PNAS
www.pnas.org

Supplementary Information for

Mosaic origin of the eukaryotic kinetochore

Eelco C. Tromer[*#], Jolien J.E. van Hooff[*], Geert J.P.L. Kops[+] and Berend Snel[+#]

[*] equal contribution as first author
[+] equal contribution as senior author
[#] corresponding authors
Email: ecet2@cam.ac.uk and b.snel@uu.nl

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Datasets S1 to S5 (hosted by PNAS)
External Data (hosted by figshare: https://figshare.com/s/884e8848b032f02eb7c3)

Data and Methods

Profile-versus-Profile searches

To find distant homologs of kinetochore proteins, we constructed HMM profiles and used genome-wide databases to apply profile-versus-profile searches. For each of the proteins included in our previous analysis [1], and for some other proteins we studied more recently (Nkp1, Nkp2, Csm1, Lsr4, Mam1, Hrr25 [2]), we aligned the sets of orthologous sequences (MAFFT, v.7.149b [3] 'einsi' or 'linsi') and used these to construct HMM profiles (www.hmmer.org, version HMMER 3.1b1) and HMM-formatted profiles. For each protein, we made such a profile from the full-length alignment. In addition, if a protein has well-annotated domains, we made separate profiles of these domains. While Zwint-1 has two RWD domains, we only used the first as a separate domain profile, because the second is very poorly conserved across species ("SI Appendix, Fig. S2"). All 147 HMM3 profiles can be found in "SI Appendix, External Data: Hidden Markov Models". We applied two different search strategies, using different tools and different search databases. For the first, we searched with full-length profiles of the kinetochore proteins and searched against a database compiled of profiles from PANTHER11.1 [4] and the kinetochore protein profiles themselves. For this search, we made use of PRC (version 1.5.6) [5]. For the second strategy, we searched with domain profiles if available for a given protein, and otherwise full-length profiles. We downloaded scop70 (March 1, 2016), pdb70 (September 14, 2016) and PfamA (version 31.0) profile databases from the HH-suite depository (http://wwwuser.gwdg.de/~compbiol/data/hhsuite/databases/hhsuite_dbs/, downloaded on July 15, 2017) and combined these profiles with the kinetochore domain/full-length profiles we also used as queries. We searched using HHsearch (version 2.0.15) [6]. For each of the search strategies, we identified which of the database profiles correspond to the kinetochore domains/proteins. Using this information, we parsed the search results identifying 'best hits' and 'bidirectional best hits' for each kinetochore domain/protein profile. We made a network using the results from the second (HHsearch-based) search strategy, applying an E-value cut-off of 1 or 10. We visualised this network in Cytoscape (version 3.5.1) [7]. The network can be found in "SI Appendix, External Data: HHsearch Network". In this network, we also added the available cellular component GO terms to the hits using SIFTS [8]. In addition, we used information from the first (PRC-based) search strategy to trace the distant homology between subunits of the Mis12 and NANO complexes ("SI Appendix, Text"). The results of both search strategies can be found in "SI Appendix, Dataset S1".

Phylogenetic trees

For inference of the phylogenetic trees presented in this manuscript, we used a variety of methods. We collected homologs by searching with our tailor-made and Pfam HMM profiles against our local proteome database [1]. The first four letters of the eukaryotic sequences represent the species, of which the full names can be found in "SI Appendix, Table S3". For the phylogenies of Zw10 and related tethering factors, HORMA, histones, RWD and Trip13, we used a subset of the species in this database, including species present in "SI Appendix, Fig. S4". The phylogeny of all eukaryotic kinases was based on sequence-based subsampling, selecting for the slowest-evolving kinases [9]. For the prokaryotic sequences in the UBC/RWD and histone phylogenies, we performed phmmer/jackhmmer online (http://ebi.ac.uk/Tools/hmmer/) and collected sequence hits from the UniProt database. For the HORMA and Trip13 phylogenies, we searched with the bacterial sequences reported by Burroughs et al. [10], from species possessing 'Bacterial-HORMA2' and 'Bacterial-HORMA1'. We added contig information for species not included by Burroughs et al. [10] ("SI Appendix, Dataset S5"). For all protein families, multiple sequence alignments were inferred using MAFFT (v.7.149b [3], 'einsi' or 'linsi') [11], and trimmed with trimAl

the function 'merge' of MAFFT (ginsi, unalignlevel 0.6). We manually scrutinized the resulting multiple sequence alignments ("SI Appendix, Dataset S3, Dataset S4") for clear misalignments based on structure-based alignments of available UBC/RWD and histones domains ("SI Appendix, Dataset S2, External Data: RWD/UBC, histone structures"). Of note: due to extensive sequence divergence, we did not include Atg7, Atg10, Med14, Med17 and Fancl-1 to the UBC/RWD phylogenetic analysis. Trees were made using RAxML (version 8.0.20, automatic substitution model selection, GAMMA model of rate heterogeneity, rapid bootstrap analysis of 100 replicates) [13] and/or IQ-TREE (version 1.6.3, extended model selection, ultrafast bootstrap (1000) and SH-like approximate likelihood ratio test) [14]. Trees were visualised and annotated using FigTree [15].

Structural similarity and secondary structure prediction

To identify potential homologs based on structural similarity with LECA kinetochore proteins, we searched both the literature and databases such as PFAM (http://pfam.xfam.org [16]), ECOD (http://prodata.swmed.edu/ecod/ [17], RCSB Protein Data Bank (http://rcsb.org [18]) and CATH (http://www.cathdb.info/ [19]). All databases were consulted between January and March 2019. Structures were visualized and processed using the python-based software package Pymol version 2.1.1 [20]. Structural alignments were performed using either 'cealign' and 'super' or were directly downloaded from the aforementioned databases and/or the DALI webserver [21]. All-versus-all structural similarity Z-scores of relevant candidate homologs, were calculated using the algorithms implemented DALI webserver [22]. An overview of the Z-scores and various hyperlinks to databases that we consulted, can be found in ("SI Appendix, Dataset S2"). A structure was considered significantly similar with a Z-score of 2.0 or higher. Average (hierarchical) linkage clustering of the DALI Z-scores was used to infer evolutionary scenarios for TBP-like and UBC/RWD (see Fig. 2 and 3). Pymol session files, containing most of the structures used for the comparison of RWD/UBC-like proteins, TBP-like domains, histones and Mis12/NANO-like are made available ("SI Appendix, External Data: RWD/UBC, TBP-like, histones, Mis12/NANO structures"). Secondary structure predictions for Zwint-1 were performed using the JPRED webserver [23], embedded in the alignment package Jalview [24].

Classifications and interpretations of homologous protein families

We identified closest homologs of kinetochore proteins based on phylogenetic trees, profile-versus-profile searches and structural similarity. We classified these closest homologs as either eukaryotic or prokaryotic. A closest eukaryotic homolog is a paralog resulting from a gene duplication before LECA [25]. In this, we distinguished kinetochore paralogs from paralogs involved in other eukaryotic cellular processes. If a protein has a kinetochore protein as its closest paralog, likely their ancestral, pre-duplication protein was already part of the primordial kinetochore. Likewise, if more than two kinetochore proteins are most closely related to one another, this group of proteins likely resulted from multiple successive duplications of a single pre-duplication kinetochore protein. We refer to this ancestral kinetochore protein as 'anc_KT' for 'ancestral kinetochore unit' ("SI Appendix, Table S1"). Hence, the 'anc_KT' is the protein that got involved in the kinetochore, and then duplicated to give rise to the paralogous kinetochore proteins. This ancestral kinetochore protein might have had also closest homolog (either eukaryotic or prokaryotic) outside of the kinetochore, which we also identified. If a LECA kinetochore protein has no closest paralog in the kinetochore, the protein itself forms the ancestral kinetochore unit.

Text

Detecting kinetochore homologs using different resources

To complete our picture of the origin of kinetochore proteins, we made use of four different sources of information: phylogenetic trees, (HMM) profile-versus-profile searches and structural information, supplemented with literature-curated evolutionary links of kinetochore proteins when available. These different information types result in different qualifications of relationships between pairs of (suspected) homologs ("SI Appendix, Table S1"). In general, the profile-versus-profile searches were in agreement with the relationships observed from the phylogenies ("SI Appendix, Dataset S1, Fig. S1). However, we noted that various kinetochore proteins, such as RWD proteins and Cep57, hit coiled-coil proteins. In general, the hits we identified as likely coiled-coil were ignored, because coiled-coil similarity might not be indicative of homology, since it could also evolve convergently [26].

