data processing

The resulting raw files were converted to mzXML files each containing separate CV (–40, –60 V) scans using FAIMS MzXML Generator (release 1.1.8003, <https://github.com/PNNL-Comp-Mass-Spec/FAIMS-MzXML-Generator>). MzXML files were analyzed in MaxQuant (v.2.2.0.0) (103) with label-free quantification (LFQ) algorithm MaxLFQ and match between runs feature activated. False Discovery Rate (FDR) was set as 0.01 at all levels. A custom protein database of *P. papillatum* proteins (43,871 sequences) supplemented with frequently observed contaminants was used for protein identification. MMTS alkylated cysteine was selected as a fixed modification [Methylthio (C), composition: H(2) C S, +45.988]. Variable modifications were oxidation (M) and acetyl (protein N-term). Downstream processing of the proteinGroups.txt file was performed in Perseus (v.2.0.6.0) (104).

ACKNOWLEDGMENTS

We acknowledge support from the Czech Grant Agency 23-06479X (to J.L.) and 23-07674S (to C.B.), a Veni Fellowship from the Dutch Research Council (NWO, VI.Veni.202.223) (to E.C.T.), a Wellcome Discovery Award (227243/Z/23/Z), and a Centre Core Grant to the Wellcome Trust Centre for Cell Biology (203149) (to B.A.). We heartily thank Matúš Valach (University of Montreal, Montreal) for providing transcriptomes for 10 diplomids. The authors would like to acknowledge the Proteomics Service Laboratory, Institute of Physiology and Institute of Molecular Genetics (Czech Academy of Sciences, Prague) and specifically Marek Vrbacký for mass spectrometry proteomics analysis, technical advice, and expertise. We thank the members of the Kops and Snel groups for feedback. The authors declare no competing financial interests.

AUTHOR AFFILIATIONS

¹Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis), Czechia

²Oncode Institute, Hubrecht Institute, Royal Academy of Arts and Sciences, Utrecht, the Netherlands

³Theoretical Biology and Bioinformatics, Department of Biology, Faculty of Science, Utrecht University, Utrecht, the Netherlands

⁴Faculty of Sciences, University of South Bohemia, České Budějovice (Budweis), Czechia

⁵Cell Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, Faculty of Science and Engineering, University of Groningen, Groningen, the Netherlands

⁶The Wellcome Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

AUTHOR ORCID*s*

Corinna Benz  <http://orcid.org/0000-0003-4367-2758>

Maximilian W. D. Raas  <http://orcid.org/0000-0003-4986-2421>

Pragya Tripathi  <http://orcid.org/0000-0002-3290-6275>

Drahomíra Faktorová  <http://orcid.org/0000-0001-9623-2233>

Eelco C. Tromer  <http://orcid.org/0000-0003-3540-7727>

Bungo Akiyoshi  <http://orcid.org/0000-0001-6010-394X>

Julius Lukeš  <http://orcid.org/0000-0002-0578-6618>

AUTHOR CONTRIBUTIONS

Corinna Benz, Data curation, Formal analysis, Investigation, Methodology | Maximilian W. D. Raas, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review and editing | Pragya Tripathi, Investigation, Visualization | Drahomíra Faktorová, Formal analysis, Methodology, Visualization | Eelco C. Tromer, Conceptualization, Data curation, Formal analysis, Funding acquisition, Supervision, Validation, Writing – original draft, Writing – review and editing | Bungo Akiyoshi, Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing – original draft, Writing – review and editing | Julius Lukeš, Conceptualization, Funding acquisition, Project administration, Writing – review and editing

DIRECT CONTRIBUTION

This article is a direct contribution from Julius Lukeš, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by David Horn, University of Dundee School of Life Sciences, and Richard McCulloch, University of Glasgow.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (105) with the data set identifier [PXD054535](https://doi.org/10.1007/s12064-013-0184-5). All original data used for, and results generated by, the homology searches as well as copies of the figures, supplementary figures, supplementary file & supplementary tables can be accessed through the online Data Repository FigShare at <https://figshare.com/s/a3c0f6af159c4545f393>.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (mBio02936-24-s0001.docx). Supplemental figures and data link.

