

Research



**Cite this article:** 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>

Received: 28 February 2021  
Accepted: 16 April 2021

**Subject Area:**  
cellular biology/microbiology/molecular biology/bioinformatics

**Keywords:**  
kinetochore, Kinetoplastida, synaptonemal complex, *Trypanosoma brucei*, recurrent evolution, SYCP2

**Authors for correspondence:**  
Eelco C. Tromer  
e-mail: e.c.tromer@rug.nl  
Bungo Akiyoshi  
e-mail: bungo.akiyoshi@bioch.ox.ac.uk

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5426780>.

# Repurposing of synaptonemal complex proteins for kinetochores in Kinetoplastida

Eelco C. Tromer<sup>1,2</sup>, Thomas A. Wemyss<sup>1</sup>, Patryk Ludzia<sup>3</sup>, Ross F. Waller<sup>1</sup> and Bungo Akiyoshi<sup>3</sup>

<sup>1</sup>Department of Biochemistry, University of Cambridge, Cambridge, UK

<sup>2</sup>Cell Biochemistry, Groningen Institute of Biomolecular Sciences & Biotechnology, University of Groningen, Groningen, The Netherlands

<sup>3</sup>Department of Biochemistry, University of Oxford, Oxford, UK

**ORCID** ECT, 0000-0003-3540-7727; TAW, 0000-0002-9762-5407; PL, 0000-0002-1678-6875; RFW, 0000-0001-6961-9344; BA, 0000-0001-6010-394X

Chromosome segregation in eukaryotes is driven by the kinetochore, a macromolecular complex that connects centromeric DNA to microtubules of the spindle apparatus. Kinetochores in well-studied model eukaryotes consist of a core set of proteins that are broadly conserved among distant eukaryotic phyla. By contrast, unicellular flagellates of the class Kinetoplastida have a unique set of 36 kinetochore components. The evolutionary origin and history of these kinetochores remain unknown. Here, we report evidence of homology between axial element components of the synaptonemal complex and three kinetoplastid kinetochore proteins KKT16-18. The synaptonemal complex is a zipper-like structure that assembles between homologous chromosomes during meiosis to promote recombination. By using sensitive homology detection protocols, we identify divergent orthologues of KKT16-18 in most eukaryotic supergroups, including experimentally established chromosomal axis components, such as Red1 and Rec10 in budding and fission yeast, ASY3-4 in plants and SYCP2-3 in vertebrates. Furthermore, we found 12 recurrent duplications within this ancient eukaryotic SYCP<sup>2-3</sup> gene family, providing opportunities for new functional complexes to arise, including KKT16-18 in the kinetoplastid parasite *Trypanosoma brucei*. We propose the kinetoplastid kinetochore system evolved by repurposing meiotic components of the chromosome synapsis and homologous recombination machinery that were already present in early eukaryotes.

## 1. Introduction

Chromosome segregation in eukaryotes is driven by spindle microtubules and kinetochores. Microtubules are dynamic polymers that consist of  $\alpha$ -/ $\beta$ -tubulin subunits, while the kinetochore is the macromolecular protein complex that assembles onto the centromeric DNA and interacts with spindle microtubules during mitosis and meiosis [1]. All studied eukaryotes use spindle microtubules to drive the chromosome movement, and  $\alpha$ -/ $\beta$ -tubulins are among the most conserved proteins in eukaryotes [2]. The kinetochore is a highly complicated structure that consists of more than 30 unique structural proteins even in a relatively simple budding yeast kinetochore [3]. CENP-A is a centromere-specific histone H3 variant that specifies kinetochore assembly sites, while the NDC80 complex (NDC80, NUF2, SPC24 and SPC25) constitutes the primary microtubule-binding activity of kinetochores [4]. Functional studies have established that CENP-A and the NDC80 complex are essential for the kinetochore function in several model eukaryotes (e.g. yeasts, worms, flies and humans) [3]. However, CENP-A is absent in some eukaryotic lineages such as holocentric insects [5], early-diverging Fungi [6] and Kinetoplastida [7]. Furthermore, it is

known that compositions of kinetochores can vary considerably among eukaryotes [8] and that these components are highly divergent at the sequence level [8–10]. The most extreme case known to date is found in Kinetoplastida, for which no apparent direct homologues of the canonical kinetochore proteins were detected [11].

Kinetoplastida comprises unicellular flagellates characterized by the presence of a conspicuous mitochondrial structure called the kinetoplast that contains a unique form of mitochondrial DNA [12]. This group belongs to the phylum Euglenozoa and is evolutionarily divergent from many popular fungal and animal model eukaryotes (all belonging to Opisthokonta) [13]. There are several medically important kinetoplastid parasites such as *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* spp., which cause neglected tropical diseases [14]. Although chromosome segregation depends on spindle microtubules in *T. brucei* [15], we previously identified 24 unique kinetoplastid kinetochore proteins (KKT1–20 and KKT22–25) that localize at centromeres in this species, which together constitute a functionally analogous kinetochore structure [16–18]. Furthermore, 12 additional proteins, namely KKT-interacting proteins (KKIP1–12), have been identified, which associated with kinetoplastid kinetochores during mitosis [19,20]. None of these 36 kinetoplastid kinetochore subunits appears to be clearly orthologous to canonical kinetochore proteins, and whether there is an evolutionary relationship between the two kinetochore systems remains unclear [21]. Interestingly, although there is a significant distance between the sister kinetochores in metaphase cells of non-kinetoplastid species (the space is called the inner centromere) [22–24], there is no clear separation between sister kinetochores in all studied kinetoplastids [25–28]. This structural difference is consistent with compositional differences between canonical and kinetoplastid kinetochores.

Tromer *et al.* traced the origin of canonical kinetochore subunits back to before the last eukaryotic common ancestor (LECA) using a combination of phylogenetic trees, profile-versus-profile homology detection and structural comparisons of its protein components [29]. They found that duplications played a major role in shaping the ancestral eukaryotic kinetochore and that its components share a deep evolutionary history with proteins of various other prokaryotic and eukaryotic pathways, e.g. ubiquitination, DNA damage repair and the flagellum [29]. In these analyses, none of the 36 KKT/KKIP proteins of the kinetoplastid kinetochore was considered because they were found only in Kinetoplastida and were therefore deemed not to have been part of the kinetochore in LECA. If KKT/KKIP proteins were not part of the kinetochore in ancestral eukaryotes, then when and from where did these new components of the kinetochore originate? (i) They may have a mosaic origin similar to the LECA kinetochore [29] and might have been pieced together and co-opted from other processes, (ii) they may have arisen from external sources (e.g. viral integrations into the genome, bacterial endosymbionts), or (iii) they may have arisen through a combination of genomic translocations, fusion of existing genes and de novo gene birth in the first ancestors of Kinetoplastida. (iv) An alternative hypothesis is that KKT/KKIP subunits might be canonical kinetochore proteins that diverged to such an extent that they cannot be readily identified through sequence searches. Indeed, a recent study found intriguing similarities between the coiled coils of the outer kinetochore protein KKIP1 and those found in the C-terminal half of two subunits of the microtubule-binding NDC80

complex (NDC80 and NUF2) [19]. However, since no sequence similarity could be detected between KKIP1 and the N-terminal Calponin homology domains of NDC80/NUF2, it remains unclear whether KKIP1 and NUF2/NDC80 are truly homologous or whether the similarity of their coiled coils is merely the result of convergent evolution. In addition, Aurora kinase and INCENP, two members of the chromosome passenger complex that localize at the inner centromere region in other eukaryotes, are found in kinetoplastids [30], suggesting that some processes or components of canonical chromosome segregation systems are conserved in kinetoplastids as well. Detailed phylogenetic analyses and sensitive homology searches of KKT/KKIP proteins will need to be performed to evaluate which of the aforementioned evolutionary scenarios applies to the kinetoplastid kinetochore.

Previous studies identified several domains in KKT/KKIP proteins that are commonly found among both eukaryotic and prokaryotic proteins, providing some clues to the functionality and evolutionary history of the proteins that make up the kinetoplastid kinetochore. These include a BRCA1 carboxy-terminal (BRCT) domain in KKT4, a forkhead-associated (FHA) domain in KKT13, a seven-bladed WD40  $\beta$ -propeller in KKT15, a Gcn5-related *N*-acetyltransferase domain in KKT23, a protein kinase domain of unknown affiliation in KKT2 and KKT3, a Cdc2-like kinase domain in KKT10 and KKT19 and large coiled coil regions (e.g. KKIP1 and KKT24) [16–19]. BRCT and FHA domains are frequently found in proteins involved in the DNA damage response [31], which is important not only for DNA damage repair in somatic cells but also for homologous recombination in meiotic cells [32]. Furthermore, it has been proposed, based on the similarity in the C-terminal polo boxes, that KKT2, KKT3 and KKT20 share common ancestry with Polo-like kinases (PLKs), which regulate various biological functions such as the cell cycle progression, the DNA damage response, kinetochores, centrosomes and synaptonemal complexes (SCs) [31,33]. PLKs localize at kinetochores in some organisms but are not thought to be a structural kinetochore component [34,35]. Although the kinase domains of KKT2 and KKT3 have been classified as unique among eukaryotic kinase subfamilies [36], they have some similarity to the kinase domain of PLKs [21]. KKT2 and KKT3 also have a unique zinc-binding domain in the central region, which promotes their centromere localization [37]. Gene duplication plays an important role in generating new functions using pre-existing proteins [38], and this process has further contributed to the kinetoplastid kinetochore development. The paralogues KKT2, KKT3 and KKT20, and KKT10 and KKT19 have arisen through gene duplications, and this also appears to be the case for KKT17 and KKT18 [16]. Although KKT17 and KKT18 are conserved even in distantly related kinetoplastids [39], previous studies did not reveal any recognizable domains in these proteins [16].

While the identification of conserved domains in kinetoplastid kinetochore proteins offers some insight into the molecular mechanism or function for these proteins, there has been previously no direct link to any pre-existing machinery from which it might be derived. In this study, we report evidence of homology between the KKT16–18 complex and the axial element components (also known as lateral elements) of the SC, a meiosis-specific tripartite structure that assembles between homologous chromosomes [40]. We speculate on the implications of this finding for the origin of kinetoplastid kinetochores.

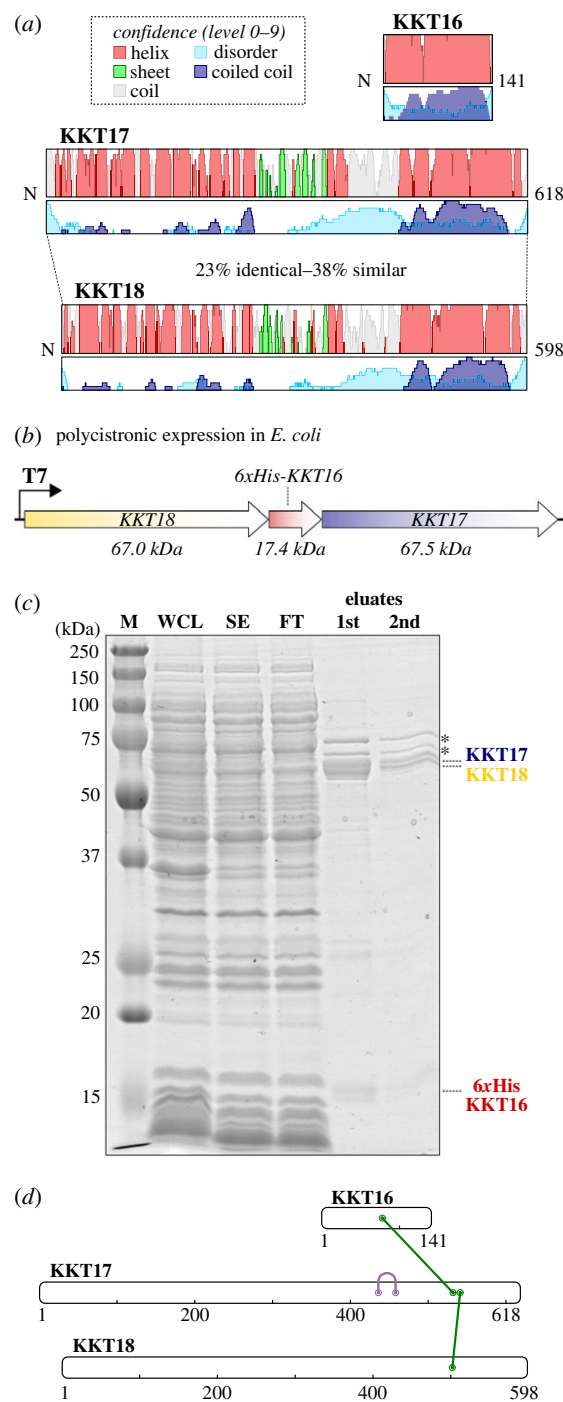
## 2. Results

### 2.1. KKT16–KKT18 form the KKT16 complex

Our previous immunoprecipitation and mass spectrometry of KKT proteins suggested that KKT16 (Tb927.11.1000), KKT17 (Tb927.3.2330) and KKT18 (Tb927.9.3800) probably form a sub-complex [16]. Consistent with this possibility, these proteins have a similar localization pattern during the cell cycle, showing diffuse nuclear signals in G1 and forming kinetochore dots from S phase to anaphase in *T. brucei* [16]. KKT17 and KKT18 have a moderate degree of sequence identity and similarity to one another (23% shared identity and 38% similarity for *T. brucei* sequences), suggesting that these two proteins are the product of a duplication event [16]. Indeed, their predicted secondary structure revealed a highly similar topology, with an *N*-terminal globular region consisting of repeated alpha helices, followed by a beta sheet-rich domain, and a *C*-terminal coiled coil, connected by a disordered linker region (figure 1*a*). KKT16 consists of a single coiled coil region, signifying the coiled coils of KKT16, KKT17 and KKT18 as a potential basis for their association (figure 1*a*). To test whether KKT16–KKT18 interact with each other, we expressed these three proteins in bacteria using a polycistronic expression system [42] (figure 1*b*). We found that KKT17 and KKT18 co-purified with 6xHis-KKT16 (figure 1*c*), showing that these three proteins on their own are sufficient to form a complex. Chemical cross-linking mass spectrometry of this sample using bis(sulfo-succinimidyl)suberate (BS<sup>3</sup>, a homo-bifunctional cross-linker that covalently links pairs of lysines that are within 26–30 Å on the protein surface) identified a cross-link between KKT16 (residue K79) and KKT17 (K532) as well as that between KKT17 (K541) and KKT18 (K502) (figure 1*d*). Together with our previous data showing that KKT17 was one of the most abundant proteins in immunoprecipitates of KKT18 and vice versa [16], these results suggest that KKT16 interacts with KKT17 and KKT18 simultaneously. We refer to these three interacting proteins as the KKT16 complex.