Proteins in LECA kinetochore & alternative eukaryotic roots

To determine which proteins were present in the LECA kinetochore (Fig. 1B), we first inferred for each protein if it was likely encoded in the genome of LECA ("SI Appendix, Table S2"). In principle, we did so based on Dollo parsimony, which states that a protein can only be invented once, hence the origin of the protein dates back to the last common ancestor of all species that have it. In applying this approach, we assume that the divergence between Opimoda (Opisthokonta and Amoebozoa) and Diphoda (Stramenopila-Alveolata-Rhizaria (SAR), Archaeplastida, Excavata) represents the root of the eukaryotes ("SI Appendix, Fig. S4") [27]. While we are well aware of the controversies about the position of the eukaryotic root [28], we think that alternative rootings would not dramatically alter our model of the LECA kinetochore. If for example the root actually lies between (a subset of) Excavata and all other eukaryotes, such as proposed by [29], the presences of kinetochore proteins in this lineage would also support their presence in LECA. This is the case for the CCAN subunits ('Cenp' proteins, Nkp1 and Nkp2, Fig. 1B), because one of the Excavata species (*Trichomonas vaginalis*) contains CenpX and CenpS [30]. These proteins likely result from duplications, and their closest paralogs are CenpW and CenpT, respectively (Fig. 3). Hence, the common ancestor of the Excavata and the other eukaryotes (LECA) likely had these four CCAN components. Furthermore, the recently sequenced genome of the amitochondriate *Monocercomonoides sp. PA203*, an excavate species distantly related to e.g. *Trichomonas vaginalis*, contains four additional CCAN subunits: CenpI, CenpK, CenpL and CenpN (unpublished data). Given that the CCAN subunits strongly co-evolve, likely LECA had the complete CCAN, also under this alternative root. The Dam1 complex would have been inferred to have been present in LECA based on Dollo parsimony, but we think it is very likely that its genes were invented later in evolution and got horizontally transferred among distantly related eukaryotic species [31]. Nkp1 and Nkp2 would not have been found to have been present in LECA under Dollo parsimony. However, we argue that, because they are homologous to subunits of the Mis12 complex (Fig. 4), and because these Mis12 complex subunits are present across the eukaryotic tree of life, Nkp1 and Nkp2 likely resulted from ancient duplications before LECA, giving rise to Nkp1 and Mis12 (Mis12 complex), and to Nkp2 and Nnf1 (Mis12 complex). We infer that Nkp1 and Nkp2 were lost in major eukaryotic lineages quickly after LECA, since we do not observe them in most Diphoda lineages ("SI Appendix, Fig. S4"). Moreover, Nkp1 and Nkp2 are part of the CCAN, which strongly co-evolves, including Nkp1 and Nkp2 [2, 30]. If a protein likely was encoded by LECA, we in principle hypothesize it to be part of the LECA kinetochore. We nevertheless exclude such a protein from the LECA kinetochore if it depends on a non-LECA protein for the kinetochore function (Hrr25), or if it seems more likely to be involved in another process, as indicated by characterizations in multiple species (Skp1). The complete list of kinetochore proteins and considerations for in/excluding them as part of the LECA kinetochore can be found in "SI Appendix, Table S2".

Kinetoplastid kinetochore proteins where not part of the LECA kinetochore

We did not analyze the origin of the non-conventional kinetochore proteins specific to kinetoplastids (KKT/KKIPs) [32, 33] because our reconstructions do not infer these proteins to have been part of the LECA kinetochore, Here we want to briefly explain why we infer a LECA kinetochore without any of the KKT/KKIPs. Most importantly all recent research on the eukaryotic tree of life places kinetoplastids next to euglenids, symbiontids and diplonemids in the group Euglenozoa [34], itself part of the excavate group Discoba, which also includes *Naegleria* [27–29, 35, 36]. Euglenids, such as *Euglena gracilis*, appear to possess 'conventional' kinetochore proteins, and no kinetoplastid kinetochore proteins [37]. *Naegleria* also harbors proteins for a nearly complete conventional kinetochore. These two observations make it the most likely scenario that (1) the kinetoplastid kinetochore is derived from an euglenozoan ancestor with a conventional kinetochore, and that (2) the common ancestor of Discoba connects to the remainder of the eukaryotes with a conventional set of kinetochore proteins. Even if, as older work has suggested, the root lies between Euglenozoa and all other eukaryotes [38], this eukaryotic phylogeny would still most parsimonously infer the absence of the kinetoplastid kinetochore proteins (KKT/KKIP proteins) from LECA, and the presence of 'conventional' kinetochore proteins in LECA. The only potential argument for considering kinetoplastid kinetochore proteins as candidates for being part of the LECA kinetochore would be that kinetoplastids diverged first from all of the other eukaryotes. However, since a body of phylogenetic research stably places kinetoplastids within the Euglenozoa, and the Euglenozoa seem to be stably part of the Discoba, we do not think this argument will hold. We are aware that a euglenozoan ancestor with a conventional kinetochore implies that during evolution it was replaced by the kinetoplastid-like kinetochore in a major evolutionary transition. Such transitions are not unique, as the LECA kinetochore itself demonstrates: this structure after all was absent from the prokaryotic ancestors of eukaryotes. And the uniqueness or presence of a transition is not an argument for or against redefining the phylogenetic tree.

Mis12/NANO-like proteins

In this study, we present the subunits of the Mis12 complex (Mis12, Nnf1, Dsn1, Nsl1) and of the NANO complex (Nkp1, Ame1/CenpU, Nkp2, Okp1/CenpQ) to be all homologous to one another, having a domain we coin "Mis12/NANO" (Fig. 1, Fig. 4). We chose to use the new complex name 'NANO' instead of COMA since the COMA complex contains CenpO and CenpP as well, which are not present in the tetramer consisting of Nkp1, CenpU, Nkp2 and CenpQ. We inferred the homology of the Mis12 and NANO complex using different sources. First, Nnf1 and Nkp2 appear homologous by being each other's bidirectional best hit in the profile-versus-profile output ("SI Appendix, Dataset S1", HHsearch E-value 10). The Nkp1 and Mis12 full-length profiles hit each other best with PRC ("SI Appendix, Dataset S1"; PRC E-value 10). In the same search, the Nkp2 profile hit the Nnf1 profile, and the Nsl1 profile hit the Mis12 profile. Moreover, the profile of Nkp1 hits that of Mis12 in HHpred online [39], albeit at very high E-value: 180. The structures of the Mis12 subunits were already shown to be similar [40, 41], therefore we propose their homology. A recent study revealed the striking structural similarity of Nkp1, Nkp2, CenpQ and CenpU to the Mis12 complex [42]. It therefore seems plausible to assume that the four-member NANO complex is homologous to the Mis12 (Fig. 4). Structural comparisons did not reveal any particular higher order similarity between the different subunits of both the Mis12 and/or NANO complex ("SI Appendix, Dataset S2"). Given the similar relative position within the complexes and the topology of the helical elements (coiled-coil and helices) of each of the four subunits per complex, we propose the Mis12 and NANO complex arose through an ancestral whole complex duplication before LECA (Fig. 4D). Furthermore, the Mis12-Nnf1 and Nkp1-Nkp2 dimer have a similar long 'head domain' and have significantly similar sequences (see above), suggesting that the Dsn1-Nsl1 and CenpU-CenpQ dimer might have a common evolutionary origin as well. Since we have no indications for

other (prokaryotic or eukaryotic) homologs, we infer that the Mis12/NANO domain was invented during the FECA-to- LECA transition and gave rise to these eight kinetochore proteins via gene duplications (see above).

Double RWD domain in Zwint-1 orthologs

In our sensitive profile-versus-profile analysis, various kinetochore RWD proteins hit each other, as well as other RWD-like and E2/UBC proteins, indicating that their sequences were sufficiently similar to confirm their homology, with the notable exceptions of the RWD domains of CenpP (full-length profiles are hit, see "SI Appendix, Dataset S1"). Interestingly, Zwint-1, the only KMN (Knl1-Mis12-Ndc80) network subunit for which a structure has not yet been determined [43], was hit by various RWD HMM profiles. Indeed, upon further inspection of the predicted secondary structure of Zwint-1 orthologs, we found that they follow a classic tandem RWD topology ("SI Appendix, Fig. S2"), similar to its direct interaction partner Knl1 and to the CenpO-CenpP dimer. Interestingly, many Zwint-1 orthologs show degeneration of the C-terminus, resulting in the loss of either the second or both RWD domains; making this a protein family that is particularly hard to predict orthologs for (see also van Hooff et al. [1]). Since all kinetochore RWD proteins form dimers through RWD-RWD interactions and since the main interactor of Zwint-1 is the double RWD protein Knl1, we predict that Zwint-1 is a bona fide double RWD kinetochore protein.

RWD/UBC evolution

Since bona fide catalytic UBCs were found in both Bacteria and Archaea, ubiquitin-like modification was likely the ancestral function of this fold in FECA [44–47]. Both catalytic [48] and non-catalytic [44] families of the UBC superfamily expanded extensively between FECA and LECA. Functionally, the non-catalytic UBC-like proteins comprise four major groups (see Fig. 2, "SI Appendix, Fig. S1E, Fig. S3, Dataset S2"): (1) ubiquitin-related non-enzymatic vesicle trafficking proteins like Uev1/Tsg101[49] and Aktip [50], (2) E2/E3-related canonical RWD proteins (RWD) involved in DNA/RNA-related processes (e.g. Gcn2 [51]), (3) orphan RWD/UBC-like proteins like FancL [52] and subunits of the Mediator complex [53], and (4) eight LECA kinetochore proteins that form hetero- or homodimers, with either a single RWD: Spc24-Spc25, Mad1-Mad1, Csm1-Csm1, or a double RWD configuration: CenpO-CenpP and Knl1-Zwint-1 (Zwint-1 is also a RWD protein, see 'Double RWD domain in Zwint-1 orthologs'). Due to the highly divergent sequence evolution of bacterial UBC-like proteins, kinetochore RWD proteins and other non-catalytic UBCs (RWD and for instance Uev1/Tsg101), we could only construct a short alignment (90 positions, minimal 30% column occupancy) for the whole UBC family ("SI Appendix, Dataset S3"). Our structural similarity searches indicated that the mediator subunits Med14, Med17 [44], the first RWD-like domain of FancL and the autophagy-related enzymes Atg7 and Atg10 [44] were part of the RWD/UBC superfamily as well ("SI Appendix, Dataset S2"). We could however not align them in a reliable manner yielding a sensible phylogenetic tree and we therefore left them out of the phylogenetic analysis and only considered their structural similarity when devising an evolutionary scenario (Fig. 2D, "SI Appendix, Fig. 3"). All in all, we performed the phylogenetic analysis using a limited amount of phylogenetic informative positions. Therefore, the parameters for the maximum likelihood methods are potentially overfitted and the resulting phylogenetic tree is therefore likely subject to artefacts such long-branch attraction (see for instance 'divergent UBC' and 'bacteria_E'), resulting in the likely misplacement of some divergent branches and distortion of the overall tree topology. Nonetheless, overall the eukaryotic UBC superfamily evolved into two distinct groups in eukaryotes: E2 ubiquitin conjugases (UBC, bootstrap:77/100) and two non-catalytic UBC-like groups (RWD and kinetochore RWD; bootstrap:96/100). The inconsistent placement of the second RWD domain of Knl1 and Zwint-1 (Knl1-2 and Zwint-1-2), precluded the conclusive inference for a single origin of kinetochore RWD domains. Whether this means that Knl1-2 and Zwint-1-2 were independently acquired compared to CenpO-2 and CenpP-2 or signify a shared and more complex origin of RWD domains is unclear. Likely the origin of kinetochore RWD domains is closely related to FancL, another