REFERENCES

1. Chodasewicz K. 2014. Evolution, reproduction and definition of life. *Theory Biosci* 133:39–45. <https://doi.org/10.1007/s12064-013-0184-5>
2. Tetz VV, Tetz GV. 2020. A new biological definition of life. *Biomol Concepts* 11:1–6. <https://doi.org/10.1515/bmc-2020-0001>

3. McIntosh JR. 2016. Mitosis. *Cold Spring Harb Perspect Biol* 8:a023218. <https://doi.org/10.1101/cshperspect.a023218>
4. McAinsh AD, Kops GJPL. 2023. Principles and dynamics of spindle assembly checkpoint signalling. *Nat Rev Mol Cell Biol* 24:543–559. <https://doi.org/10.1038/s41580-023-00593-z>
5. Keeling PJ, Burki F. 2019. Progress towards the Tree of Eukaryotes. *Curr Biol* 29:R808–R817. <https://doi.org/10.1016/j.cub.2019.07.031>
6. Drinnenberg IA, Akiyoshi B. 2017. Evolutionary lessons from species with unique kinetochores. *Prog Mol Subcell Biol* 56:111–138. https://doi.org/10.1007/978-3-319-58592-5_5
7. Tromer EC, van Hooff JJE, Kops GJPL, Snel B. 2019. Mosaic origin of the eukaryotic kinetochore. *Proc Natl Acad Sci U S A* 116:12873–12882. <https://doi.org/10.1073/pnas.1821945116>
8. Musacchio A, Desai A. 2017. A molecular view of kinetochore assembly and function. *Biology (Basel)* 6:5. <https://doi.org/10.3390/biology6010005>
9. McAinsh AD, Marston AL. 2022. The four causes: the functional architecture of centromeres and kinetochores. *Annu Rev Genet* 56:279–314. <https://doi.org/10.1146/annurev-genet-072820-034559>
10. Meraldi P, McAinsh AD, Rheinbay E, Sorger PK. 2006. Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol* 7:R23. <https://doi.org/10.1186/gb-2006-7-3-r23>
11. van Hooff JJ, Tromer E, van Wijk LM, Snel B, Kops GJ. 2017. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep* 18:1559–1571. <https://doi.org/10.15252/embr.201744102>
12. Kops G, Snel B, Tromer EC. 2020. Evolutionary dynamics of the spindle assembly checkpoint in eukaryotes. *Curr Biol* 30:R589–R602. <https://doi.org/10.1016/j.cub.2020.02.021>
13. Ishii M, Akiyoshi B. 2022. Plasticity in centromere organization and kinetochore composition: Lessons from diversity. *Curr Opin Cell Biol* 74:47–54. <https://doi.org/10.1016/j.cub.2021.12.007>
14. Drinnenberg IA, deYoung D, Henikoff S, Malik HS. 2014. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *eLife* 3:e03676. <https://doi.org/10.7554/eLife.03676>
15. Navarro-Mendoza MI, Pérez-Arques C, Panchal S, Nicolás FE, Mondo SJ, Ganguly P, Pangilinan J, Grigoriev IV, Heitman J, Sanyal K, Garre V. 2019. Early diverging fungus *Mucor circinelloides* lacks centromeric histone CENP-A and displays a mosaic of point and regional centromeres. *Curr Biol* 29:3791–3802. <https://doi.org/10.1016/j.cub.2019.09.024>
16. Cortes-Silva N, Ulmer J, Kiuchi T, Hsieh E, Cornilleau G, Ladid I, Dingli F, Loew D, Katsuma S, Drinnenberg IA. 2020. CenH3-independent kinetochore assembly in lepidoptera requires CCAN, including CENP-T. *Curr Biol* 30:561–572. <https://doi.org/10.1016/j.cub.2019.12.014>
17. Salas-Leiva DE, Tromer EC, Curtis BA, Jerlström-Hultqvist J, Kolisko M, Yi Z, Salas-Leiva JS, Gallot-Lavallée L, Williams SK, Kops G, Archibald JM, Simpson AGB, Roger AJ. 2021. Genomic analysis finds no evidence of canonical eukaryotic DNA processing complexes in a free-living protist. *Nat Commun* 12:6003. <https://doi.org/10.1038/s41467-021-26077-2>
18. Zeeshan M, Pandey R, Ferguson DJP, Tromer EC, Markus R, Abel S, Brady D, Daniel E, Limenitakis R, Bottrill AR, Le Roch KG, Holder AA, Waller RF, Guttery DS, Tewari R. 2020. Real-time dynamics of *Plasmodium* NDC80 reveals unusual modes of chromosome segregation during parasite proliferation. *J Cell Sci* 134:jcs245753. <https://doi.org/10.1242/jcs.245753>
19. Brusini L, Dos Santos Pacheco N, Tromer EC, Soldati-Favre D, Brochet M. 2022. Composition and organization of kinetochores show plasticity in apicomplexan chromosome segregation. *J Cell Biol* 221:e202111084. <https://doi.org/10.1083/jcb.202111084>
20. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renaud H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, et al. 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309:416–422. <https://doi.org/10.1126/science.1112642>