### 2.2. KKT16–KKT18 are similar to axial element components of the synaptonemal complex

To identify potential homologues of the KKT16 complex proteins and to detect proteins or protein complexes with similar domain architectures, we used hidden Markov modelling methods (see §5). We generated multiple-sequence alignments (MSAs) of previously detected homologues of KKT16–18 found among distantly related (non-parasitic) kinetoplastids [16,39]. We then constructed profile hidden Markov models (HMM) based on these MSAs and performed secondary structure-aware profile-versus-profile HMM searches against databases of known conserved domains and structures (see §5; figures 2*a* and 3*a*) using HHsearch [45]. Because KKT17 and KKT18 appeared to be paralogues, we included all kinetoplastid homologues of these proteins into a single MSA, aiming to increase the similarity detection sensitivity. By using this approach, we found that KKT17 and KKT18 consist of three domains frequently found in eukaryotic proteins (figure 2*a*; see electronic supplementary material, file S1 for output files): (i) an *N*-terminal Armadillo repeat region (ARM, many high probability homologues: Prob > 80%, *E*-value ~ 1), followed by (ii) a  $\beta$ -barrel Pleckstrin homology (PH) domain found in chromatin



**Figure 1.** KKT16, KKT17 and KKT18 form the KKT16 complex. (a) Cartoon of the KKT16, KKT17 and KKT18 proteins, including secondary structure predictions based on multiple-sequence alignments of kinetoplastid homologues with *T. brucei* KKT proteins as a query. Height of each track indicates the confidence of each prediction (§5). Confidence levels are discretized into 10 levels (0–9). Identity percentages for *T. brucei* KKT17 and KKT18 were calculated using the Needleman–Wunsch global alignment algorithm, with the BLOSUM62 matrix to derive their similarity score [41]. (b) Schematic of the construct used to co-express the three sub-units. See electronic supplementary material, figure S1 for expression tests. (c) Expression and purification of the KKT16 complex from *E. coli*. Coomassie-stained 12% acrylamide gel of the purification is shown. Asterisks indicate common contaminants; M: marker; WCL: whole cell lysate; SE: soluble extract; FT: flow through. (d) Chemical crosslinking mass spectrometry of the KKT16 complex using BS<sup>3</sup>. The green lines indicate pairs of inter-molecule crosslinks, while the purple line indicates an intra-molecule crosslink. The list of identified crosslinked peptides is shown in electronic supplementary material, table S1.

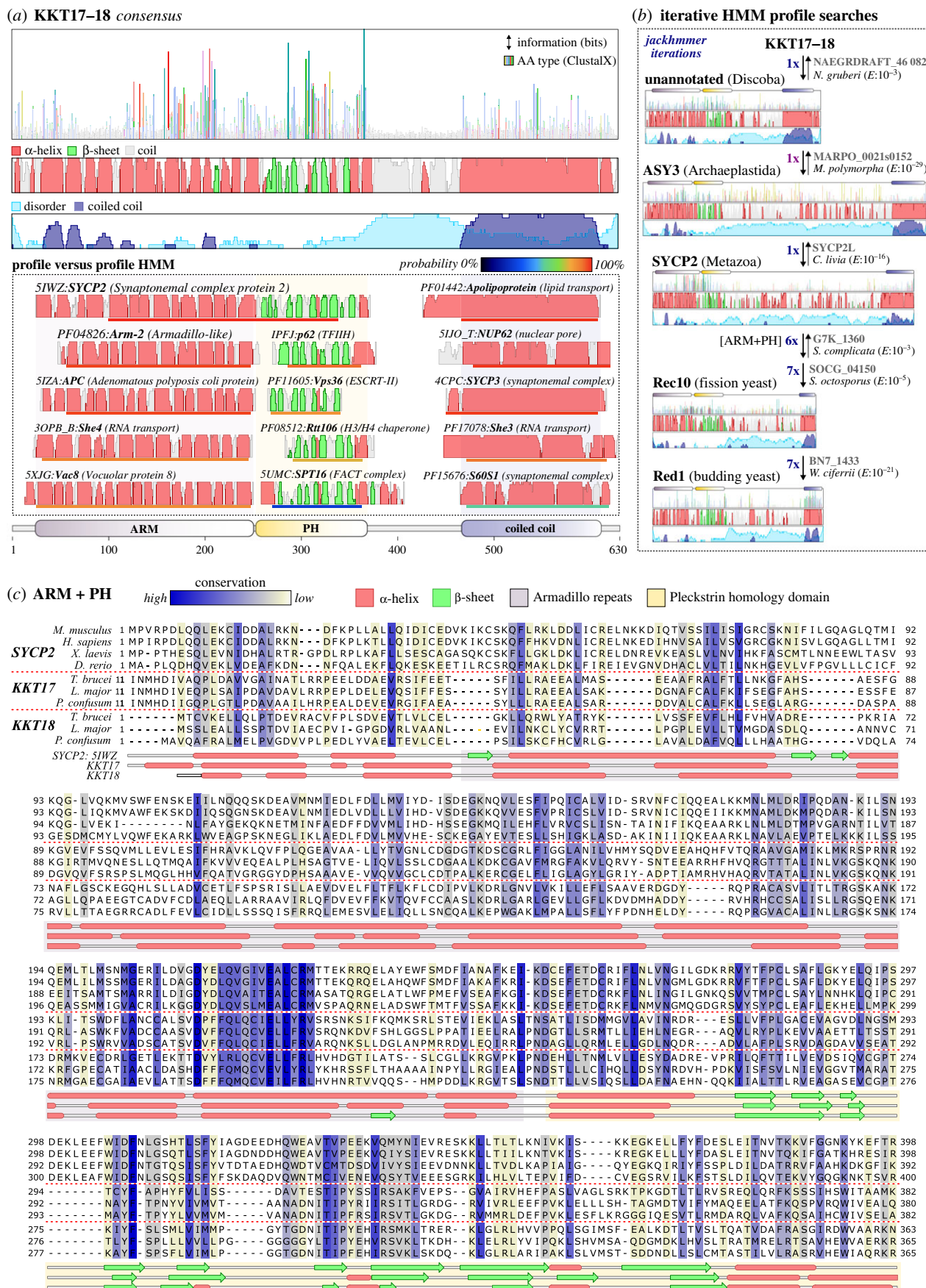


Figure 2. (Caption overleaf.)

(e.g. histone chaperone Rtt106, pfam-A: PF08512, Prob: 75%,  $E \sim 10$ ) and membrane-associated proteins (e.g. ESCRT complex protein Vps36, pfam-A: PF11605, Prob: 84%,  $E \sim 10$ ), which are linked by a structurally disordered region to (iii) a C-terminal coiled coil that has similarity to many higher-order-forming

coiled coil proteins, with the extracellular lipid-binding apolipo-proteins as strongest scoring HMM profile (APO, pfam-A: PF01442, Prob: 97%,  $E \sim 10^{-1}$ ). Searches with the HMM profile of KKT16 revealed similarity with a large number of coiled coil components of various eukaryotic protein complexes (figure 3a;

**Figure 2.** (*Overleaf.*) KKT17 and KKT18 are orthologues of the eukaryotic SC proteins SYCP2, ASY3, Red1 and Rec10. (a) Cartoon of the conserved domain architecture and sequence features of a consensus of homologues of KKT17 and KKT18 (see electronic supplementary material, file S3 for MSA, and electronic supplementary material, file S6 for details of the sequence annotation). From top to bottom: (1) collapsed sequence logo with different conserved amino acids for each position represented by the ClustalX colouring scheme. Height indicates the bits of information per position and is used as a proxy for conservation of amino acids at given positions; (2) secondary structure prediction, similar to figure 1. Height indicates the probability of the prediction for each position (level 0–9); (3) overview of selected results (best scoring hits—top, generic hits—towards the bottom) from a profile-versus-profile HMM search (HHsearch, see S5) of KKT17 and KKT18 against the PFAM (conserved domains) and PDB (3D structures) database. Identifiers starting with ‘PF’ indicate PFAM domains and four digit identifiers (e.g. 5IWZ) indicate PDB entries. Terms between parentheses indicate relevant functional or domain annotation. For each hit, only a proportion of the total protein/domain is shown around the region that has a significant similarity with the KKT17 and KKT18 HMM. Coloured bars below each of the proteins/domains indicate the HHsearch probability score; and (4) topology diagram of conserved domains in KKT17 and KKT18; ARM: Armadillo repeats; PH: Pleckstrin homology domain. (b) Cartoon of the similarity detection protocol (S5) for establishing homology between highly divergent synaptonemal complex proteins KKT17, KKT18, SYCP2, ASY3, Red1 and Rec10. Similarity detection does not necessarily have to be performed in this order, but has been visualized in a linear manner to showcase the searchpath from KKT17 and KKT18 to other eukaryotic SYCP2-type proteins. The seed HMM for KKT17 and KKT18 is based on the same multiple-sequence alignment as shown in (a). Thick arrows and *E*-values indicate the direction and the significance of the similarity searches. Thin arrows in the reverse direction indicate that reciprocal searches yield similar homologous connections. Dark purple numbers indicate the number of iterations of HMM searches needed to include a particular sequence. Uniprot identifiers (grey) on the right indicate the highest scoring orthologues after each iteration belonging to a larger group of related sequences (e.g. SYCP2-type proteins in Metazoa) for which separate conservation and architecture cartoons are presented as in (a) (see electronic supplementary material, file S3 for cartoons of all SYCP2-type HMMs). The positions of the ARM and PH domain and coiled coils are projected on top of the cartoons. Brackets indicate when only selected domains are used for similarity detection. Full species names: *Naegleria gruberi*, *Marchantia polymorpha*, *Columba livia*, *Saitoella complicata*, *Schizosaccharomyces octosporus*, *Wickerhamomyces ciferrii*. (c) Multiple-sequence alignment of the *N*-terminal ARM repeats and PH domain of SYCP2, KKT17 and KKT18 homologues (see for colour-coding settings *jalview* session files electronic supplementary material, file S5). Columns with single amino acid occupancy were removed. Secondary structure prediction by PSIPRED based on a multiple-sequence alignment of KKT17 and KKT18 using the *T. brucei* proteins as seed sequence (S5). The secondary structure of SYCP2 is mapped based on the PDB file 5IWZ (mouse SYCP2) [43]. Full species names: *Mus musculus*, *Homo sapiens*, *Xenopus laevis*, *Danio rerio*, *Trypanosoma brucei*, *Leishmania major*, *Paratrypanosoma confusum*.

see electronic supplementary material, file S1 for raw output files), similar to KKT17 and KKT18, and with apolipoprotein as the highest-scoring HMM (APOE, pdb: 2L7B, Prob: 95%,  $E \sim 1$ ).

Although the KKT16 complex subunits consist of generic domains, the specific ARM-PH topology of KKT17 and KKT18 was only detected in the HMM profile of one metazoan protein (figure 2a): the synaptonemal complex protein 2 (SYCP2, pdb:5IWZ [43], Prob > 99%,  $E \sim 10^{-5}$ ). In addition, the C-terminal coiled coil domain of KKT17 and KKT18 showed similarity to a known SYCP2 multimerization partner [46], the single coiled coil protein synaptonemal complex protein 3 (SYCP3, pdb:4CPC [44], Prob:95%,  $E \sim 1$ ). SYCP2:SYCP3 multimers constitute the axial elements of the SC, a zipper-like structure that forms the linkage between parental chromosomes to facilitate homologous recombination during meiosis [32,47,48]. The significant similarity ( $E < 10^{-2}$ ) and high HHsearch probability scores (Prob > 85%) suggested that the KKT16 complex and the axial elements of the SC shared a common ancestor, providing an important clue to the possible origin of kinetoplastid kinetochores.