double RWD protein that found in eukaryotic genomes, and subunits of the Mediator complex, Med14-17 (structural similarity) and Med15, for which we uncovered the presence of a C-terimnal RWD domain in this study. How bacterial UBCs are related to archaeal and eukaryotic UBCs and RWDs is not clear from our studies. Many different classes of modification systems that operate UBC-like folds are also present in Bacteria (bact_UBC_A-D) and even a non-catalytic domain is found in some lineages (bact_UBC_E) [44]. In addition, a number of bacterial UBCs cluster together with archaeal and eukaryotic UBCs. Likely, bacterial UBCs represent an ancient protein modification system (bacteria_UBC_A-E) that has been optimized in archaeal lineage that are closely associated with FECA. Concomitantly, our trees suggest multiple horizontal gene transfers of archaeal/eukaryotic UBCs to bacteria. Various non-catalytic UBC proteins can be found in the Asgard Archaea (closest to the archaeal ancestor of eukaryotes, Fig. 1A) [45], which are phylogenetically affiliated with both RWD and UBC ("SI Appendix, Fig. S1E, Fig. 3"). Therefore, an RWD-like protein could have already been present in the archaeal ancestor of eukaryotes. Given that UBCs of archaeal descent extensively radiated in eukaryotes [48], we think it is likely that RWD and kinetochore RWD were also part of this radiation and are thus of archaeal descent.

Reconstruction of histone fold evolution

Although the canonical nucleosomal histones are amongst the most highly conserved eukaryotic proteins at the amino acid level, most other histones in eukaryotes are highly divergent, including the TFIID-related TBP-associated factors (TAFs), SAGA-related proteins (SUPTs), CCAAT-binding complex/nuclear transcription factor (CBF/NF), Negative coregulator 2 (NC2), subunits of the DNA polymerase epsilon (DPOE), Chromatin Accessibility Complex (CHRAC) and the kinetochore histones (CenpA, CenpS, CenpT, CenpX, CenpW) ("SI Appendix, Fig. S1I"). To produce an informative alignment, we first made separate alignments of slowly evolving orthologs (manually curated) of each of the LECA histone proteins, which we subsequently aligned (corrected based on structural alignments, see "SI Appendix, Data and Methods, Dataset S2, Dataset S4, External Data"). In addition, we added archaeal and bacterial histone-like sequences, which we acquired through jackhmmer runs against archaeal and bacterial UniProt databases online (see "SI Appendix, Data and Methods"), using known archaeal histones such as HMf [54], the reported Asgard histone-like proteins [45] and the 'DUF1931' protein family as queries. Due to the limited amount of positions (69) and highly divergent nature of the histone family, we could not fully resolve histone evolution as different algorithms (RAxML and IQTREE) and various models gave inconsistent results: (1) a number of different histone groups (H2B, Taf3/8/Supt7, Taf12) appeared at different positions in the tree, (2) the duplication order within for instance the TAF clade was often different with various low bootstrap support values, (3) the position of bacterial and archaeal taxa varied, and (4) the exact placement of CenpX and CenpW relative to each other changed (further detailed below). We here present one of the trees in which CenpX and CenpW are each other's closest paralog ("SI Appendix, Fig. S1I"). In archaea and bacteria, histones are found that are affliated to different eukaryotic histone groups. The presence of a high number of bacterial histone-like proteins surprised us. Although our analyses did not give a consistent result, it is likely that histone-like proteins in Bacteria were acquired through horizontal gene transfers from either archaeal or eukaryotic lineages. In general, we observe that CenpX and CenpW cluster together with one of two major histone groups: (1) Taf11, H2A, NFY, NC2, DPOE and Chrac1, while CenpS, CenpT and CenpA are more similar to (2) H2B, H3, H4 and all the other TAFs and Supts (not Taf11). The duplication of CenpA and H3 is not always supported and in a number of cases CenpA branches from within the H3 clade. The duplication of CenpT and CenpS is overall well supported (boostrap:87-99/100). The position of CenpX and CenpW varied. The various trees (using different evolutionary models) suggested a closest paralog for CenpX, i.e. Taf11, H2A, NC2A and a bacterial clade, and for CenpW, i.e. Taf12 or NC2A/NFY. Apart from the duplication of the kinetochore histone dimer partners, the NFYA-NFYB, NC2A-NC2B and Dpoe3-Doe4/Chrac1 histones seem to originate from an internal duplication as well. The order of duplication

for the TFIID and SAGA complex TAF/SUPTs could however not be easily reconciled with the known dimer pairs (see "SI Appendix, Fig. S1I": heterodimers are Taf3/8-Taf10, Taf4-Taf12, Taf6-Taf9, Taf11-Taf13, Supt7-Supt3, Supt3-Taf10 and Ada1-Taf12).

Figure S1. Phylogenetic trees of domains present in kinetochore proteins.

Kinetochore proteins are indicated in green, closest homologs (e.g. eukaryotic closest paralogs). Details on tree inference methods can be found in "SI Appendix, Text, Data and Methods". The sequences were obtained from our local proteome database in combination with bacterial and archaeal entries from the Uniprot database, harvest through online jackhmmer tool (see "SI Appendix Data and Methods"). The first four letters of the protein name indicate the species as listed in "SI Appendix, Table S3". Due the size of the tree, the IDs in E (RWD) and I (Histones) are not displayed.

(A) Vps51-like tethering complex subunits (Zw10)
(B) WD40 (Cdc20)
(C) WD40 (Bub3)
(D) kinases (MadBub, Mps1, Aurora, Plk)
(E) RWD (Zwint-1, Knl1, CenpO, CenpP, Spc24, Spc25, Mad1, Csm1)
(F) HORMA (Mad2, p31^{comet})
(G) AAA+ ATPase (Trip13)
(H) WD40-NRH-Sec39 (Rod)
(I) histones (CenpA, CenpT, CenpS, CenpX, CenpW) – support is shown for internal nodes that correspond to the origin of new orthologous groups up to the LECA level.

Figure S1A. Vps51-like tethering complex subunits (Zw10)

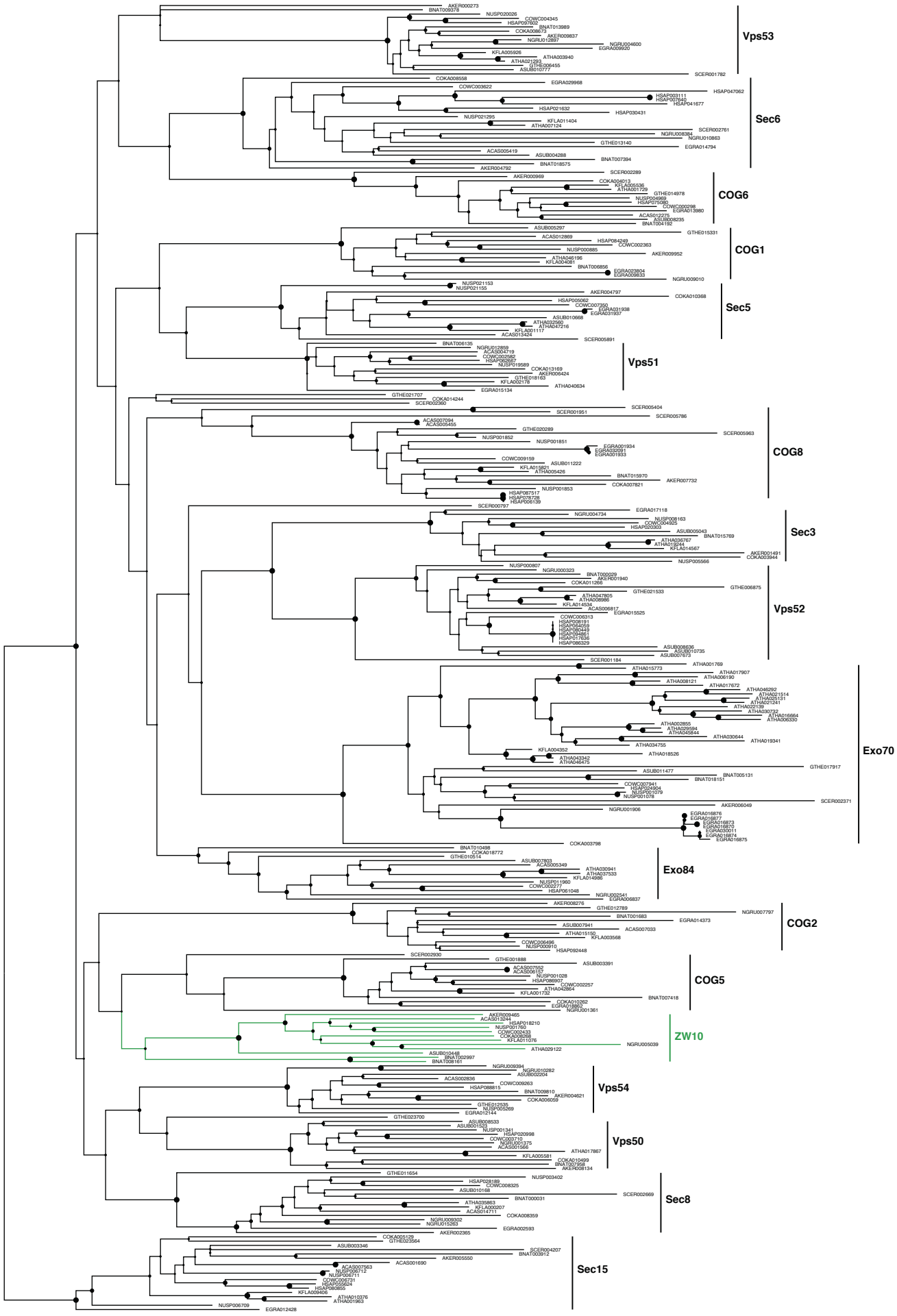


Figure S1B. WD40 (Cdc20)