21. Akiyoshi B, Gull K. 2013. Evolutionary cell biology of chromosome segregation: insights from trypanosomes. *Open Biol* 3:130023. <https://doi.org/10.1098/rsob.130023>
22. Akiyoshi B, Gull K. 2014. Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156:1247–1258. <https://doi.org/10.1016/j.cell.2014.01.049>
23. Nerusheva OO, Akiyoshi B. 2016. Divergent polo box domains underpin the unique kinetoplastid kinetochore. *Open Biol* 6:150206. <https://doi.org/10.1098/rsob.150206>
24. D'Archivio S, Wickstead B. 2017. Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J Cell Biol* 216:379–391. <https://doi.org/10.1083/jcb.201608043>
25. Nerusheva OO, Ludzia P, Akiyoshi B. 2019. Identification of four unconventional kinetoplastid kinetochore proteins KKT22-25 in *Trypanosoma brucei*. *Open Biol* 9:190236. <https://doi.org/10.1098/rsob.190236>
26. Brusini L, D'Archivio S, McDonald J, Wickstead B. 2021. Trypanosome KKI1 dynamically links the inner kinetochore to a kinetoplastid outer kinetochore complex. *Front Cell Infect Microbiol* 11:641174. <https://doi.org/10.3389/fcimb.2021.641174>
27. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, et al. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* 38:D457–D462. <https://doi.org/10.1093/nar/gkp851>
28. Jackson AP, Otto TD, Aslett M, Armstrong SD, Bringaud F, Schlacht A, Hartley C, Sanders M, Wastling JM, Dacks JB, Acosta-Serrano A, Field MC, Ginger ML, Berriman M. 2016. Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr Biol* 26:161–172. <https://doi.org/10.1016/j.cub.2015.11.055>
29. Butenko A, Opperdoes FR, Flegontova O, Horák A, Hampl V, Keeling P, Gawryluk RMR, Tikhonenkov D, Flegontov P, Lukeš J. 2020. Evolution of metabolic capabilities and molecular features of diplomonids, kinetoplastids, and euglenids. *BMC Biol* 18:23. <https://doi.org/10.1186/s12915-020-0754-1>
30. Tikhonenkov DV, Gawryluk RMR, Mylnikov AP, Keeling PJ. 2021. First finding of free-living representatives of Prokinetoplastina and their nuclear and mitochondrial genomes. *Sci Rep* 11:2946. <https://doi.org/10.1038/s41598-021-82369-z>
31. Tromer EC, Wemyss TA, Ludzia P, Waller RF, Akiyoshi B. 2021. Repurposing of synaptonemal complex proteins for kinetochores in Kinetoplastida. *Open Biol* 11:210049. <https://doi.org/10.1098/rsob.210049>
32. Cavalier-Smith T. 2016. Higher classification and phylogeny of Euglenozoa. *Eur J Protistol* 56:250–276. <https://doi.org/10.1016/j.ejop.2016.09.003>
33. Kostygov AY, Karnkowska A, Votýpka J, Tashyreva D, Maciszewski K, Yurchenko V, Lukeš J. 2021. Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol* 11:200407. <https://doi.org/10.1098/rsob.200407>
34. Lax G, Kolisko M, Eglit Y, Lee WJ, Yubuki N, Karnkowska A, Leander BS, Burger G, Keeling PJ, Simpson AGB. 2021. Multigene phylogenetics of euglenids based on single-cell transcriptomics of diverse phagotrophs. *Mol Phylogenet Evol* 159:107088. <https://doi.org/10.1016/j.ympev.2021.107088>
35. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, et al. 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605. <https://doi.org/10.1126/science.1261605>
36. Flegontova Olga, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, Bowler C, Lukeš J, Horák A. 2016. Extreme diversity of diplomonid eukaryotes in the ocean. *Curr Biol* 26:3060–3065. <https://doi.org/10.1016/j.cub.2016.09.031>
37. Tashyreva D, Simpson AGB, Prokopchuk G, Škodová-Sveráková I, Butenko A, Hammond M, George EE, Flegontova O, Záhonová K, Faktorová D, Yabuki A, Horák A, Keeling PJ, Lukeš J. 2022. Diplonemids - A review on “new” flagellates on the oceanic block. *Protist* 173:125868. <https://doi.org/10.1016/j.protis.2022.125868>
38. Flegontova O, Flegontov P, Londoño PAC, Walczowski W, Šantić D, Edgcomb VP, Lukeš J, Horák A. 2020. Environmental determinants of the distribution of planktonic diplomonids and kinetoplastids in the oceans. *Environ Microbiol* 22:4014–4031. <https://doi.org/10.1111/1462-2920.15190>
39. Tashyreva D, Týč J, Horák A, Lukeš J. 2023. Ultrastructure and 3D reconstruction of a diplomonid protist (Diplonemea) and its novel membranous organelle. *mBio* 14:e0192123. <https://doi.org/10.1128/mbio.01921-23>