### 2.3. KKT16–KKT18 are part of the highly divergent SYCP<sup>2–3</sup> gene family, including the SC components Red1, Rec10:Rec27, ASY3:ASY4 and SYCP2:SYCP3 found in fungi, Archaeplastida and Metazoa

Although structures of the SC are highly conserved among eukaryotes [40], previous sequence similarity searches using BLAST, mostly found homologues of SYCP2 and SYCP3 in Metazoa and failed to detect significant sequence similarity among the wide range of eukaryotes [49]. Based on our detection of similarity between the KKT16 complex in Kinetoplastida and the SYCP2:SYCP3 multimer in Metazoa, we sought to test for the

presence of homologues among diverse eukaryotes using a more sensitive HMM-based approach. For ease of reference, we use the term ‘SYCP<sup>2–3</sup>’ to indicate all genes similar to SYCP2 and SYCP3 and KKT16–KKT18. The term ‘SYCP2-type’ is used for those SYCP<sup>2–3</sup> genes with an ‘ARM-PH-coiled coil’ domain topology (e.g. KKT17 and SYCP2). ‘SYCP3-type’ refers to other coiled coil-only genes (e.g. KKT16 and SYCP3).

To exploit sequence diversity in our search for putative SYCP<sup>2–3</sup> homologues, we collated a large sequence database of 343 genomes and transcriptomes broadly sampled from the eukaryotic tree of life, but specifically focussed on including lineages more closely related to Kinetoplastida, such as Diplonemida, Euglenida and other species from the superphylum Discoba (for full species list see electronic supplementary material, table S2). By using an iterative HMM profile ‘hopping’ protocol for homology detection (see for detailed description S5), we identified candidate homologues throughout the eukaryotic tree of life (figure 2b,c and figure 3b,c; for overview see electronic supplementary material, table S2 and file S2). We found that significant similarities ( $E < 10^{-2}$ ) to the region spanning the last part of the ARM repeats and the full PH domain in combination with the presence of a C-terminal coiled coil were the best criteria to distinguish SYCP2-type homologues from other eukaryotic ARM repeat or PH domain-bearing proteins. Importantly, we established homology based on the significant sequence similarity between the KKT16 complex, SYCP2 and SYCP3 and previously identified SC proteins Red1 (*Saccharomyces cerevisiae*), Rec10 (*Schizosaccharomyces pombe*) and ASY3:ASY4 (*Arabidopsis thaliana*) [46,50–52] (see for examples of linear homology detection paths: figures 2b and 3b). Red1 and Rec10 have a highly divergent *N*-terminal ARM-PH domain, but lack significant similarity between their C-terminal coiled coils and those in other SYCP<sup>2–3</sup> genes (figure 2b). It is unclear whether coiled coils in these proteins evolved convergently or diverged beyond recognition. However, based on the finding that Red1

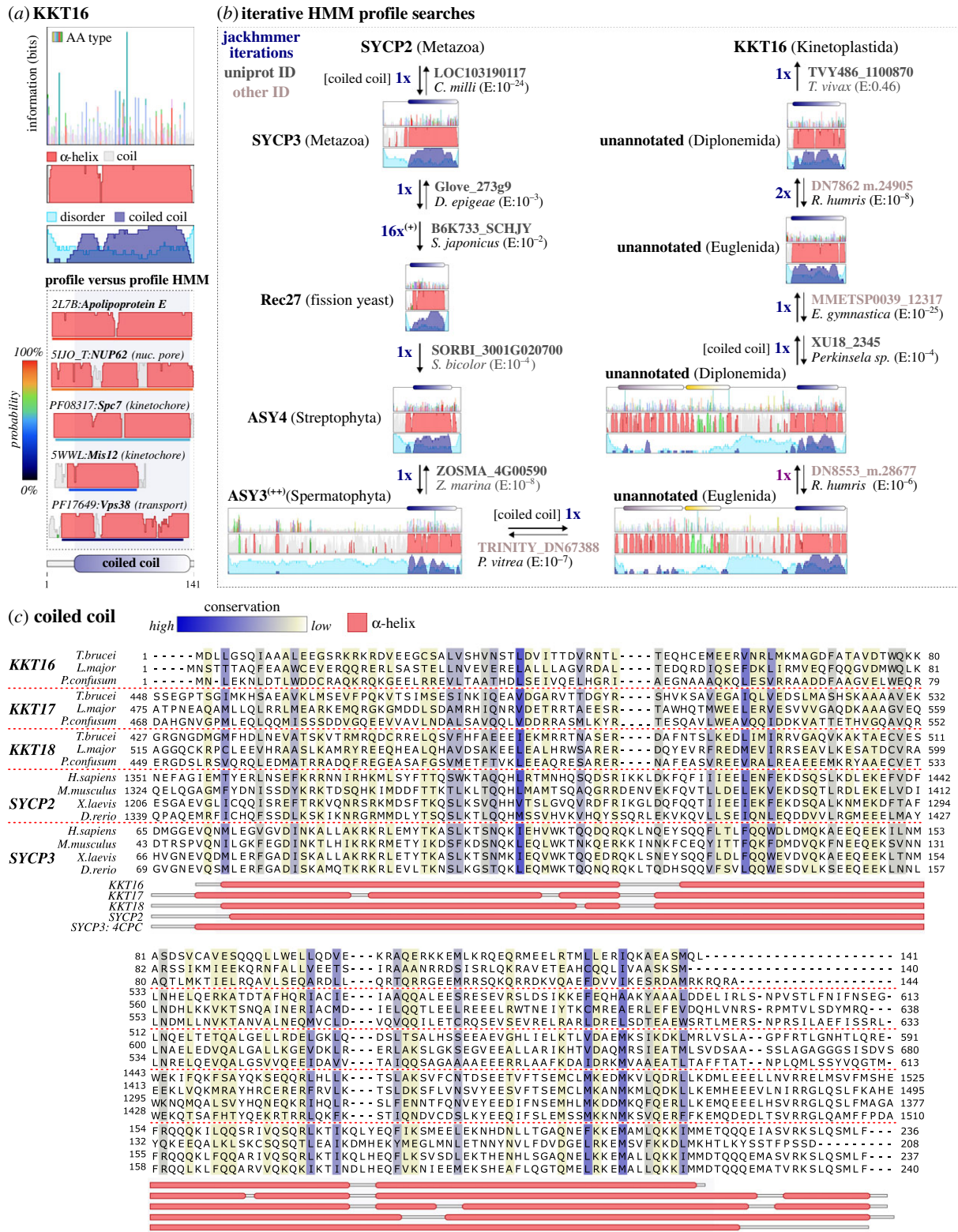


Figure 3. (Caption overleaf.)

and ASY3:ASY4 are structurally and functionally analogous to metazoan SYCP2:SYCP3 multimers [46], it is likely that these SC proteins share common ancestry with SYCP2 and SYCP3.

It is known that coiled coils are often unsuitable for sequence similarity searches and phylogenetic analyses due to a high degree of sequence redundancy. For example, searches with some clade-specific SYCP<sup>2-3</sup> HMM profiles (e.g. metazoan SYCP3 and KKT17 and KKT18) yielded homologues of the metazoan-specific extracellular apolipoproteins after several iterations (§5), which are unlikely to be

SYCP<sup>2-3</sup> homologues based on their known function and location [53]. In most other searches we performed, coiled coil domains of SYCP<sup>2-3</sup> genes were sufficient to identify homologues using reciprocal iterative searches. To restrict the inclusion of false-positive coiled coil proteins, we only considered candidates with bidirectional best similarity to (1) *bona fide* eukaryotic SC components (Red1, Rec10, Rec27, ASY3, ASY4, SYCP2 and SYCP3), (2) KKT16 complex subunits, and/or (3) an SYCP2- or SYCP3-type gene of the same clade when a candidate for both is present (e.g. SYCP<sup>2-3</sup> genes in Diplonemida and Euglenida; figure 3b).

**Figure 3.** (*Overleaf.*) KKT16 is a coiled coil protein homologous to the C-terminus of SYCP2 and KKT17 and KKT18. (a) Cartoon of the conserved domain architecture and sequence features of KKT16. See figure 2a for explanation. The best scoring domains, similar to KKT16, are all part of coiled coil proteins, but no clear connection could be established between KKT16 and SYCP2-SYCP3 or any other synaptonemal complex protein using profile-versus-profile HMM searches. (b) Cartoon of the similarity detection protocol (see §5) for establishing homology between the coiled coils of highly divergent SYCP<sup>2-3</sup> proteins (SYCP2 and SYCP3; Red1, Rec10 and Rec27; ASY3 and ASY4 and KKT17 and KKT18) and KKT16. Similarity detection does not necessarily have to be performed in this order, but has been visualized in a linear manner to specifically showcase the long path towards establishing homology between the coiled coils of SYCP2 and SYCP3 proteins to KKT16 (only a 'grey zone hit' with *E*-value 0.46). The seed HMM for metazoan SYCP2 is the same as shown in figure 2b. Thick arrows and *E*-values indicate the direction and the significance of the similarity searches. Thin arrows in the reverse direction indicate that reciprocal searches yield similar homologous connections. Dark blue numbers indicate the number of iterations of HMM searches needed to include a particular sequence. Uniprot (dark grey) and non-Uniprot identifiers (lighter grey, searches performed on local sequence database; for sources see electronic supplementary material, table S2) indicate the highest scoring homologues after each iteration belonging to a larger phylogenetic group for which separate conservation and architecture cartoons are presented as shown in (a) (see electronic supplementary material, file S3 for cartoons of all SYCP<sup>2-3</sup> HMMs). The position of the ARM and PH domain and the coiled coil are projected on top of the cartoons (colours correspond to (a)). Brackets indicate when only selected domains are used for similarity detection. <sup>(+)</sup>HMM used to detect all SYCP2-type and SYCP3-type homologues in Opisthokonta based on 16 iterations (fungi and animals, see electronic supplementary material, file S3) <sup>(++)</sup>ASY3 homologues in Spermatophyta (e.g. in *A. thaliana*) lost the N-terminal ARM and PH domains. Full species names: *Callorhinchus milii*; *Diversispora epigaea*; *Schizosaccharomyces japonicus*; *Sorghum bicolor*; *Zostera marina*; *Ploeotia vitrea*; *Rhynchopus humris*; *Eutreptiella gymnastica*; *Trypanosoma vivax*. (c) Multiple-sequence alignment of the coiled coils of KKT16–KKT18 and SYCP2 and SYCP3 homologues. See figure 2c and electronic supplementary material, file S5 for further explanation on annotation and the conservation colouring scheme. The secondary structure of SYCP3 is mapped based on the PDB file 4CPC (human SYCP3) [44].

Interestingly, searches with HMM profiles of clade-specific SYCP2- and SYCP3-type genes often had overlapping candidate homologues in reciprocal searches, signifying potential duplications of their C-terminal coiled coils. For instance, closely overlapping candidate homologues were found between SYCP2 and SYCP3 in Metazoa and between ASY3 and ASY4 in Streptophyta (figure 3b). Similarly, we found that the *S. pombe* linear element (SC-like structure) components Rec10 and Rec27 probably share a common ancestor (figures 2b and 3b) [54]. KKT17 and KKT18 also showed evidence of homology to KKT16, albeit with borderline significance ( $0.01 < E < 1$ , figure 3b). The homology of the C-terminal coiled coils of SYCP<sup>2-3</sup> genes suggested that the KKT16-18 and Rec10:Rec27 complexes are likely based on the same coiled coil-mediated interactions found for SYCP2:SYCP3 (mammals), ASY3:ASY4 (plants) and Red1 (budding yeast) heteromultimers or homomultimers [46].