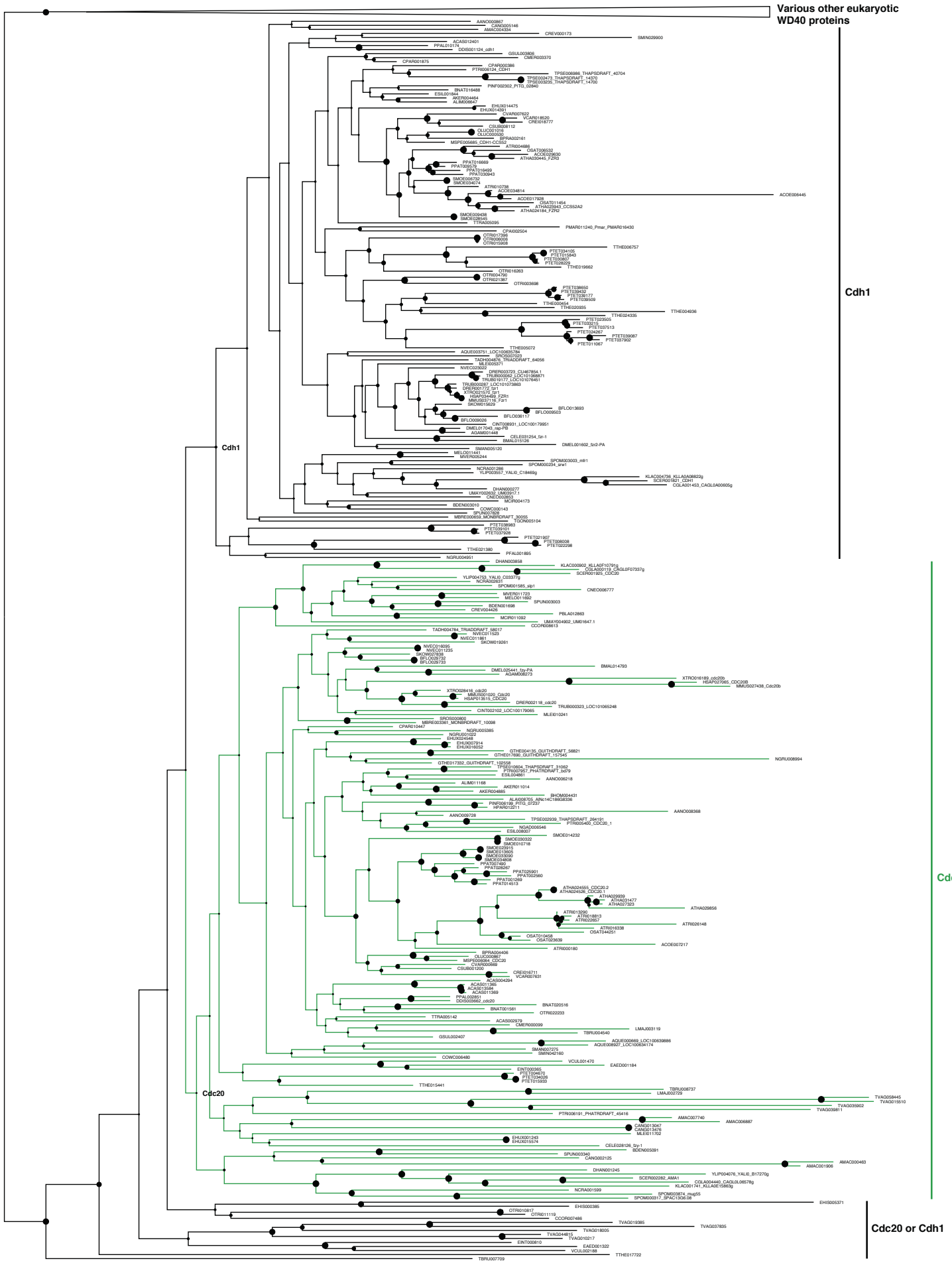
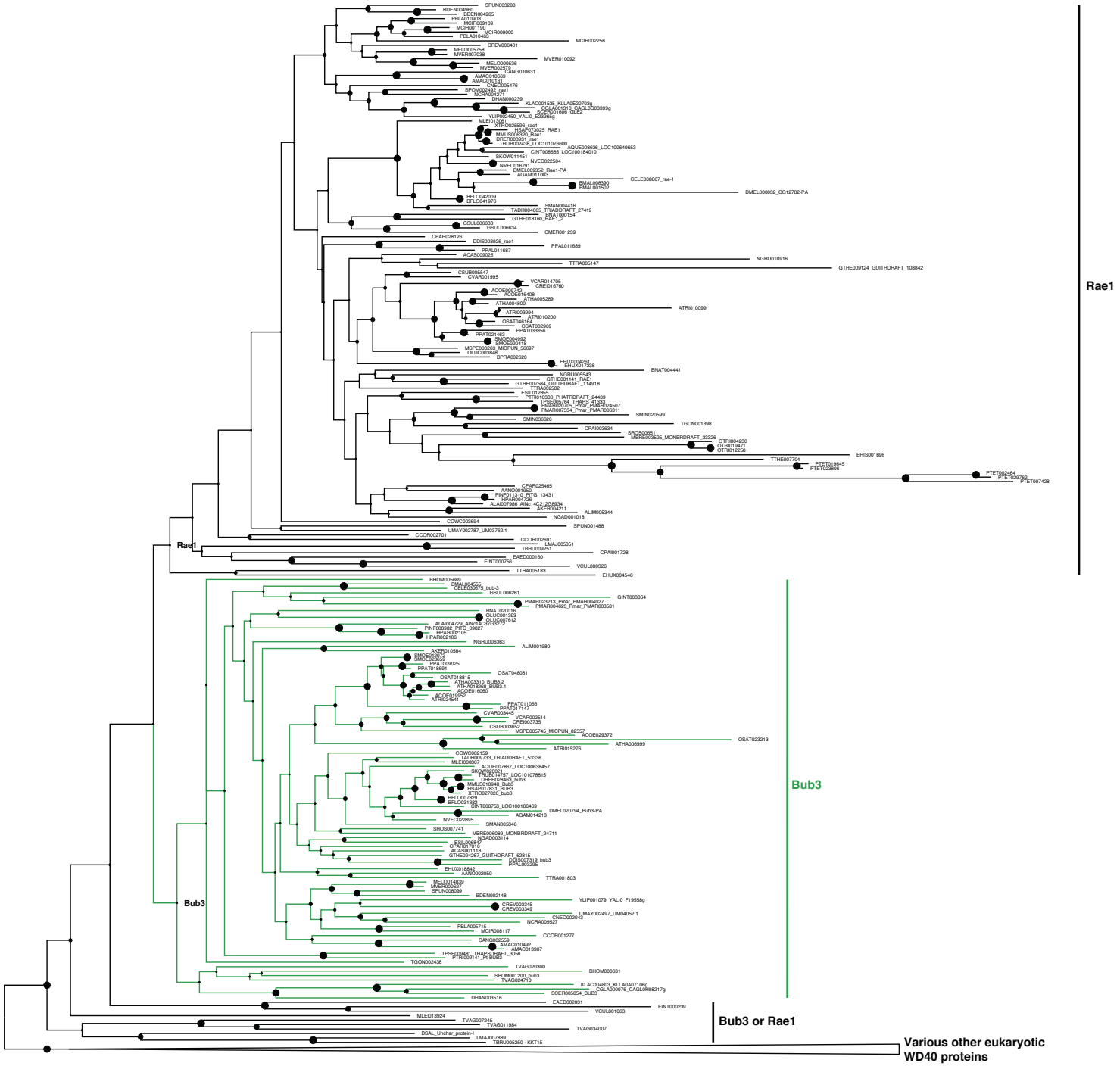


Figure S1C. WD40 (Bub3)



Support (rapid bootstraps)

0.4

Model: LG + G (RAxML)

Figure S1D. kinases (MadBub, Mps1, Aurora, Plk)

Support (rapid bootstraps)

0.8

Model: LG + G (RAxML)

Figure S1E. RWD (Zwint-1, Knl1, CenpO, CenpP, Spc24, Spc25, Mad1, Csm1)