40. van Hooff JJE, Raas MWD, Tromer EC, Eme L. 2024. Shaping up genomes: prokaryotic roots and eukaryotic diversification of SMC complexes. *bioRxiv*. <https://doi.org/10.1101/2024.01.07.573240>
41. Kaur B, Valach M, Peña-Díaz P, Moreira S, Keeling PJ, Burger G, Lukeš J, Faktorová D. 2018. Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine microeukaryotes Diplonemida (Euglenozoa). *Environ Microbiol* 20:1030–1040. <https://doi.org/10.1111/1462-2920.14041>
42. Faktorová D, Nisbet RER, Fernández Robledo JA, Casacuberta E, Sudek L, Allen AE, Ares M Jr, Aresté C, Balestreri C, Barbrook AC, et al. 2020. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat Methods* 17:481–494. <https://doi.org/10.1038/s41592-020-0796-x>
43. Faktorová D, Kaur B, Valach M, Graf L, Benz C, Burger G, Lukeš J. 2020. Targeted integration by homologous recombination enables in situ tagging and replacement of genes in the marine microeukaryote *Diplonema papillatum*. *Environ Microbiol* 22:3660–3670. <https://doi.org/10.1111/1462-2920.15130>
44. Faktorová D, Záhonová K, Benz C, Dacks JB, Field MC, Lukeš J. 2023. Functional differentiation of Sec13 paralogues in the euglenozoan protists. *Open Biol* 13:220364. <https://doi.org/10.1098/rsob.220364>
45. Valach M, Benz C, Aguilar LC, Gahura O, Faktorová D, Zíková A, Oeffinger M, Burger G, Gray MW, Lukeš J. 2023. Miniature RNAs are embedded in an exceptionally protein-rich mitoribosome via an elaborate assembly pathway. *Nucleic Acids Res* 51:6443–6460. <https://doi.org/10.1093/nar/gkad422>
46. Valach M, Moreira S, Petitjean C, Benz C, Butenko A, Flegontova O, Nenarokova A, Prokopchuk G, Batstone T, Lapébie P, Lemogo L, Sarrasin M, Stretenowich P, Tripathi P, Yazaki E, Nara T, Henrissat B, Lang BF, Gray MW, Williams TA, Lukeš J, Burger G. 2023. Recent expansion of metabolic versatility in *Diplonema papillatum*, the model species of a highly speciose group of marine eukaryotes. *BMC Biol* 21:99. <https://doi.org/10.1186/s12915-023-01563-9>
47. Škodová-Sveráková I, Záhonová K, Juricová V, Danchenko M, Moos M, Baráth P, Prokopchuk G, Butenko A, Lukáčová V, Kohútová L, Bučková B, Horák A, Faktorová D, Horváth A, Šimek P, Lukeš J. 2021. Highly flexible metabolism of the marine euglenozoan protist *Diplonema papillatum*. *BMC Biol* 19:251. <https://doi.org/10.1186/s12915-021-01186-y>
48. Wells JN, Marsh JA. 2019. A graph-based approach for detecting sequence homology in highly diverged repeat protein families, p 251–261. In Sikosek T (ed), *Computational Methods in Protein Evolution*. Springer, New York, NY.
49. Schou KB, Andersen JS, Pedersen LB. 2014. A divergent calponin homology (NN-CH) domain defines a novel family: implications for evolution of ciliary IFT complex B proteins. *Bioinformatics* 30:899–902. <https://doi.org/10.1093/bioinformatics/btt661>