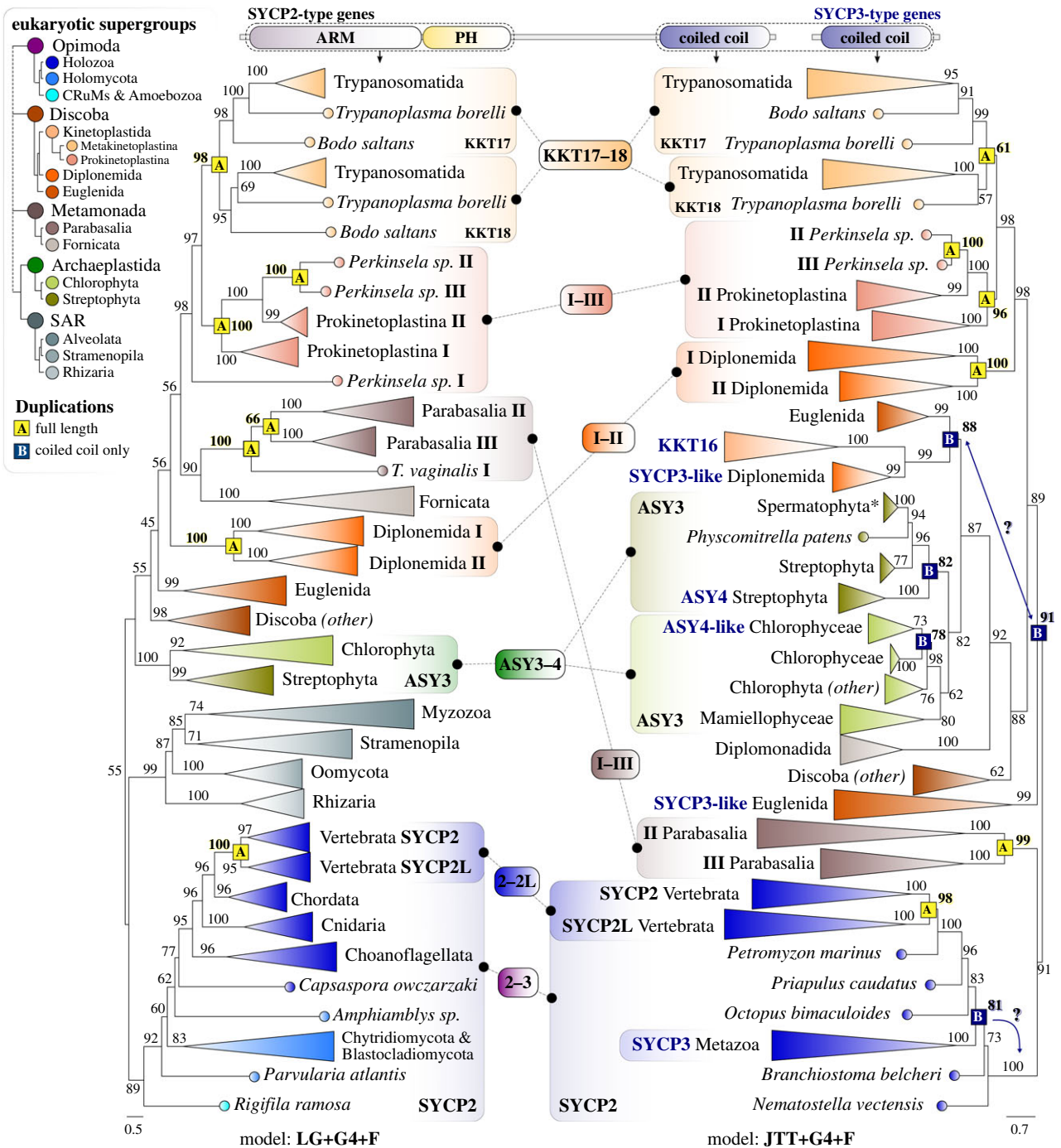
#### 2.4. The SYCP<sup>2-3</sup> gene family evolved through recurrent duplications

Our identification of different numbers of homologues of both SYCP2 and SYCP3-type proteins among distant eukaryotic lineages raised the question of what evolutionary events led to the diversification of the SYCP<sup>2-3</sup> gene family. How and when did these paralogues arise? Are they the result of lineage-specific duplication events or do they point to a more ancient origin for the SYCP2 and SYCP3-type genes? To answer these questions, we inferred separate phylogenetic trees for the N-terminal ARM-PH region of SYCP2-type genes and the coiled coil domains of all SYCP<sup>2-3</sup> gene family members. Due to the divergent nature of SYCP<sup>2-3</sup> homologues, we adopted a previously used alignment strategy [29] that generates a super-alignment of distantly related sequences through iteratively aligning increasingly less similar clade-specific MSAs (see §5). Some coiled coil domains of the SYCP<sup>2-3</sup> homologues were too divergent to yield MSAs of sufficient quality and were excluded from our analyses (e.g. the coiled coil of the SYCP<sup>2-3</sup> genes in Fungi and SAR supergroups; see §5). To subsequently infer phylogenetic trees, we used the maximum-likelihood phylogenomics software IQ-

Tree [55] (1000x Ultrafast bootstrap replicates, automated model selection; see §5). To resolve duplications, we reconciled the resulting phylogenetic trees of both the ARM-PH and coiled coil tree with the known eukaryotic species tree [56]. Although the highly divergent nature of the SYCP<sup>2-3</sup> sequences and the short length of their coiled coils precluded the faithful recovery of ancient patterns of eukaryotic evolution, they did provide evidence of more recent instances of gene duplication within well-resolved lineages (figure 4).

In total, we found evidence for seven independent duplications of full-length SYCP2-type homologues that were consistent between both ARM-PH and the coiled coil trees (see yellow A in figure 4). Of these, we found three recurrent duplications among Kinetoplastida. One duplication in the common ancestor of the subclass Metakinetoplastina gave rise to KKT17 and KKT18, including in Trypanosomatida and Bodonida. Separate duplication events gave rise to the three SYCP2-type paralogues (I–III) present in the prokinetoplastid endosymbiont *Perkinsella* sp. (high support, bootstrap >95). Interestingly, we also detected an independent duplication event in the sister lineage to Kinetoplastida, the poorly described flagellate order Diplonemida. Why these duplications occur, especially in these two euglenozoan lineages, is unclear. It is known that SYCP2L, the vertebrate-specific paralogue of SYCP2 (bootstrap support >95), localizes at centromeres during specific stages of meiosis [57], indicating that the centromeric localization of SYCP<sup>2-3</sup> paralogues is not unique to KKT17 and KKT18. Furthermore, we found two duplications that gave rise to the three SYCP2-type paralogues (I–III) present in the metamonad parasite *Trichomonas vaginalis* of the class Parabasalia. Finally, we also found a species-specific duplication of the SYCP2-type gene ASY3 (I–II) in *Chlamydomonas reinhardtii* (not shown in figure 4; see electronic supplementary material, file S4).

In a similar fashion, we traced the origins of SYCP3-type genes to four independent duplications of the coiled coil domains of clade-specific SYCP2-type ancestors, albeit generally with lower bootstrap support and positions less clearly reconcilable with the eukaryotic tree of life (see dark blue B, figure 4). In Archaeplastida, two independent duplications gave rise to an SYCP3-type paralogue in both Streptophyta (ASY4) and the green algal class Chlorophyceae (ASY4L). In

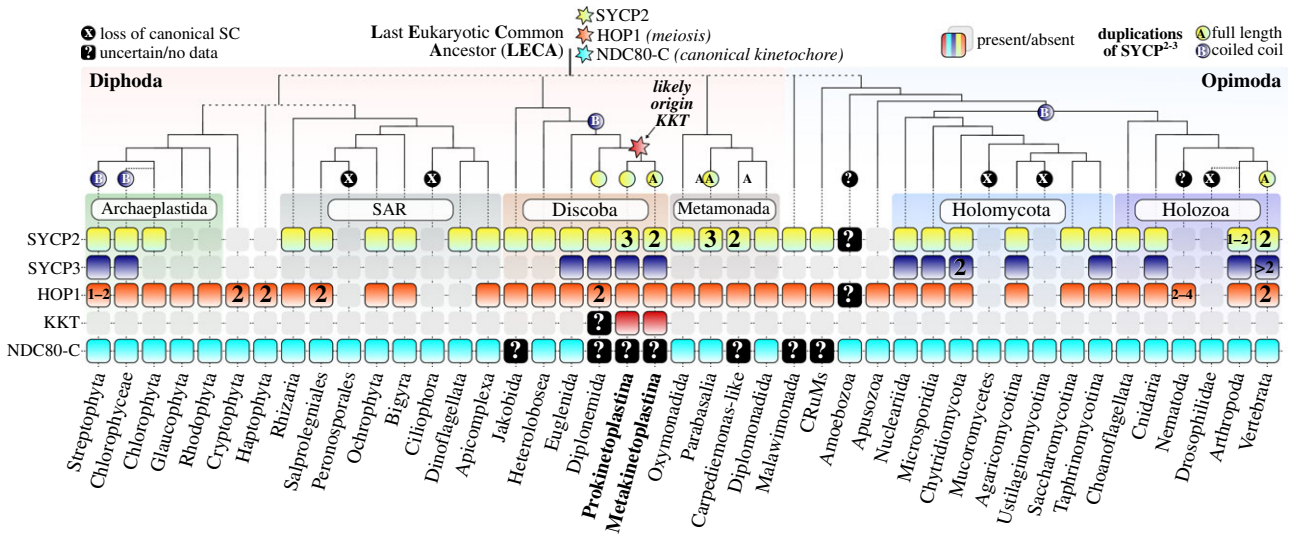


**Figure 4.** The SYCP<sup>2-3</sup> gene family expanded through independent duplications. Mirrored phylogenetic trees (IQ-Tree: maximum likelihood, see S5) of the *N*-terminal ARM repeats and PH domain of SYCP2-type genes (left), and coiled coils of SYCP2 and SYCP3-type genes (right, names are in dark blue). A and B indicate independent duplications of either the full-length SYCP2-type genes (yellow, (a) or the specific duplication of the C-terminal coiled coil (dark blue, (b)). Colours indicate taxonomic levels to which the various lineages belong (see legend top left). Values at branches indicate UltraFast Bootstrap support (1000x replicates: see S5), those associated with duplications are in bold and highlighted in yellow (full-length duplications) and dark blue (coiled coil duplications). Models of sequence evolution that were used to infer the phylogenetic trees are shown below each phylogram. Branch lengths are scaled and indicate the number of substitutions per site (see scale bar below each phylogram). A full overview of the uncollapsed phylograms, phylogenetic analyses details and the underlying alignments can be found in electronic supplementary material, file S4. \*ASY3 in Spermatophyta (e.g. in *A. thaliana*) lost the *N*-terminal ARM repeats and PH domain.

accordance with our HMM searches (figures 2*a* and 3*b*), the closest homologue of SYCP3 was SYCP2, signifying a duplication of the coiled coil domain in the common ancestor of Metazoa (figure 4). The close association of KKT16 in Kinetoplastida with SYCP3-type genes in Diplonemida (bootstrap support: 99) and SYCP2-type homologues from Euglenida (bootstrap support: 88) suggested that these duplications must have occurred in the ancestral Euglenozoa and that both SYCP2- and SYCP3-type genes were present in this ancestor (figure 4).

Altogether, our phylogenetic analysis indicated that SYCP2-type paralogues and SYCP3-type genes originated independently through recurrent duplications of multiple SYCP2-type ancestors at various taxonomic levels. Because we could not detect any SYCP<sup>2-3</sup> proteins in prokaryotes, we considered all eukaryotic members of the SYCP<sup>2-3</sup> gene family a single orthologous group, which was likely descendant from a single SYCP2-type gene present in the LECA. A graphical overview of the evolutionary scenario for SYCP<sup>2-3</sup>





**Figure 5.** Evolutionary scenario for the SYCP<sup>2-3</sup> gene family in the light of meiosis and kinetochores. Phylogenetic profiles of SYCP<sup>2-3</sup> and HOP1 orthologues, and the kinetoplastid and canonical kinetochore throughout the eukaryotic tree of life. See for sequences electronic supplementary material, table S2 and file S2. Canonical kinetochore (cyan, harbouring NDC80-based kinetochores), kinetoplastid kinetochore (dark red) and question marks (uncertain/no data) indicate evidence for the type of kinetochore proteins found among each eukaryotic lineage (see electronic supplementary material, table S2 for references and comments). Cartoon and classification of the eukaryotic tree of life, and the position of the LECA is guided by Burki *et al.* [56]. Numbers indicate the amount of paralogues found among particular lineages. 'A' and 'B' refer to the two types of independent duplications found for the SYCP<sup>2-3</sup> gene family similar to figure 4. (X) indicates co-loss of SYCP<sup>2-3</sup> and HOP1 (e.g. Ciliophora and Ustilaginomycotina). Hexagrams (stars) indicate the likely origin of each feature/gene family.

duplications, phylogenetic profiles and inferred ancestral states is presented in figure 5.