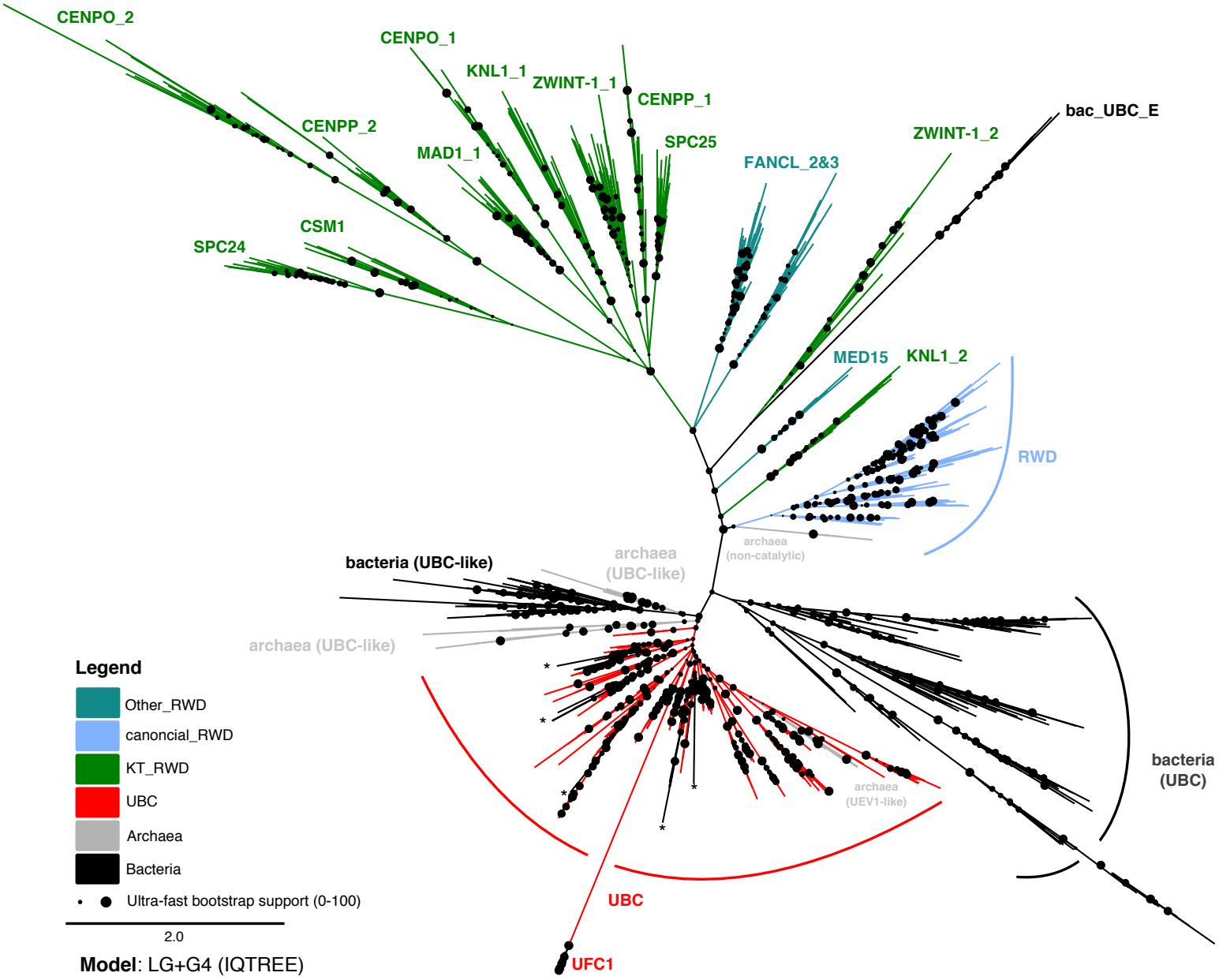


Figure S1F HORMA (Mad2, p31^{comet})

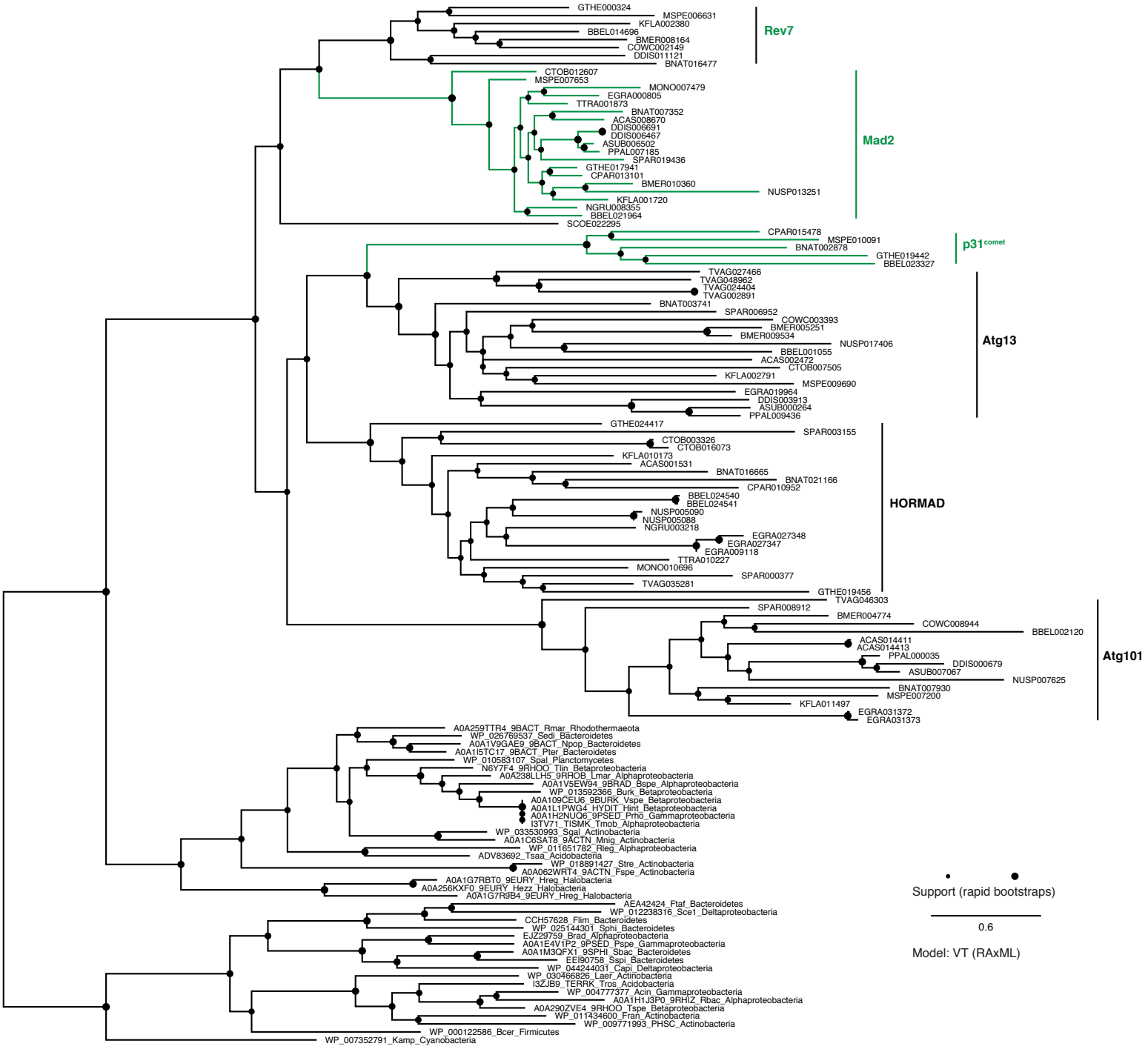
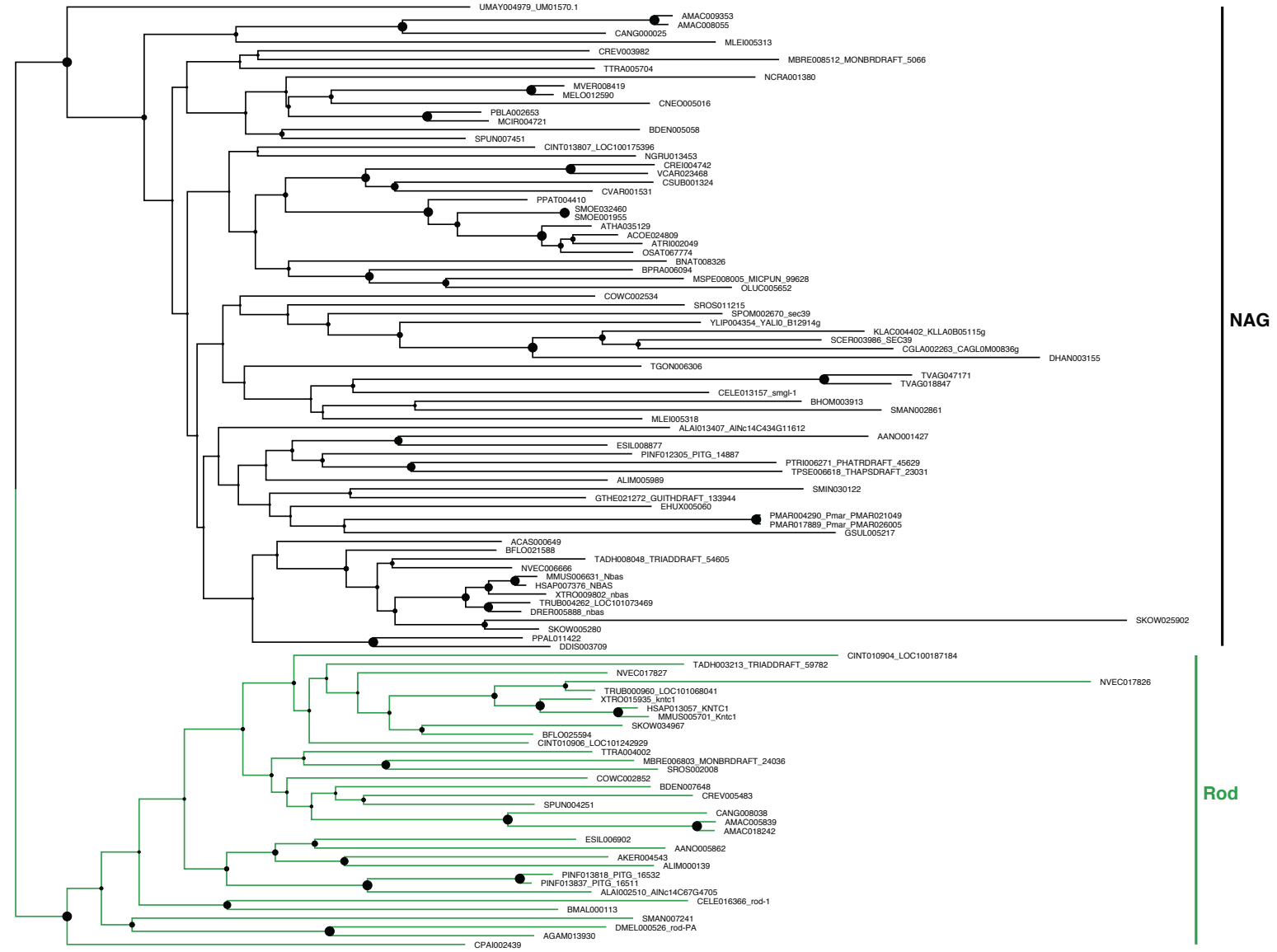


Figure S1G. AAA+ ATPase (Trip13)

Figure S1H. WD40-NRH-Sec39 (Rod)