50. Holden JM, Koreny L, Obado S, Ratushny AV, Chen W-M, Chiang J-H, Kelly S, Chait BT, Aitchison JD, Rout MP, Field MC. 2014. Nuclear pore complex evolution: a trypanosome Mlp analogue functions in chromosomal segregation but lacks transcriptional barrier activity. *Mol Biol Cell* 25:1421–1436. <https://doi.org/10.1091/mbc.E13-12-0750>
51. Ballmer D, Carter W, van Hooff JJE, Tromer EC, Ishii M, Ludzia P, Akiyoshi B. 2024. Kinetoplastid kinetochore proteins KKT14-KKT15 are divergent Bub1/BubR1-Bub3 proteins. *Open Biol* 14:240025. <https://doi.org/10.1098/rsob.240025>
52. van Wijk LM, Snel B. 2020. The first eukaryotic kinome tree illuminates the dynamic history of present-day kinases. *bioRxiv*. <https://doi.org/10.1101/2020.01.27.920793>
53. Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. 2015. Bacterial proteins pinpoint a single eukaryotic root. *Proc Natl Acad Sci U S A* 112:E693–9. <https://doi.org/10.1073/pnas.1420657112>
54. Stok C, Tsaridou S, van den Tempel N, Everts M, Wierenga E, Bakker FJ, Kok Y, Alves IT, Jae LT, Raas MWD, et al. 2023. FIRR/CM1orf112 is synthetic lethal with PICH and mediates RAD51 dynamics. *Cell Rep* 42:112668. <https://doi.org/10.1016/j.celrep.2023.112668>
55. van Rooijen LE, Tromer EC, van Hooff JJE, Kops GJPL, Snel B. 2023. Increased sampling and intracomplex homologies favor vertical over horizontal inheritance of the Dam1 complex. *Genome Biol Evol* 15:evad017. <https://doi.org/10.1093/gbe/evad017>
56. Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. *Trends Ecol Evol* 35:43–55. <https://doi.org/10.1016/j.tree.2019.08.008>
57. Mansfeld J, Collin P, Collins MO, Choudhary JS, Pines J. 2011. APC15 drives the turnover of MCC-CDC20 to make the spindle assembly checkpoint responsive to kinetochore attachment. *Nat Cell Biol* 13:1234–1243. <https://doi.org/10.1038/ncb2347>
58. Osman F, Whitby MC. 2013. Emerging roles for centromere-associated proteins in DNA repair and genetic recombination. *Biochem Soc Trans* 41:1726–1730. <https://doi.org/10.1042/BST20130200>
59. Malik HS, Henikoff S. 2003. Phylogenomics of the nucleosome. *Nat Struct Biol* 10:882–891. <https://doi.org/10.1038/nsb996>
60. Luo X, Tang Z, Rizo J, Yu H. 2002. The Mad2 spindle checkpoint protein undergoes similar major conformational changes upon binding to either Mad1 or Cdc20. *Mol Cell* 9:59–71. [https://doi.org/10.1016/S1097-2765\(01\)00435-x](https://doi.org/10.1016/S1097-2765(01)00435-x)
61. Aravind L, Koonin EV. 1998. The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair. *Trends Biochem Sci* 23:284–286. [https://doi.org/10.1016/S0968-0004\(98\)01257-2](https://doi.org/10.1016/S0968-0004(98)01257-2)
62. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>