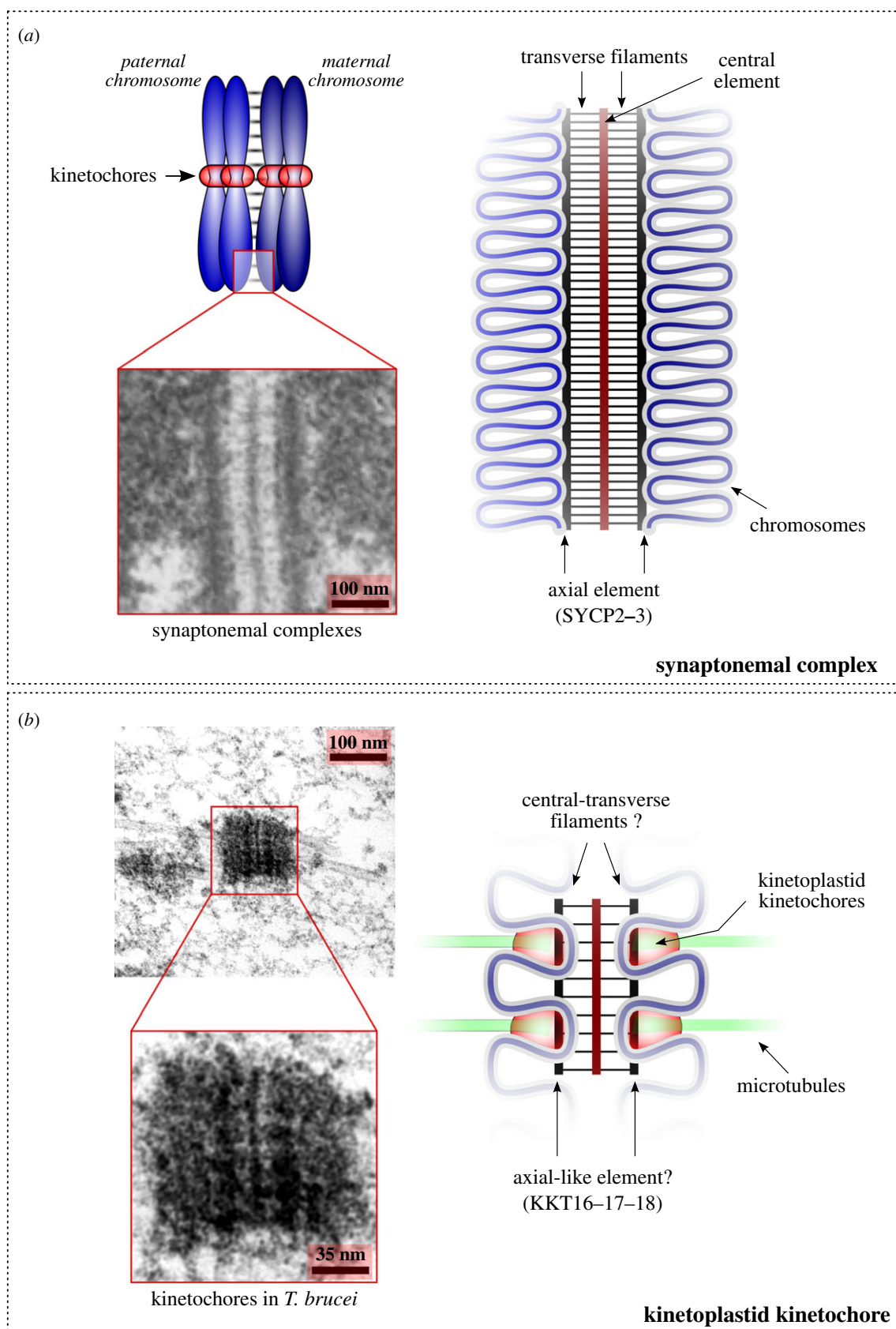
### 3. Discussion

#### 3.1. Widespread conservation of the SYCP<sup>2-3</sup> gene family in diverse eukaryotes points to an ancient origin for axial elements of the synaptonemal complex

This study revealed similarities in KKT16–KKT18 to the axial element components of SCs. This was an unexpected finding because KKT16–KKT18 are kinetochore proteins present in mitotic trypanosomes cells, while the SC is a strictly meiotic structure [40]. During meiosis, chromosome axes assemble on each pair of sister chromatids and serve as a platform from which chromatin loops radiate. The SC subsequently assembles in between the axes of homologous chromosomes along their lengths, forming a conspicuous zipper-like ultrastructure visible by electron microscopy (figure 6a). [32,47,48]. The structure of the SC is widely conserved among diverse eukaryotes and appears as a tripartite protein structure that consists of two axial elements (also known as lateral elements) that flank a central element to which they are connected via transverse filaments (figure 6a) [59]. In vertebrates, components of axial elements include SYCP2, SYCP3, meiotic HOP1/HORMAD family proteins and cohesin complexes [60]. HOP1/HORMADs and cohesins are conserved in many eukaryotes (figure 5) [8,61,62]. By contrast, it has been difficult to detect homologues for SYCP2 and SYCP3 or to establish their significant sequence similarity with SC components identified in other model systems (e.g. Red1 in *S. cerevisiae*, Rec10 in *S. pombe* and ASY3:ASY4 in *A. thaliana*) [49].

In this study, we used a remote homology detection protocol combining both profile-versus-profile and iterative HMM searches to identify highly divergent orthologues of the metazoan SYCP2:SYCP3 multimers and subunits of the KKT16 complex (figures 2 and 3). Specifically, we made use of the 'hopping' strategy, which was previously employed to detect divergent homologues of canonical kinetochore proteins [8,63,64]. The 'hopping' strategy uses the following logic: 'if A is homologous to B and B is homologous to C, then A must be homologous to C'. This approach has the potential to find many more orthologues of highly divergent gene families such as those involved in meiosis and the kinetochore, but inclusion criteria other than sequence similarity must be considered to prevent false-positive candidate orthologues. In particular, for SYCP<sup>2-3</sup> proteins, which harbour generic domains such as ARM repeats and coiled coils, we selected only those candidate sequences that either had the full-length SYCP2-type ARM-PH-coiled coil topology or coiled coil-only proteins that showed significant similarity in reciprocal HMM searches with either known SC components or KKT16-18 (e.g. ASY3, KKT17 or SYCP3) and/or a full-length SYCP2-type sequence of the same eukaryotic lineages. These stricter criteria potentially resulted in the exclusion of SYCP3-type coiled coil-only genes among eukaryotic lineages, for instance in Apicomplexa and other phyla from the SAR supergroup (see absences in figure 5). Further experimental validation of the localization and molecular function of such candidates from the SAR supergroup will be needed to gain more confidence that these genes would be part of the SYCP<sup>2-3</sup> gene family and exert a function in the SC.

In addition to showing that experimentally verified axial SC components of several models (animals, fungi and plants) and KKT16–KKT18 belong to a common gene family, we identified SYCP<sup>2-3</sup> orthologues in all eukaryotic supergroups (figure 5), although not in any prokaryote. The widespread presence of these proteins suggests that axial elements were part of the ancient eukaryotic SC and that SYCP<sup>2-3</sup> genes



**Figure 6.** Genomic and microscopic similarity of kinetoplastid kinetochores and the synaptonemal complex suggests a common origin. (a) Schematic of the meiotic synapsis (left) and the synaptonemal complexes (right). SYCP2 and SYCP3 are components of the axial element. Zoom in is an electron micrograph of a mouse spermatocyte, showing synaptonemal complexes in between homologous chromosomes. Reproduced from ref. [58] under CC-BY. (b) Left: electron micrograph of a mitotic trypanosome cell, showing that the putative kinetochore structure attaches spindle microtubules from opposite poles. Adapted from ref. [28] with permission from Springer Nature. Right: hypothetical model showing that the KKT16 complex may form an axial element-like structure in the kinetoplastid kinetochores.

were probably present in the LECA (figure 5). We detected highly divergent SYCP<sup>2-3</sup> orthologues in Metamonada (e.g. *Giardia intestinalis* and *T. vaginales*), Microsporidia (e.g.

*Encephalitozoon intestinalis*) and a wide variety of fungi (e.g. *Neurospora crassa*, *Fusarium oxysporum*, *Spizellomyces punctatus*; see electronic supplementary material, table S2 and file S2).

We specifically searched for SYCP<sup>2-3</sup> genes in *Drosophila*-related lineages since they seem to have a largely analogous SC and SYCP<sup>2-3</sup> genes that were previously not identified [65]. We found divergent SYCP<sup>2-3</sup> orthologues in various insect lineages (Lepidoptera, Diptera), but none in Drosophilidae, signifying the specific loss of SYCP<sup>2-3</sup> genes in this lineage (figure 5). Interestingly, we also did not detect any SYCP<sup>2-3</sup> homologues in lineages, which were previously described to lack SC structures or pathways of meiotic recombination, such as Ciliophora [66,67], Amoebozoa [68] and Ustilaginomycotina [69,70]. The status of Amoebozoa was somewhat unclear as we found one SYCP2-type gene in the amoeba *Planoprotostelium fungivorum* (figure 5), but no significant sequences were found in the reciprocal similarity searches. By contrast, we detected candidate homologues in lineages with previously described SC-like structures during meiosis, such as Apicomplexa [71] and Oxymonadida [72]. The general concordance between the presence of SYCP<sup>2-3</sup> genes and SC-like structures suggests that these genes could be used as a good predictor for the presence of canonical SCs. However, we also found cases where SC-like structures were described, but no SYCP<sup>2-3</sup> genes were detected, such as Bacillariophyta (diatoms) and in Rhodophyta (red algae) [59,73,74]. To examine the possibility of missed detection of SYCP<sup>2-3</sup>, we searched for orthologues of the meiotic HORMA domain protein Hop1/HORMAD, which is a meiosis-specific interactor of SYCP2-type proteins [46,60] and is typically expected to co-occur with the presence of canonical SC and meiosis [51,65]. The phylogenetic profiles of SYCP<sup>2-3</sup> and HOP1 corresponded well (31/43 shared presences and absences, figure 5), but we found six lineages (Rhodophyta, Glaucophyta, Cryptophyta, Haptophyta, Apusozoa and Nematoda) that do contain HOP1, but not SYCP<sup>2-3</sup> (figure 5). Conversely, we detected several highly divergent SYCP2-type proteins among dinoflagellates, in contrast with a recent report [75], while such lineages lacked HOP1 (see electronic supplementary material, table S2 and file S2). Whether these discordances between HOP1 and SYCP<sup>2-3</sup> point to a lack of homology detection or whether SC-like structures in these lineages contain analogous SC components like in Nematoda [76] and Drosophilidae [65] remains unclear. In any case, the absence of HOP1 and SYCP<sup>2-3</sup> in Mucoromycetes, Zoopagomycota (see electronic supplementary material, table S2) and the oomycote class Peronosporales (potato blight pathogen *Phytophthora infestans*) probably signifies the absence of a canonical SC structure in these lineages.

We detected 12 recurrent duplications for the SYCP<sup>2-3</sup> gene family. Why these duplications occurred remains unclear. In the case of the lineages-specific coiled coil-only duplications that gave rise to different SYCP3-type genes (i.e. SYCP3, KKT16, ASY4, ASY4L), we speculate a specific need for the coiled coil to facilitate the formation of axial element-like structures apart from the function of the N-terminal domain ARM-PH, which is currently unknown. In the case of SYCP2-type genes, there are only limited data available on the two paralogues present in vertebrates: SYCP2 and SYCP2L. SYCP2L is expressed specifically in oocytes and localizes at SCs and centromeres [57], although its function remains unclear. It is noteworthy that both Kinetoplastida and Diplonemida show recurrent duplications of SYCP2-type genes (figures 4 and 5). This apparent increase in paralogues might correlate with new functions of these proteins in the kinetochore rather than in the SC. It will therefore be of interest to study SYCP<sup>2-3</sup>

homologues in Diplonemida and assess whether these proteins play a role in the kinetochore and/or the SC.

### 3.2. Hypothesis: kinetoplastids repurposed meiotic structures to build kinetochores

Beyond the shared ancestry of the KKT16 complex and axial element components of the SC, other KKT proteins have conserved domains with relevance to homologous recombination and chromosome synapsis (i.e. BRCT domain, FHA domain, and Polo-like kinases). We therefore propose that ancient kinetoplastid ancestors repurposed parts of the meiotic machinery to assemble a kinetochore-like structure by restricting its formation to one chromosomal region and acquiring microtubule-binding activities (figure 6*b*). This hypothesis could explain the unique organization of kinetoplastid kinetochores that lack a clear gap between sister kinetochores even at metaphase [25–28] and, indeed, are strikingly similar in structure to SCs (figure 6) [58].

Functions of the KKT16 complex at the kinetoplastid kinetochore remain unknown. Our previous mass spectrometry analysis of the KKT16 complex purifications from mitotically growing cells did not reveal significant co-purification of cohesin subunits or HOP1 [16]. It is therefore currently unclear whether the KKT16 complex has a similar function to SYCP<sup>2-3</sup> homologues found in other eukaryotes. Because KKT16–KKT18 are the only members of the SYCP<sup>2-3</sup> gene family present in kinetoplastids, it will be interesting to examine whether KKT16–KKT18 are also used as components of SCs during meiosis, which takes place when trypanosomes transmit in the salivary glands of the tsetse fly vector [77]. Although the main functions of the SC are to hold homologous chromosomes together and promote recombination, it is known that SC components have non-canonical functions in certain lineages. For example, some organisms rely on the SC or its components for connection between homologous chromosomes beyond prophase I (when the SC is disassembled in most organisms). In the female silkworm *Bombyx mori* that lacks meiotic recombination, homologous chromosomes are joined together by the retention of modified SCs until metaphase I [78]. Similarly, some SC components remain at the centromeric or noncentromeric region and promote biorientation of non-exchange chromosomes [79]. Functions of the KKT16 complex remain unknown, so it will be interesting to test whether it plays any role in connecting and properly orienting sister chromatids in trypanosomes.

### 3.3. Origins of kinetoplastid kinetochores in the light of early eukaryotic evolution

Why do kinetoplastids have unique kinetochores, while other eukaryotes have canonical kinetochore proteins? The absence of canonical kinetochore proteins among Kinetoplastida provides several hypothetical scenarios for the evolution of the kinetoplastid kinetochore. In the case that the first common ancestors of Kinetoplastida possessed the canonical kinetochore system, it must have been secondarily lost and presumably replaced by the new unique kinetochore system now found in this group (figure 5). This scenario seems to be most consistent with the current consensus on the eukaryotic tree of life, where Kinetoplastida are part of the phylum Euglenozoa, which also includes Diplonemida,

Symbiontida and Euglenida [56]. Furthermore, most eukaryotes have the canonical kinetochore system [8], including euglenids [80] (figure 5). It is noteworthy that an initial survey of diplomonid transcriptomes found only limited evidence for the presence of a canonical kinetochore system with putative candidates for the centromeric H3 variant CENP-A, but no subunits of the NDC80 complex or other structural kinetochore components [39] (figure 5). Intriguingly, no clear orthologues of the kinetoplastid kinetochore proteins could be identified either [39], suggesting that Diplonemida might potentially have yet another kinetochore system. Identification of kinetochore proteins in Diplonemida and in-depth sequence analyses of these and kinetoplastid kinetochore proteins using our sensitive homology detection workflow will be needed to shed further light on how ancestral kinetoplastids acquired a unique kinetochore system.