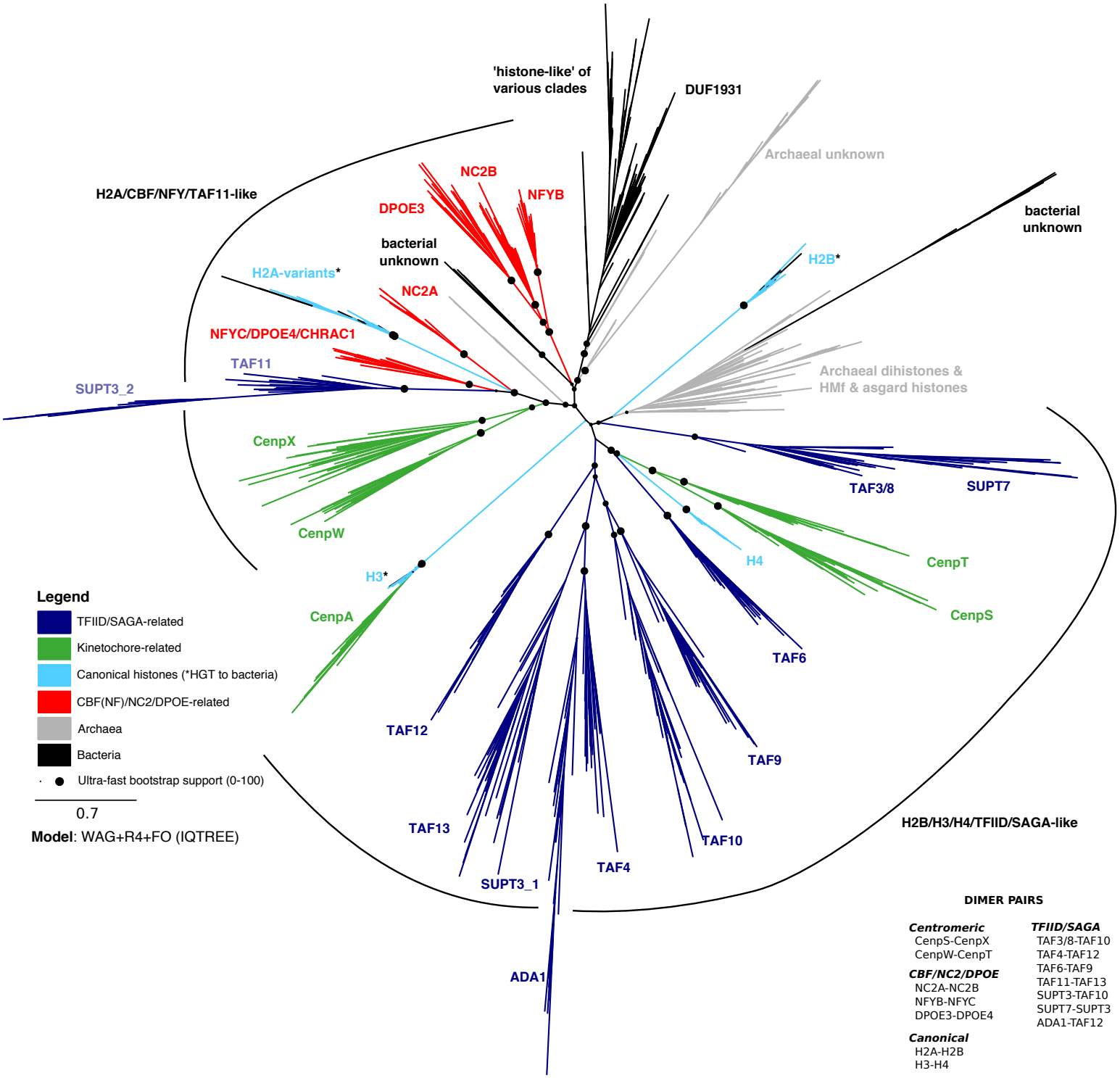


Support (rapid bootstraps)

0.5

Model: LG + G (RAxML)

Figure S1I. histones (CenpA, CenpT, CenpS, CenpX, CenpW)



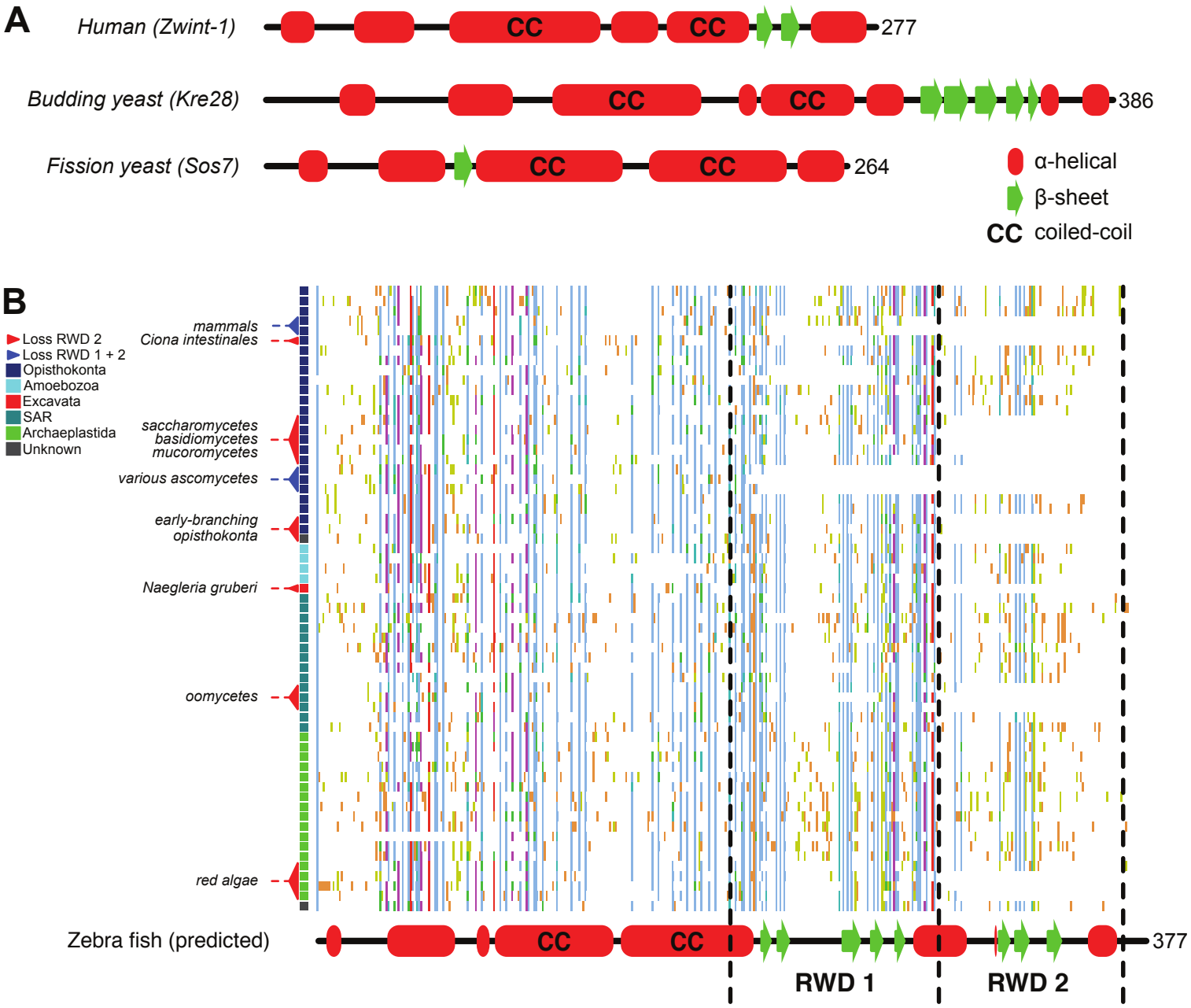


Figure S2. Recurrent loss of RWD domains during the evolution of Zwint-1.

(A) Secondary structure prediction of three Zwint-1 orthologs: Zwint-1 (human), Kre28 (budding yeast) and Sos7 (fission yeast), reveal a highly divergent C-terminal region. **(B)** Overview of a multiple sequence alignment of 83 Zwint-1 orthologs that was based on the de novo predicted Zwint-like ortholog in zebra fish (*Danio rerio*). The colors of the alignment are based on the classic Clustal coloring scheme and the alignment is condensed so that the letters of the residues are not visible anymore. Note that the second RWD domain is more degenerated and therefore less conserved throughout the eukaryotic lineages that have a Zwint-1 ortholog. On the left the small colored blocks indicate the supergroup to which each species belongs. The red and blue clades indicate which of the orthologs lost the RWD-2 or both the RWD-1 and RWD-2 domain respectively.