63. Carmena M, Wheelock M, Funabiki H, Earnshaw WC. 2012. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat Rev Mol Cell Biol* 13:789–803. <https://doi.org/10.1038/nrm3474>
64. Li Z, Lee JH, Chu F, Burlingame AL, Günzl A, Wang CC. 2008. Identification of a novel chromosomal passenger complex and its unique localization during cytokinesis in *Trypanosoma brucei*. *PLoS ONE* 3:e2354. <https://doi.org/10.1371/journal.pone.0002354>
65. Ballmer D, Akiyoshi B. 2024. Dynamic localization of the chromosomal passenger complex in trypanosomes is controlled by the orphan kinesins KIN-A and KIN-B. *eLife* 13:RP93522. <https://doi.org/10.7554/eLife.93522>
66. Adams RR, Wheatley SP, Gouldsworthy AM, Kandels-Lewis SE, Carmena M, Smythe C, Gerloff DL, Earnshaw WC. 2000. INCENP binds the Aurora-related kinase AIRK2 and is required to target it to chromosomes, the central spindle and cleavage furrow. *Curr Biol* 10:1075–1078. [https://doi.org/10.1016/S0960-9822\(00\)00673-4](https://doi.org/10.1016/S0960-9822(00)00673-4)
67. Sessa F, Mapelli M, Ciferri C, Tarricone C, Arecas LB, Schneider TR, Stukenberg PT, Musacchio A. 2005. Mechanism of Aurora B activation by INCENP and inhibition by hesperadin. *Mol Cell* 18:379–391. <https://doi.org/10.1016/j.molcel.2005.03.031>
68. Komaki S, Tromer EC, De Jaeger G, De Winne N, Heese M, Schnittger A. 2022. Molecular convergence by differential domain acquisition is a hallmark of chromosomal passenger complex evolution. *Proc Natl Acad Sci U S A* 119:e2200108119. <https://doi.org/10.1073/pnas.2200108119>
69. Zickler D, Kleckner N. 2023. Meiosis: dances between homologs. *Annu Rev Genet* 57:1–63. <https://doi.org/10.1146/annurev-genet-061323-044915>
70. Ishii M, Akiyoshi B. 2020. Characterization of unconventional kinetochore kinases KKT10 and KKT19 in *Trypanosoma brucei*. *J Cell Sci* 133:jcs240978. <https://doi.org/10.1242/jcs.240978>
71. Hardie DG. 1999. Plant protein serine/threonine kinases: classification and functions. *Annu Rev Plant Physiol Plant Mol Biol* 50:97–131. <https://doi.org/10.1146/annurev.arplant.50.1.97>
72. Lindberg MF, Meijer L. 2021. Dual-specificity, tyrosine phosphorylation-regulated kinases (DYRKs) and cdc2-like kinases (CLKs) in human disease, an overview. *Int J Mol Sci* 22:6047. <https://doi.org/10.3390/ijms22116047>
73. Garcia-Silva M-R, Sollelis L, MacPherson CR, Stanojčić S, Kuk N, Crobu L, Bringaud F, Bastien P, Pagès M, Scherf A, Sterkers Y. 2017. Identification of the centromeres of *Leishmania major*: revealing the hidden pieces. *EMBO Rep* 18:1968–1977. <https://doi.org/10.15252/embr.2017144216>
74. Geoghegan V, Carnielli JBT, Jones NG, Saldivia M, Antoniou S, Hughes C, Neish R, Dowle A, Mottram JC. 2022. CLK1/CLK2-driven signalling at the *Leishmania* kinetochore is captured by spatially referenced proximity phosphoproteomics. *Commun Biol* 5:1305. <https://doi.org/10.1038/s42003-022-04280-1>