An alternative scenario is that early kinetoplastid ancestors never possessed the canonical kinetochore system. There is a controversial hypothesis that places the root of the eukaryotic tree of life between Euglenozoa (or deeply within Euglenozoa) and all other eukaryotes [81]. In this ‘Euglenozoa-first’ scenario (discussed in [21]), it is possible that kinetoplastid kinetochores and canonical kinetochores were invented independently, and they are both derived systems, meaning that ancestral eukaryotes might have possessed a chromosome segregation machinery that does not exist anymore today. It is still unclear whether mitosis or meiosis evolved first [82–86]. By contrast, it is known that some species of Archaea are capable of homologous recombination and cell fusion [87]. Although we have been unable to find any SYCP<sup>2–3</sup> genes in Archaea or bacteria, the widespread presence of SCs among eukaryotes, including Euglenozoa, suggests that SCs were likely present in the LECA (figure 5). Under the Euglenozoa-first hypothesis, our findings that KKT16 complex subunits have similarities to SC components raise a possibility that some features of meiosis (i.e. chromosome synapsis and genetic exchange) might have evolved before an active chromosome segregation mechanism that relies on kinetochores and spindle microtubules.

## 4. Concluding remarks

Although the kinetochore is at the heart of chromosome segregation, substantial compositional diversity and rapid sequence evolution of its subunits are widespread throughout the eukaryotic tree of life. This presents us with fundamental questions: how can kinetochores be essential and divergent at the same time? How (and why) do cells replace one kinetochore system with another? While Drinnenberg *et al.* used the elegant analogy with the ‘ship of Theseus’ to explain this remarkable evolutionary behaviour of kinetochores [10], the radically different composition of the kinetoplastid kinetochore seems to be at odds with such a piece-by-piece replacement model. Our study provides a new concept for understanding such an extreme jump in the evolution of kinetochores in eukaryotes, namely the apparent ability to adapt and repurpose meiotic complexes for mitotic functions. Further functional and evolutionary characterization of divergent kinetochores in Kinetoplastida and other eukaryotes will thus not only benefit our understanding of their inner workings but also shed light on how this core cellular system has been allowed to diverge in such a radical fashion.

## 5. Data and methods

### 5.1. Primers and plasmids

Primers used in this study are listed in electronic supplementary material, table S3. To make pBA198 (6His-KKT16), KKT16 was amplified from genomic DNA using BA509/BA510 and cloned into the EcoRI/HindIII sites of the pST44 polycistronic expression vector (RRID: Addgene\_64007) [42]. To make pBA200 (KKT18, 6His-KKT16), KKT18 was amplified from genomic DNA using BA511/BA512 and cloned into pBA198 using XbaI/BglII sites. To make pBA202 (KKT18, 6His-KKT16, KKT17), KKT17 was amplified from genomic DNA using BA513/BA514 and cloned into pBA200 using KpnI/MluI sites.

### 5.2. Expression and purification of the recombinant KKT16 complex

pBA202 (KKT18, 6His-KKT16 and KKT17) was transformed into Rosetta 2(DE3)pLysS *Escherichia coli* cells (Novagen, 71403). Cells were grown in Lysogeny broth media (Fisher Scientific, BP1426-2) at 37°C to an OD<sub>600</sub> of approximately 0.6 and protein expression was induced by 0.2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich, I6758) at 18°C overnight. Recombinant proteins were purified using an Ni-NTA Fast Start Kit under native condition (Qiagen, 30600). To check the expression of KKT16 complex subunits, protein expression was induced by 0.2 mM IPTG at 37°C for 3 h. For cross-linking mass spectrometry experiments, the KKT16 complex (pBA202) was expressed in Rosetta 2(DE3)pLysS *E. coli* cells and purified as follows. Cells were grown in auto induction media (Formedium, AIMLB0205) [88] at 37°C to an OD<sub>600</sub> of 0.55 and then grown overnight at 18°C. Cells were pelleted at 3400 g at room temperature, and the cell pellet was frozen in liquid nitrogen and stored at –80°C. Cells were resuspended in P500a lysis buffer (50 mM sodium phosphate pH 7.5 (Sigma-Aldrich, S0876 and S0751), 500 mM NaCl (Sigma-Aldrich, S9888) and 10% glycerol (Sigma-Aldrich, G5516)) supplemented with protease inhibitors (20 μg ml<sup>–1</sup> leupeptin (Merck, E18), 20 μg ml<sup>–1</sup> pepstatin (Merck, 516481-100MG), 20 μg ml<sup>–1</sup> E-64 (Peptanova, 4096-100) and 2 mM benzamide (Sigma-Aldrich, 434760)) and 0.5 mM TCEP (Sigma-Aldrich, C4706) and were sonicated on ice. Lysed cells were spun at 48 000 g at 4°C for 30 min. The supernatant was incubated with TALON beads (Takara Clontech, 635503) for 1 h at 4°C. We washed the beads with lysis buffer, eluted proteins with an elution buffer (50 mM sodium phosphate, pH 7.5, 500 mM NaCl, 10% glycerol and 250 mM imidazole (Sigma-Aldrich, 56750)) with 0.5 mM TCEP. The sample was stored at –80°C.

### 5.3. Chemical cross-linking mass spectrometry (XL-MS)

BS<sup>3</sup> cross-linker (bis(sulfosuccinimidyl)suberate) (Fisher Scientific, 10066323) was equilibrated at room temperature for 2 h and then resuspended to 0.87 mM in distilled water. Then, immediately 2 μl of the cross-linker was mixed with 18 μl of 0.5 mg ml<sup>–1</sup> KKT16-17-18 in 25 mM sodium phosphate, pH 7.5, 250 mM NaCl, 5% glycerol, 125 mM imidazole and 0.25 mM TCEP. The cross-linking reaction occurred on ice for 60 min. Following the incubation, the sample was boiled for

10 min and resolved on a NuPAGE 4–12% gradient polyacrylamide gel (Fisher Scientific, NP0322). The gel was stained using SimplyBlue (Fisher Scientific, LC6060), and bands corresponding to cross-linked species were cut out and subjected to mass spectrometry as described previously [89]. Mass spectrometric data were converted into mgf format using pParse and searched by the pLink software [90] (version 2.3.5) as described previously [89]. Search parameters were as follows: maximum number of missed cleavages = 3, fixed modification = carbamidomethyl-Cys and variable modification = Oxidation-Met. Cross-links that have score  $< 1 \times 10^{-6}$  were visualized using xiNET [91] (electronic supplementary material, table S1). All raw files relating to cross-linking mass spectrometry have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [92] with the dataset identifier PXD025220.

#### 5.4. Secondary structure-guided profile-versus-profile HMM comparisons and visualization

Distant homologues of KKT16 complex members (KKT16–KKT18) were extracted from a previous extensive survey of kinetoplastid kinetochore proteins by Butenko *et al.* [39]. For each subunit, we constructed MSAs using MAFFT (v. 7.475 [93], RRID:SCR\_011811: option ‘eins-i’ or ‘lins-i’). Secondary structure-annotated profile HMMs were constructed based on MSAs of both full-length sequences and domains (ARM repeats, PH and coiled coil) using the ‘hhmake’ script from HHsuite3 (RRID: SCR\_010277 [45]). To map the domain architecture and potentially uncover highly divergent homologues of KKT16 complex members, we searched the pre-compiled pdb70 (RRID:SCR\_012820) and pfam-A (RRID: SCR\_004726) profile HMM databases from the HHsuite repository (link, downloaded 1 November 2020) using the KKT16 complex subunit HMM models (see electronic supplementary material, file S1 for HHsearch text output for full-length and subdomain HMM-versus-HMM profile searches). All HMMs (including iterative HMM searches, see below) used during this study were annotated and visualized using custom python scripts (see electronic supplementary material, file S3 for HMM and MSA text files and visualization, and electronic supplementary material, file S6 for relevant settings and sources for visualization). The following predictions/annotations were included: (1) amino acid conservation (Shannon entropy: bits of information), derived via the Skylign API, RRID: SCR\_001176 [94]), (2) secondary structure (PSIPRED, RRID:SCR\_010246 [95]), (3) coiled coil (DeepCoil version 1.0: <https://github.com/labstructbioinf/DeepCoil>) [96]), and (4) intrinsic structural disorder (IUPRED version 2a, RRID: SCR\_014632 [97]). MSA columns that were not present in either the first or the MSA consensus sequence were removed to ensure gapless HMM visualization. Plots were generated for each HMM and manually compiled into figures using the open-source scalable vector graphics editor Inkscape 1.0rc1 for macOS (Inkscape Project 2020, retrieved from <https://inkscape.org>, RRID: SCR\_014479).

#### 5.5. Sequence database

We compiled a large sequence database consisting of 343 (single cell) genomes and transcriptomes of a wide variety of eukaryotes [8,98,99] (see electronic supplementary material,

table S2 for sources). We specifically focussed on including lineages closely related to Kinetoplastida, such as Diplone-mida, Euglenida and other Discoba [39,100–103], and taxa related to lineages that lack canonical SC proteins (e.g. nematodes, Drosophilidae, Ciliophora and Amoebozoa). For several species, it was not possible to obtain annotated protein-coding regions. In these cases, we used TransDecoder v5.5.0 to predict open reading frames (<https://github.com/TransDecoder/TransDecoder>).

#### 5.6. Supervised remote homology detection (hopping) protocol

Because of the highly divergent sequence composition of KKT16-18 and other axial element components of the SC found in diverse eukaryotes, we employed a supervised homology detection protocol to optimize multiple HMMs based on iterative reciprocal similarity searches using jackhmmmer and hmmsearch (HMMER 3.1 and 3.3; RRID:SCR\_005305 [104]). Searches were executed with standard inclusion thresholds until convergence, unless otherwise specified. HMMs were constructed using ‘hmmbuild’. Our protocol was based on the following steps/considerations: (i) to increase the initial search sensitivity, we constructed profile HMMs of automatically defined clade-specific orthologous groups (OrthoFinder 2.0, RRID: SCR\_017118 [105], standard settings) of KKT16-18 and other axial element components (e.g. SYCP2,3 and ASY3,4). We used both full-length and subdomain (ARM, PH, coiled coil) HMMs as seeds for iterative sequence searches. (ii) When queries using seed HMMs returned few hits, we searched Uniprot (RRID: SCR\_002380) with the jackhmmmer web server (<https://www.ebi.ac.uk/Tools/hmmer/search/jackhmmmer> [106]) for multiple iterations until no new putative candidate homologues could be included ( $E < 0.01$ , domain  $E < 0.03$ ). (iii) Iterative HMM searching is highly sensitive, and either too many homologues or other non-homologous sequences can be included by mistake due to the presence of highly common domains or due to local sequence biases such as coiled-coil regions. To prevent the inclusion of potentially false-positive candidates, we used either the full-length sequence or other non-common domain/coiled coil regions of these new candidates as a query for jackhmmmer searches. In case neither searches yielded reciprocal hits after numerous iterations or hits were clearly part of non-homologous proteins (with other domains or lacking either the ARM repeats or coiled coil), they were discarded as putative homologues. For example, we noted that multiple iterations with full-length SYCP2-type HMM profiles resulted in the frequent eventual inclusion of homologues of long ARM repeat proteins such as Vac8p and APC (also found to be similar based on HHsearch; figure 2a). Such ARM repeat similarities point to more ancient homologous connections between these groups of proteins, but the absence of a PH domain and C-terminal coiled coil prompted us to remove these sequences to prevent their inclusion as candidate SYCP2-type genes in subsequent iterations. Altogether, we found that significant similarities with the region spanning the last part of the ARM repeats and the PH domain in combination with the presence of a C-terminal coiled coil were the best inclusion criteria to distinguish between SYCP2-type homologues and other eukaryotic ARM repeat or PH domain-bearing proteins. Our profile HMM searches

frequently returned metazoan apolipoprotein sequences as putative homologues of SYCP2 and SYCP3-type proteins (see also the HHsearch output in figures 2*a* and 3*a*) through similarity of their coiled coil regions. While these extracellular lipid-binding proteins could formally be homologous to SC proteins, we deemed it more likely that the coiled coils of these proteins evolved convergently. To restrict further inclusion of false-positive coiled coil homologues, we only included candidates with bidirectional best similarity to experimentally verified eukaryotic SC components (Red1, Rec10 and ASY3,4, SYCP2,3), subunits of the KKT16 complex and/or an SYCP2- or SYCP3-type gene of the same clade when a candidate for both is present. (iv) Due to the highly divergent nature of SC components and KKT proteins, we could not establish a single optimized profile HMM that captured all orthologous sequences. Instead in many instances, we used the sequence ‘HMM hopping’ method, which follows from the logic that homology is a transitive feature by nature: if A is homologous to B and B is homologous to C, then A is homologous to C. Also possible reciprocal ‘HMM hopping’ searches were performed to increase our confidence in distant homologous relationships. In the case only unidirectional searches yielded new candidates or experimentally verified SC components, searches were repeated with more permissive bit scores (18–25) or *E*-values (0.1–1) to achieve reciprocal homologous relationships. Such cases were specifically scrutinized for similarity in the secondary structure as well as their sequence composition. Examples of profile HMM hopping schemes for establishing homology between SC components and KKT16–18 are visualized in figures 2*b* and 3*b*. (v) If any of the iterative (reciprocal) searches using all of the (clade-specific) seed and optimized profile HMMs yielded overlapping hits and met the criteria mentioned earlier, we included these sequences as orthologues (see electronic supplementary material, file S2 for SYCP2 and SYCP3-type sequences). HMMs of clade-specific SYCP2–3 genes used to establish the homology between SYCP2 and SYCP3, ASY3 and ASY4, Red1, Rec10 and Rec27 and KKT16–KKT18 can be found in electronic supplementary material, file S3.