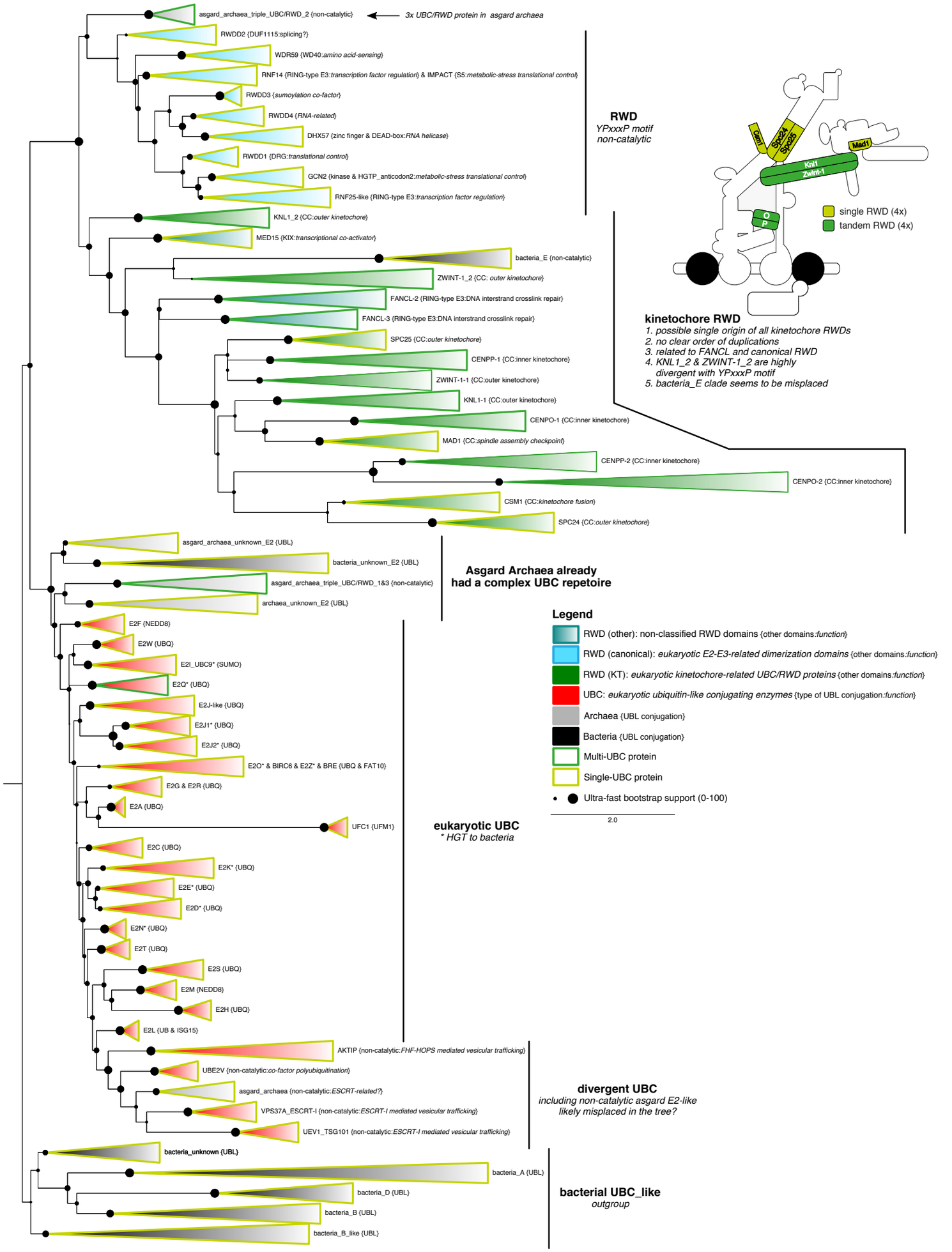


Figure S3. UBC superfamily evolution

Annotated phylogenetic tree of the UBC superfamily (based on "SI Appendix, Fig. S1E"). '{}' denote other domains present and general function for RWD proteins (kinetochore [KT] & other), while for UBC-like proteins it signifies potential substrates. See "SI Appendix, Text – RWD/UBC evolution" for discussion. In short: : the RWD and UBC proteins in eukaryotes are likely descendant from a bona fide ubiquitin-like modification system [47] that can be found in lineages that are phylogenetically affiliated with the archaeal ancestor of eukaryotes. Subsequent duplications and sub/neofunctionalization can be classified into four groups: (1) UBC (E2 ubiquitin conjugases), (2) canonical RWD ('other', E2/E3-associated proteins), and (3) orphan non-canonical RWD/UBC-like proteins and (4) RWD kinetochore proteins (KT), which are likely the closest homolog to the triple RWD protein FancL and members of the Mediator complex (Med14, Med15 and Med17). The latter three RWD-like groups are likely highly related. Note the presence of a RWD-like domain in an Asgard protein with 3x RWD/UBC domains. Note that the RWD-like domains FancL-1, Med14-(1-2-3), Med17 and the UBC-like enzymes Atg7/Atg10 are not included in this tree ("SI Appendix, Text"). The stroke in light/dark green correspond to the RWD configuration present in each UBC/RWD group (single versus multiple).

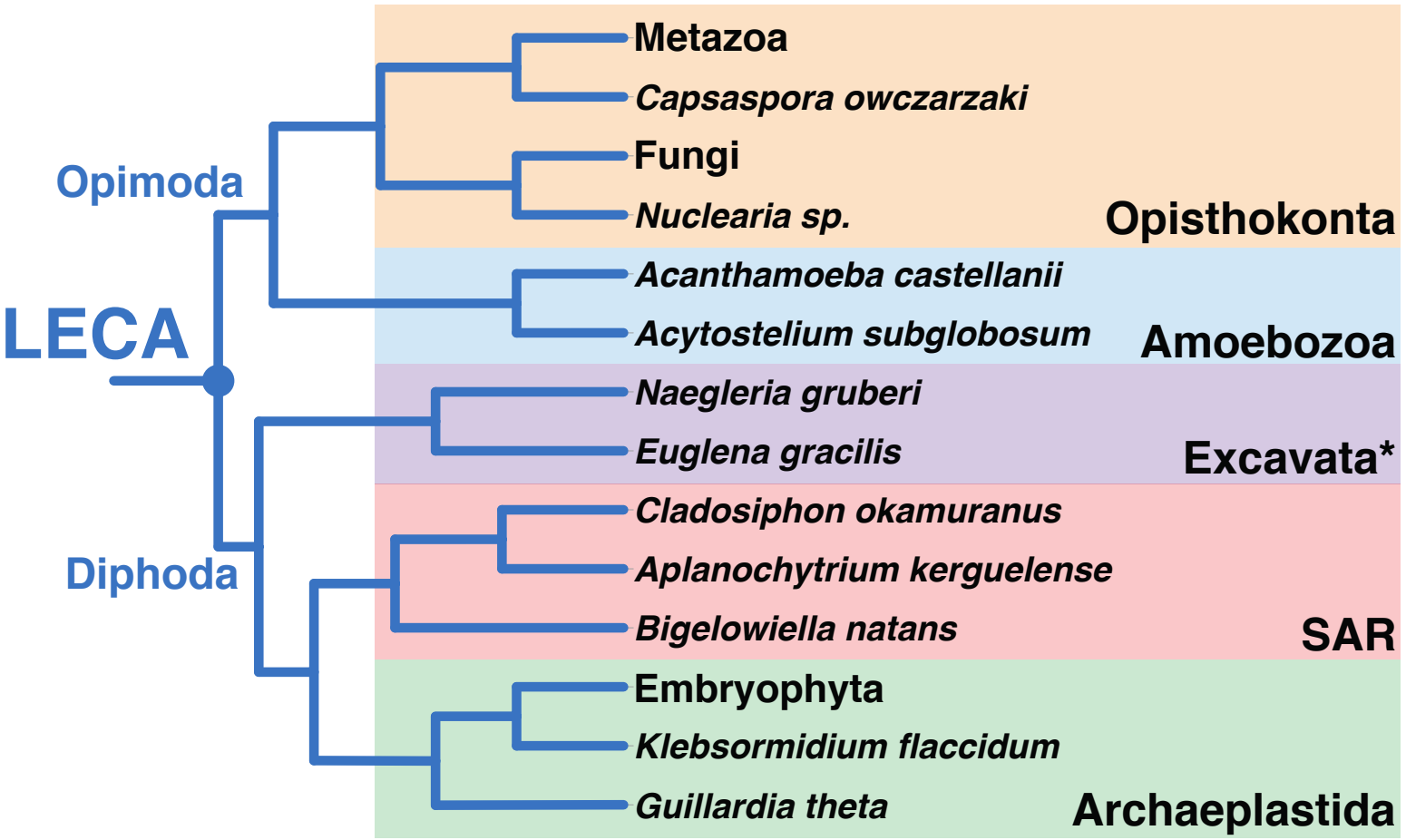


Figure S4. Phylogeny of eukaryotes.
Cartoon of the eukaryotic species tree with Opimoda-Diphoda root and eukaryotic supergroups [27]. This topology was used to infer whether a protein was likely present in LECA ("SI Appendix, Table S2", Fig. 1B). * = Discoba + Metamonads, but excluding Malawimonas.