75. Porter D. 1973. *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electronmicroscopic study. *J Protozool* 20:351–356. <https://doi.org/10.1111/j.1550-7408.1973.tb00895.x>
76. Triemer RE. 1992. Ultrastructure of mitosis in *Diplonema ambulator* Larsen and Patterson (euglenozoa). *Eur J Protistol* 28:398–404. [https://doi.org/10.1016/S0932-4739\(11\)80003-9](https://doi.org/10.1016/S0932-4739(11)80003-9)
77. Lowell JE, Cross GAM. 2004. A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. *J Cell Sci* 117:5937–5947. <https://doi.org/10.1242/jcs.01515>
78. Alfieri C, Chang L, Barford D. 2018. Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13. *Nature* 559:274–278. <https://doi.org/10.1038/s41586-018-0281-1>
79. Butterfield ER, Obado SO, Scutts SR, Zhang W, Chait BT, Rout MP, Field MC. 2024. A lineage-specific protein network at the trypanosome nuclear envelope. *Nucleus* 15:2310452. <https://doi.org/10.1080/19491034.2024.2310452>
80. Ebenezer TE, Zoltner M, Burrell A, Nenarokova A, Novák Vanclová AMG, Prasad B, Soukal P, Santana-Molina C, O'Neill E, Nankissoor NN, Vadakedath N, Daiker V, Obado S, Silva-Pereira S, Jackson AP, Devos DP, Lukeš J, Lebert M, Vaughan S, Hampl V, Carrington M, Ginger ML, Dacks JB, Kelly S, Field MC. 2019. Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol* 17:11. <https://doi.org/10.1186/s12915-019-0626-8>
81. Pastor F, Shkreta L, Chabot B, Durantel D, Salvetti A. 2021. Interplay between CMGC kinases targeting SR proteins and viral replication: splicing and beyond. *Front Microbiol* 12:658721. <https://doi.org/10.3389/fmicb.2021.658721>
82. Lukeš J, Speijer D, Ziková A, Alfonso JD, Hashimi H, Field MC. 2023. Trypanosomes as a magnifying glass for cell and molecular biology. *Trends Parasitol* 39:902–912. <https://doi.org/10.1016/j.pt.2023.08.004>
83. Kaur B, Záhonová K, Valach M, Faktorová D, Prokopchuk G, Burger G, Lukeš J. 2020. Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplomemids. *Nucleic Acids Res* 48:2694–2708. <https://doi.org/10.1093/nar/gkz1215>
84. Valach M, Moreira S, Hoffmann S, Stadler PF, Burger G. 2017. Keeping it complicated: mitochondrial genome plasticity across diplomemids. *Sci Rep* 7:14166. <https://doi.org/10.1038/s41598-017-14286-z>
85. George EE, Tashyreva D, Kwong WK, Okamoto N, Horák A, Husnik F, Lukeš J, Keeling PJ. 2022. Gene transfer agents in bacterial endosymbionts of microbial eukaryotes. *Genome Biol Evol* 14:evac099. <https://doi.org/10.1093/gbe/evac099>
86. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <https://doi.org/10.1038/nprot.2013.084>
87. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682. <https://doi.org/10.1093/bioinformatics/btq003>
88. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
89. Wheeler RJ. 2021. A resource for improved predictions of *Trypanosoma* and *Leishmania* protein three-dimensional structure. *PLoS ONE* 16:e0259871. <https://doi.org/10.1371/journal.pone.0259871>
90. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 35:1026–1028. <https://doi.org/10.1038/nbt.3988>
91. Richter DJ, Berney C, Strasser JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, de Vargas C. 2022. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer J* 2:e56. <https://doi.org/10.24072/peerjournal.173>
92. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* 19:679–682. <https://doi.org/10.1038/s41592-022-01488-1>
93. Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>
94. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>
95. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
96. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20:473. <https://doi.org/10.1186/s12859-019-3019-7>
97. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
98. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 42:243–246. <https://doi.org/10.1038/s41587-023-01773-0>
99. de Potter B, Raas MWD, Seidl MF, Verrijzer CP, Snel B. 2023. Uncoupled evolution of the Polycomb system and deep origin of non-canonical PRC1. *Commun Biol* 6:1144. <https://doi.org/10.1038/s42003-023-05501-x>
100. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
101. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
102. Masuda T, Tomita M, Ishihama Y. 2008. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* 7:731–740. <https://doi.org/10.1021/pr700658q>
103. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372. <https://doi.org/10.1038/nbt.1511>
104. Tyanova S, Cox J. 2018. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research, p 133–148. In Stechow L (ed), *Cancer systems biology: methods and protocols*. Springer, New York, NY.
105. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaino JA. 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50:D543–D552. <https://doi.org/10.1093/nar/gkab1038>