## 5.7. Phylogenetic analyses

Due to the highly divergent nature of the SYCP2–3 gene family, generating a high-quality MSA including all sequences was not feasible. We therefore adopted a previously used [29] iterative alignment protocol to generate a super alignment consisting of separate clade-specific MSAs (see §5 for definition of clade-specific MSAs) using the ‘merge’ option in MAFFT (MAFFT v7.475 [93], RRID: SCR\_011811: merge, ginsi, unalignlevel 0.6). Before addition to the super alignment, each clade-specific MSA was trimmed with trimAl (v1.4.rev22, RRID: SCR\_017334 [107]) to remove unconserved positions. The order of MSA merging was determined based on bidirectional next-best sequence recovery using hmmsearch [104]. For instance, the coiled coils of metazoan SYCP2 and SYCP3 were closest, as reciprocal searches using the SYCP2 HMM yielded SYCP3 homologues as best next hits, and vice versa. MSAs were manually scrutinized for apparent misalignments and re-run using either the MAFFT ‘eins-i’ or the ‘linsi-i’ option to yield better

alignments. This procedure was performed separately for the *N*-terminal ARM repeats and PH domain of SYCP2-type homologues and C-terminal coiled coils of SYCP2 and SYCP3-type homologues. The coiled coils of SYCP2 and SYCP3-type proteins in species of the SAR supergroup, Fungi (e.g. Rec10, Rec27 and Red1), *Bodo saltans* KKT18, *Perkinsela* sp. I and *T. vaginales* I and the *N*-terminal ARM-PH domain of Red1/Rec10-like homologues were too divergent to yield MSAs of sufficient quality. Furthermore, phylogenetic tree inference including these sequences showed signs of long-branch attraction and were therefore left out of the phylogenetic analysis. For the final super alignments, only positions with column occupancy higher than 30% (ARM repeats + PH domain) and 70% (coiled coil) were considered for further analysis. Maximum-likelihood phylogenetic analyses (phylograms shown in figure 4; for data files see electronic supplementary material, file S4) were performed with IQ-Tree (version 1.6.12, RRID: SCR\_017254) with automatic substitution model selection using ModelFinder [108], a GAMMA model of rate heterogeneity and 1000 Ultrafast bootstrap and SH-like approximate likelihood ratio test replicates [55]. Parameters for the final phylogenetic analyses are as follows: ARM repeats + PH domain (alignment: 424 positions, model: LG + G4 + F); coiled coil (alignment: 165 positions, model: JTT + G4 + F). Trees were visualized and annotated using FigTree v1.4.4 [109].

**Data accessibility.** All data, annotations and scripts are included in the manuscript. The scripts to execute the visualization of the secondary structure, conservation and the HHsearch output cannot be shared directly due to local computational environment dependencies. A detailed description/manual is available to replicate these analyses/visualizations (see §5 and electronic supplementary material, file S6 for description). The same electronic supplementary material files were also deposited in the FigShare repository at doi:10.6084/m9.figshare.13725787 [110].

**Authors’ contributions.** E.C.T. performed all the bioinformatic analyses and compiled the scripts to generate the figures. T.A.W. assisted in setting up the evolutionary analysis using the HHsearch algorithm and wrote the associated python script for visualization of the results. R.F.W. supervised the project. B.A. reconstituted the KKT16 complex and noticed the similarity between the SC and trypanosome kinetochore architectures in electron microscopy images. P.L. purified the KKT16 complex and performed cross-linking mass spectrometry. B.A. and E.C.T. wrote the paper together with input from R.F.W. and T.A.W.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Competing interests.** We declare we have no competing interests.

**Funding.** E.C.T. was funded through a personal Postdoctoral Research Fellowship awarded by the Herchel Smith Fund at the University of Cambridge (United Kingdom). P.L. was supported by the Boehringer Ingelheim Fonds. B.A. was supported by a Wellcome Trust Senior Research Fellowship (grant no. 210622/Z/18/Z) and the European Molecular Biology Organisation Young Investigator Program. RFW was funded by the Wellcome Trust (grant no. 214298/Z).

**Acknowledgments.** We thank Alastair Simpson, Gordon Lax and Julius Lukeš for providing access to Euglenida, Diplonemida and/or Kinetoplastida transcriptomes and genomes before publication; Svenja Hester in the Advanced Proteomics Facility for processing mass spectrometry samples; Shabaz Mohammed for advice on crosslinking mass spectrometry and Keith Gull for providing an original electron micrograph of *T. brucei* kinetochores. We also thank Kim Nasmyth and David Sherratt for discussion. We thank members of the Akiyoshi and Waller labs for feedback.

## References

- McIntosh JR. 2016 Mitosis. *Cold Spring Harb. Perspect. Biol.* **8**, a023218. (doi:10.1101/cshperspect.a023218)
- Findeisen P, Mühlhausen S, Dempewolf S, Hertzog J, Zietlow A, Carlomagno T, Kollmar M. 2014 Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol. Evol.* **6**, 2274–2288. (doi:10.1093/gbe/evu187)
- Biggins S. 2013 The composition, functions, and regulation of the budding yeast kinetochore. *Genetics* **194**, 817–846. (doi:10.1534/genetics.112.145276)
- Musacchio A, Desai A. 2017 A molecular view of kinetochore assembly and function. *Biology* **6**, 5. (doi:10.3390/biology6010005)
- Drinneberg IA, deYoung D, Henikoff S, Malik HS. 2014 Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife* **3**, e03676. (doi:10.7554/eLife.03676)
- Navarro-Mendoza MI *et al.* 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.e6. (doi:10.1016/j.cub.2019.09.024)
- Lowell JE, Cross GAM. 2004 A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. *J. Cell Sci.* **117**, 5937–5947. (doi:10.1242/jcs.01515)
- 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. (doi:10.15252/embr.201744102)
- Pontremoli C, Forni D, Pozzoli U, Clerici M, Cagliani R, Sironi M. 2021 Kinetochore proteins and microtubule-destabilizing factors are fast evolving in eutherian mammals. *Mol. Ecol.* **30**, 1505–1515. (doi:10.1111/mec.15812)
- Drinneberg IA, Henikoff S, Malik HS. 2016 Evolutionary turnover of kinetochore proteins: a Ship of Theseus? *Trends Cell Biol.* **26**, 498–510. (doi:10.1016/j.tcb.2016.01.005)
- Berriman M *et al.* 2005 The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422. (doi:10.1126/science.1112642)
- d'Ávila-Levy CM *et al.* 2015 Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem. Inst. Oswaldo Cruz* **110**, 956–965. (doi:10.1590/0074-02760150253)
- Keeling PJ, Burki F. 2019 Progress towards the Tree of Eukaryotes. *Curr. Biol.* **29**, R808–R817. (doi:10.1016/j.cub.2019.07.031)
- Field MC *et al.* 2017 Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. *Nat. Rev. Microbiol.* **15**, 217–231. (doi:10.1038/nrmicro.2016.193)
- Ersfeld K, Gull K. 1997 Partitioning of large and minichromosomes in *Trypanosoma brucei*. *Science* **276**, 611–614. (doi:10.1126/science.276.5312.611)
- Akiyoshi B, Gull K. 2014 Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156**, 1247–1258. (doi:10.1016/j.cell.2014.01.049)
- Nerusheva OO, Akiyoshi B. 2016 Divergent polo box domains underpin the unique kinetoplastid kinetochore. *Open Biol.* **6**, 150206. (doi:10.1098/rsob.150206)
- Nerusheva OO, Ludzia P, Akiyoshi B. 2019 Identification of four unconventional kinetoplastid kinetochore proteins KKT22–25 in *Trypanosoma brucei*. *Open Biol.* **9**, 190236. (doi:10.1098/rsob.190236)
- D'Archivio S, Wickstead B. 2017 Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J. Cell Biol.* **216**, 379–391. (doi:10.1083/jcb.201608043)
- Brusini L, D'Archivio S, McDonald J, Wickstead B. 2019 Ndc80/Nuf2-like protein KPIP1 connects a stable kinetoplastid outer kinetochore complex to the inner kinetochore and responds to metaphase tension. bioRxiv.
- Akiyoshi B. 2016 The unconventional kinetoplastid kinetochore: from discovery toward functional understanding. *Biochem. Soc. Trans.* **44**, 1201–1217. (doi:10.1042/BST20160112)
- Goshima G, Yanagida M. 2000 Establishing biorientation occurs with precocious separation of the sister kinetochores, but not the arms, in the early spindle of budding yeast. *Cell* **100**, 619–633. (doi:10.1016/S0092-8674(00)80699-6)
- Fujiwara T, Tanaka K, Kuroiwa T, Hirano T. 2013 Spatiotemporal dynamics of condensins I and II: evolutionary insights from the primitive red alga *Cyanidioschyzon merolae*. *Mol. Biol. Cell* **24**, 2515–2527. (doi:10.1091/mbc.e13-04-0208)
- Bloom KS. 2014 Centromeric heterochromatin: the primordial segregation machine. *Annu. Rev. Genet.* **48**, 457–484. (doi:10.1146/annurev-genet-120213-092033)
- Solari AJ. 1980 The 3-dimensional fine structure of the mitotic spindle in *Trypanosoma cruzi*. *Chromosoma* **78**, 239–255. (doi:10.1007/BF00328395)
- Ureña F. 1986 Three-dimensional reconstructions of the mitotic spindle and dense plaques in three species of *Leishmania*. *Z. Parasitenkd.* **72**, 299–306. (doi:10.1007/BF00928739)
- Solari AJ. 1995 Mitosis and genome partition in trypanosomes. *Biocell* **19**, 65–84.
- Ogbadoyi E, Ersfeld K, Robinson D, Sherwin T, Gull K. 2000 Architecture of the *Trypanosoma brucei* nucleus during interphase and mitosis. *Chromosoma* **108**, 501–513. (doi:10.1007/s004120050402)
- Tromer EC, van Hooff JJE, Kops GJPL, Snel B. 2019 Mosaic origin of the eukaryotic kinetochore. *Proc. Natl. Acad. Sci. USA* **116**, 12 873–12 882. (doi:10.1073/pnas.1821945116)
- Li Z, Umeiyama T, Wang CC. 2008 The chromosomal passenger complex and a mitotic kinesin interact with the Tousled-like kinase in trypanosomes to regulate mitosis and cytokinesis. *PLoS ONE* **3**, e3814. (doi:10.1371/journal.pone.0003814)
- Reinhardt HC, Yaffe MB. 2013 Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response. *Nat. Rev. Mol. Cell Biol.* **14**, 563–580. (doi:10.1038/nrm3640)
- Hunter N. 2015 Meiotic recombination: the essence of heredity. *Cold Spring Harb. Perspect. Biol.* **7**, a016618. (doi:10.1101/cshperspect.a016618)
- Jordan PW, Karppinen J, Handel MA. 2012 Polo-like kinase is required for synaptonemal complex disassembly and phosphorylation in mouse spermatocytes. *J. Cell Sci.* **125**, 5061–5072. (doi:10.1242/jcs.105015)
- Archambault V, Glover DM. 2009 Polo-like kinases: conservation and divergence in their functions and regulation. *Nat. Rev. Mol. Cell Biol.* **10**, 265–275. (doi:10.1038/nrm2653)
- Zitouni S, Nabais C, Jana SC, Guerrero A, Bettencourt-Dias M. 2014 Polo-like kinases: structural variations lead to multiple functions. *Nat. Rev. Mol. Cell Biol.* **15**, 433–452. (doi:10.1038/nrm3819)
- Parsons M, Worthey EA, Ward PN, Mottram JC. 2005 Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* **6**, 127. (doi:10.1186/1471-2164-6-127)
- Marcianò G, Nerusheva OO, Ishii M, Akiyoshi B. 2019 Unconventional kinetochore kinases KKT2 and KKT3 have a unique zinc finger that promotes their kinetochore localization. bioRxiv.
- Ohno S. 1970 *Evolution by gene duplication*. Berlin, Germany: Springer.
- Butenko A *et al.* 2020 Evolution of metabolic capabilities and molecular features of diplomonads, kinetoplastids, and euglenids. *BMC Biol.* **18**, 23. (doi:10.1186/s12915-020-0754-1)
- Zickler D, Kleckner N. 1999 Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.* **33**, 603–754. (doi:10.1146/annurev.genet.33.1.603)
- Rice P, Longden I, Bleasby A. 2000 EMBOS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277. (doi:10.1016/S0168-9525(00)02024-2)
- Tan S, Kern RC, Selleck W. 2005 The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* **40**, 385–395. (doi:10.1016/j.pep.2004.12.002)
- Feng J, Fu S, Cao X, Wu H, Lu J, Zeng M, Liu L, Yang X, Shen Y. 2017 Synaptonemal complex protein 2 (SYCP2) mediates the association of the centromere with the synaptonemal complex. *Protein Cell* **8**, 538–543. (doi:10.1007/s13238-016-0354-6)
- Syrjänen JL, Pellegrini L, Davies OR. 2014 A molecular model for the role of SYCP3 in meiotic chromosome organisation. *Elife* **3**, e02963. (doi:10.7554/eLife.02963)

45. 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. (doi:10.1186/s12859-019-3019-7)
46. West AMV *et al.* 2019 A conserved filamentous assembly underlies the structure of the meiotic chromosome axis. *eLife* **8**, e40372. (doi:10.7554/elife.40372)
47. Zickler D, Kleckner N. 2015 Recombination, pairing, and synapsis of homologs during meiosis. *Cold Spring Harb. Perspect. Biol.* **7**, a016626. (doi:10.1101/cshperspect.a016626)
48. Page SL, Hawley RS. 2004 The genetics and molecular biology of the synaptonemal complex. *Annu. Rev. Cell Dev. Biol.* **20**, 525–558. (doi:10.1146/annurev.cellbio.19.111301.155141)
49. Grishaeva TM, Bogdanov YF. 2014 Conservation and variability of synaptonemal complex proteins in phylogenesis of eukaryotes. *Int. J. Evol. Biol.* **2014**, 856230. (doi:10.1155/2014/856230)
50. Offenberger HH, Schalk JA, Meuwissen RL, van Aalderen M, Kester HA, Dietrich AJ, Heyting C. 1998 SCP2: a major protein component of the axial elements of synaptonemal complexes of the rat. *Nucleic Acids Res.* **26**, 2572–2579. (doi:10.1093/nar/26.11.2572)
51. Lorenz A, Wells JL, Pryce DW, Novatchkova M, Eisenhaber F, McFarlane RJ, Loidl J. 2004 *S. pombe* meiotic linear elements contain proteins related to synaptonemal complex components. *J. Cell Sci.* **117**, 3343–3351. (doi:10.1242/jcs.01203)
52. Ferdous M *et al.* 2012 Inter-homolog crossing-over and synapsis in *Arabidopsis* meiosis are dependent on the chromosome axis protein AtASY3. *PLoS Genet.* **8**, e1002507. (doi:10.1371/journal.pgen.1002507)
53. Hatters DM, Peters-Libeu CA, Weisgraber KH. 2006 Apolipoprotein E structure: insights into function. *Trends Biochem. Sci.* **31**, 445–454. (doi:10.1016/j.tibs.2006.06.008)
54. Davis L, Rozalén AE, Moreno S, Smith GR, Martín-Castellanos C. 2008 Rec25 and Rec27, novel linear-element components, link cohesin to meiotic DNA breakage and recombination. *Curr. Biol.* **18**, 849–854. (doi:10.1016/j.cub.2008.05.025)
55. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
56. Burki F, Roger AJ, Brown MW, Simpson AGB. 2019 The new Tree of Eukaryotes. *Trends Ecol. Evol.* **35**, 43–55. (doi:10.1016/j.tree.2019.08.008)
57. Zhou J *et al.* 2015 Accelerated reproductive aging in females lacking a novel centromere protein SYCP2L. *Hum. Mol. Genet.* **24**, 6505–6514. (doi:10.1093/hmg/ddv359)
58. Kouznetsova A, Benavente R, Pastink A, Höög C. 2011 Meiosis in mice without a synaptonemal complex. *PLoS One* **6**, e28255. (doi:10.1371/journal.pone.0028255)
59. Westergaard M, von Wettstein D. 1972 The synaptonemal complex. *Annu. Rev. Genet.* **6**, 71–110. (doi:10.1146/annurev.ge.06.120172.000443)
60. Láscares-Lagunas L, Martínez-García M, Colaiácovo M. 2020 SnapShot: meiosis—prophase I. *Cell* **181**, 1442–1442.e1. (doi:10.1016/j.cell.2020.04.038)
61. Schleiffer A, Kaitna S, Maurer-Stroh S, Glotzer M, Nasmyth K, Eisenhaber F. 2003 Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Mol. Cell* **11**, 571–575. (doi:10.1016/S1097-2765(03)00108-4)
62. Ramesh MA, Malik S-B, Logsdon Jr JM. 2005 A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr. Biol.* **15**, 185–191.
63. van Hooff JJE, Tromer E, van Dam TJP, Kops GJPL, Snel B. 2019 Inferring the evolutionary history of your favorite protein: a guide for molecular biologists. *Bioessays* **41**, e1900006. (doi:10.1002/bies.201900006)
64. Zeeshan M *et al.* 2020 Real-time dynamics of Plasmodium NDC80 reveals unusual modes of chromosome segregation during parasite proliferation. *J. Cell Sci.* **134**, jcs245753. (doi:10.1242/jcs.245753)
65. Lake CM, Hawley RS. 2012 The molecular control of meiotic chromosomal behavior: events in early meiotic prophase in *Drosophila* oocytes. *Annu. Rev. Physiol.* **74**, 425–451. (doi:10.1146/annurev-physiol-020911-153342)
66. Mochizuki K, Novatchkova M, Loidl J. 2008 DNA double-strand breaks, but not crossovers, are required for the reorganization of meiotic nuclei in *Tetrahymena*. *J. Cell Sci.* **121**, 2148–2158. (doi:10.1242/jcs.031799)
67. Chi J, Mahé F, Loidl J, Logsdon J, Dunthorn M. 2014 Meiosis gene inventory of four ciliates reveals the prevalence of a synaptonemal complex-independent crossover pathway. *Mol. Biol. Evol.* **31**, 660–672. (doi:10.1093/molbev/mst258)
68. Tekle YI, Wood FC, Katz LA, Cerón-Romero MA, Gorfu LA. 2017 Amoebozoans are secretly but ancestrally sexual: evidence for sex genes and potential novel crossover pathways in diverse groups of amoebae. *Genome Biol. Evol.* **9**, 375–387.
69. Holloman WK, Schirawski J, Holliday R. 2008 The homologous recombination system of *Ustilago maydis*. *Fungal Genet. Biol.* **45**(Suppl 1), S31–S39. (doi:10.1016/j.fgb.2008.04.006)
70. Fletcher HL. 1981 A search for synaptonemal complexes in *Ustilago maydis*. *J. Cell Sci.* **50**, 171–180.
71. del Cacho E, Pages M, Gallego M, Monteagudo L, Sánchez-Acedo C. 2005 Synaptonemal complex karyotype of *Eimeria tenella*. *Int. J. Parasitol.* **35**, 1445–1451. (doi:10.1016/j.ijpara.2005.06.009)
72. Hollande A, Carruette-Valentin J. 1970 Appariement chromosomique et complexes synaptonématiques dans les noyaux en cours de dépolyploidisation chez *Pyronympha flagellata*: le cycle évolutif des *Pyronymphines* symbiontes de *Reticulitermes lucifugus*. *CR Acad. Sci. Paris* **270**, 2550–2555.
73. Patil S *et al.* 2015 Identification of the meiotic toolkit in diatoms and exploration of meiosis-specific SPO11 and RAD51 homologs in the sexual species *Pseudo-nitzschia multistriata* and *Seminavis robusta*. *BMC Genomics* **16**, 1–21. (doi:10.1186/s12864-015-1983-5)
74. Broadwater S, Scott J, Pobiner B. 1986 Ultrastructure of meiosis in *Dasya baillouiviana* (Rhodophyta). I. Prophase I. *J. Phycol.* **22**, 490–500. (doi:10.1111/j.1529-8817.1986.tb02493.x)
75. Shah S, Chen Y, Bhattacharya D, Chan CX. 2020 Sex in Symbiodiniaceae dinoflagellates: genomic evidence for independent loss of the canonical synaptonemal complex. *Sci. Rep.* **10**, 9792. (doi:10.1038/s41598-020-66429-4)
76. Rog O, Dernburg AF. 2013 Chromosome pairing and synapsis during *Caenorhabditis elegans* meiosis. *Curr. Opin. Cell Biol.* **25**, 349–356. (doi:10.1016/j.ceb.2013.03.003)
77. Peacock L, Bailey M, Carrington M, Gibson W. 2014 Meiosis and haploid gametes in the pathogen *Trypanosoma brucei*. *Curr. Biol.* **24**, 181–186. (doi:10.1016/j.cub.2013.11.044)
78. Rasmussen SW. 1977 The transformation of the synaptonemal complex into the ‘elimination chromatin’ in *Bombyx mori* oocytes. *Chromosoma* **60**, 205–221. (doi:10.1007/BF00329771)
79. Obeso D, Pezza RJ, Dawson D. 2014 Couples, pairs, and clusters: mechanisms and implications of centromere associations in meiosis. *Chromosoma* **123**, 43–55. (doi:10.1007/s00412-013-0439-4)
80. Ebenezer TE *et al.* 2019 Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol.* **17**, 11. (doi:10.1186/s12915-019-0626-8)
81. Cavalier-Smith T. 2010 Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol. Lett.* **6**, 342–345. (doi:10.1098/rsbl.2009.0948)
82. Cavalier-Smith T. 2010 Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biol. Direct* **5**, 7. (doi:10.1186/1745-6150-5-7)
83. Garg SG, Martin WF. 2016 Mitochondria, the cell cycle, and the origin of sex via a syncytial eukaryote common ancestor. *Genome Biol. Evol.* **8**, 1950–1970. (doi:10.1093/gbe/evw136)
84. Cleveland LR. 1947 The origin and evolution of meiosis. *Science* **105**, 287–289. (doi:10.1126/science.105.2724.287)
85. Hurst LD, Nurse P. 1991 A note on the evolution of meiosis. *J. Theor. Biol.* **150**, 561–563. (doi:10.1016/S0022-5193(05)80447-3)
86. Solari AJ. 2002 Primitive forms of meiosis: the possible evolution of meiosis. *Biocell* **26**, 1–13.
87. Wagner A, Whitaker RJ, Krause DJ, Heilers J-H, van Wolferen M, van der Does C, Albers S-V. 2017 Mechanisms of gene flow in archaea. *Nat. Rev. Microbiol.* **15**, 492–501. (doi:10.1038/nrmicro.2017.41)
88. Studier FW. 2005 Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234. (doi:10.1016/j.pep.2005.01.016)
89. Ludzia P, Lowe E, Marciàno G, Mohammed S. 2020 Structural characterisation of KKT4, an unconventional microtubule-binding kinetochore protein. bioRxiv.



90. Chen Z-L *et al.* 2019 A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404. (doi:10.1038/s41467-019-11337-z)
91. Combe CW, Fischer L, Rappsilber J. 2015 xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147. (doi:10.1074/mcp.0114.042259)
92. Perez-Riverol Y *et al.* 2019 The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450. (doi:10.1093/nar/gky1106)
93. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
94. Wheeler TJ, Clements J, Finn RD. 2014 Skyalign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* **15**, 7. (doi:10.1186/1471-2105-15-7)
95. Buchan DWA, Jones DT. 2019 The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407. (doi:10.1093/nar/gkz297)
96. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. 2019 DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **35**, 2790–2795. (doi:10.1093/bioinformatics/bty1062)
97. Dosztányi Z, Csizmok V, Tompa P, Simon I. 2005 IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434. (doi:10.1093/bioinformatics/bti541)
98. Deutekom ES, Vosseberg J, van Dam TJP, Snel B. 2019 Measuring the impact of gene prediction on gene loss estimates in eukaryotes by quantifying falsely inferred absences. *PLoS Comput. Biol.* **15**, e1007301. (doi:10.1371/journal.pcbi.1007301)
99. Richter DJ, Berney C, Strasser JFH, Burki F, de Vargas C. 2020 EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. bioRxiv.
100. Starns D *et al.* 2019 TriTrypDB: A web based resource offering the improvement of structural gene models of pathogenic kinetoplastids through community annotation using Apollo. *F1000Res* **8**, 434.
101. Tashyreva D *et al.* 2018 Phylogeny and morphology of new diplomonads from Japan. *Protist* **169**, 158–179. (doi:10.1016/j.protis.2018.02.001)
102. Lax G, *et al.* 2021 Multigene phylogenetics of euglenids based on single-cell transcriptomics of diverse phagotrophs. *Mol. Phylogenet. Evol.* **159**, 107088. (doi:10.1016/j.ympev.2021.107088)
103. 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. (doi:10.1038/s41598-021-82369-z)
104. Eddy SR. 2011 Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195. (doi:10.1371/journal.pcbi.1002195)
105. Emms DM, Kelly S. 2018 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. (doi:10.1186/s13059-019-1832-y)
106. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015 HMMER web server: 2015 update. *Nucleic Acids Res.* **43**, W30–W38. (doi:10.1093/nar/gkv397)
107. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. (doi:10.1093/bioinformatics/btp348)
108. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
109. Rambaut A. 2012 Figtree v1. 4. Molecular evolution, phylogenetics and epidemiology. See <http://tree.bio.ed.ac.uk/software/figtree>.
110. Tromer EC, Wemys TA, Ludzia P, Waller RF, Akiyoshi B. 2021 Repurposing of synaptonemal complex proteins for kinetochores in kinetoplastida. FigShare.