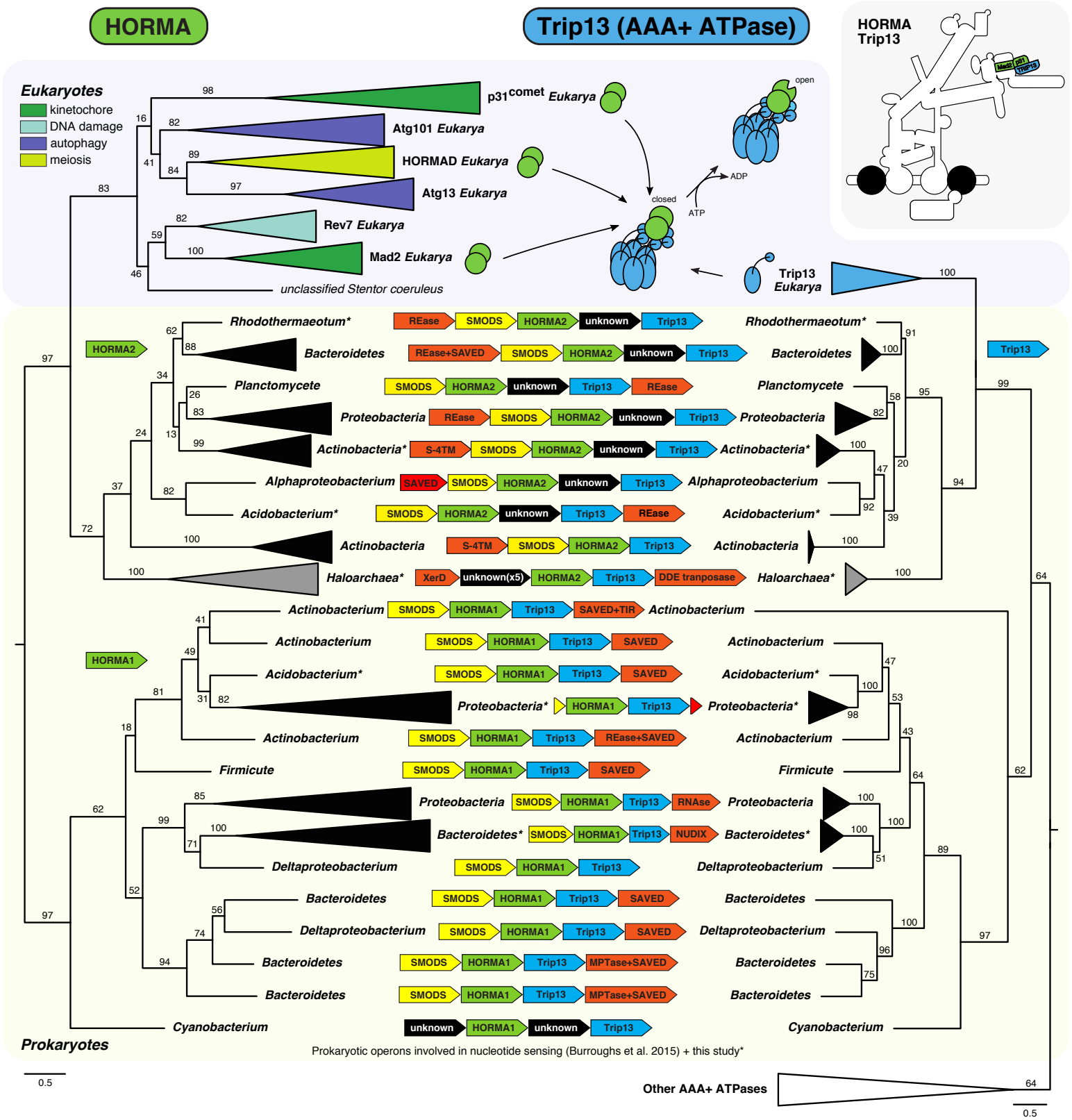


Figure S5. The HORMA-Trip13 module likely has a prokaryotic origin.
Full phylogenetic trees of HORMA domain proteins and AAA+ ATPases. In eukaryotes, HORMAD, Mad2 and p31comet are structurally modified (open/closed) by a Trip13 hexamer (upper panel, right side). In prokaryotes, HORMA (two types: 1, 2) and Trip13 are present in a single operon, strongly suggesting that they also interact in these species and thus that this interaction is ancient. The phylogenetic trees indeed suggest that the eukaryotic HORMA domain and Trip13 were derived from prokaryotes. In addition, the prokaryotic operons include proteins that are involved in nucleotide signaling (yellow: nucleotide transferase [SMODS], red: tranposase-related [SAVED and others], and black: unknown), signifying that prokaryotic HORMA and Trip13 are affiliated to this process [10]. The uncollapsed trees can be found in "SI Appendix, Fig. S1F, S1G". Asterisks indicate the species for which we discovered a HORMA-Trip13 operon (see Dataset S5 for annotation). The collapsed red and yellow domains in Proteobacteria (HORMA1) indicate the presence of various different domains in the operons amongst these lineages (see "SI Appendix, Dataset S5").

Table S1 Ancient homologs of kinetochore domains and their functions (Figure 1, 6)

Note that if multiple domains have a shared evolutionary history, we regard them as a single unit in this table (kinase-polo box, WD40-NRH-Sec39). Some domains were recruited to the kinetochore before they duplicated to give rise to multiple kinetochore proteins. Those initial kinetochore entities are the 'ancestral kinetochore units'. If a protein does not have closely related homologs in the kinetochore, the protein itself was the ancestral unit that got involved in the kinetochore. For all relationships, we indicate which type of evidence we have for it. A: phylogenetic tree, B: hit in profile-profile search, C: structure and/or literature. *The phylogeny of Ska1, Ska2 and Ska3 cannot be rooted, therefore it is unknown which are the each other's closest paralog. **The BIR domain is involved in multiple processes in animals, but the kinetochore (inner centromere) function might be the ancestral one, because this is also reported in budding and fission yeast, which only have one BIR domain protein.

domain	LECA KT	closest pre-LECA kinetochore homolog(s)	ancestral kinetochore unit	closest pre-LECA non-kinetochore homolog(s)	other closely-related pre-LECA non-kinetochore homolog(s)
kinase +(polo)	Plk	-	Plk	centrosome (Plk4) **A**	diverse cellular functions (AGC kinases) **A**
	Aurora	-	Aurora	centrosome (Plk-Plk4) **A**	
kinase	MadBub	-	MadBub	uncharacterized **A**	unresolved **A**
	Mps1	-	Mps1	chromatin assembly & DNA repair (Tlk) **A**	unresolved **A**
TPR	MadBub	Mps1 **B**	anc_KT_TPR	unresolved **AB**	spliceosome (Syf1/Crnl1) **B**
	Mps1	MadBub **B**			
histone	CenpA	-	CenpA	nucleosome (H3) **A**	TBP-associated factors (TFIID, SAGA) **A** nucleosome (H3/H4/H2B) **A**
	CenpS CenpT	CenpT **A** CenpS **A**	anc_KT_histone_1	unresolved **A** (low bootstrap)	
	CenpW CenpX	CenpX **A** CenpW **A**	anc_KT_histone_2	unresolved **A** (low bootstrap)	DNA repair (DPOE3/4) **A** transcription (CBF and NC2 complex) **A** nucleosome (H2A) **A**
WD40	Cdc20	-	Cdc20	cell cycle progression (Cdh1) **A**	unresolved **A**
	Bub3	-	Bub3	nuclear mRNA transport (Rae1) **A**	unresolved **A**
HORMA	Mad2	-	Mad2	DNA repair (Rev7) **A**	autophagy (Atg13, Atg101) **A** meiosis (HORMAD) **A** prokaryotic nucleotide signaling (HORMA1) **A**
	p31^{comet}	-	p31^{comet}	unresolved (low bootstrap) **A**	
KT_RWD (1x)	Spc24 Spc25 Mad1 Csm1	unresolved order of duplications **ABC**	anc_KT_RWD	DNA repair (FancL) **ABC** transcription (Med15) **AB** transcription (Med14, Med17) **C**	ubiquitin-like conjugases (E2) **ABC** canonical RWD (e.g. Gcn2) **ABC** autophagy (Atg7, Atg10) **C**
KT_RWD (2x)	Zwint1 Knl1 CenpO CenpP				
NN-CH	Ndc80 Nuf2	Nuf2 **BC** Ndc80 **BC**	anc_KT_NN-CH	unresolved [55] **BC**	intraflagellar transport (Ift54-81, Cluap1) **BC** RNA splicing & transport (Fam98, Rtraf) **BC** vesicle transport (Ccdc22, Ccdc93) **BC**

TBP-like	CenpN CenpL	CenpL **C** CenpN **C**	anc_KT_TBP-like	unresolved **C**	transcription (TBP, Med18, Med20) **C** binding of cargo (AP-2b/4b, COPb/g1) **C** enzymes (e.g. RNaseH3, spermine synthase) **C**
Mis12 NANO	Mis12 Nkp1 Nnf1 Nkp2 Dsn1 CenpU Nsl1 CenpQ	Nkp1 **BC** Mis12 **BC** Nkp2 **BC** Nnf1 **BC** CenpU **C** Dsn1 **C** CenpQ **C** Nsl1 **C**	anc_KT_Mis12/NANO	kinetochore-specific	
Ska-like	Ska1 Ska2 Ska3	unresolved* **A**	anc_KT_Ska	kinetochore-specific	
winged-helix	Ska1	-	Ska1	unresolved **C**	transcription and replication related proteins (e.g. RepA (prokaryotic), Cdt1, Orc1, Cdc6) **C**
AAA+ ATPase	Trip13	-	Trip13	prokaryote nucleotide signaling (Trip13) **A**	-
WD40-NRH-Sec39	Rod	-	Rod	Golgi-to-ER retrograde transport (Nag) **A**	vesicle coats (Clathrin, COPII, NUPs) **C**
vesicle tether	Zw10	-	Zw10	intra-Golgi transport (Cog5) **A**	vesicle tethering (COG, GARP and exocyst) **A**
BIR	Survivin**	-	Survivin	-	-
cupin	CenpC	-	CenpC	unresolved **ABC**	metabolic enzymes, transcription factors **ABC**
HEAT	CenpI	-	CenpI	unresolved **BC**	nuclear transport (importin beta) **C** coatomers (COPb, AP-1g) **C** DNA repair (FancI) **C** Signaling (PP2A) **C**
GTPase	CenpM	-	CenpM	unresolved **AB**	Nuclear transport (Ran) **BC** ER-Golgi transport (Rab1A) **BC**
kinesin	CenpE	-	CenpE	unresolved [56] **A**	Spindle formation (Kif15) **B** Golgi-to-ER retrograde transport (Kif1C) **B**
zinc finger	BugZ	-	BugZ	unresolved **BC**	transcription regulation (Znf879) **B**
novel domains	Zwilch, Incenp, Borealin, Sgo, Cep57, CenpH, CenpK				

Table S2. Kinetochore proteins from human and/or yeast and their inferred presence in the LECA kinetochore (Figure 1B).

For each of these proteins, we determined the orthologs across eukaryotic species [1] and determined whether it was encoded by the LECA genome (a 'LECA protein'), based on Dollo parsimony (present in both Opimoda and Diphoda, "SI Appendix, Fig. S4"), or based on the inference of a pre-LECA duplication that gave rise to this protein. In addition, we assessed how likely this protein was part of the LECA kinetochore ('LECA KT protein'). Model species (h: human, y: budding yeast). Supergroup presence (Opisthokonta (O), Amoebozoa (Am), Excavata (E), SAR (S), Archaeplastida (Ar)).

Protein	Model species	Kinetochore complex	Known domains	Supergroup presence	LECA (y/n: reason)	LECA KT (y/n: reason)
Mad1	h, y	Mad1-Mad2	RWD [57]	O, Am, E, S, Ar	yes: parsimony	yes
Mad2	h, y	Mad1-Mad2, MCC	HORMA [58]	O, Am, E, S, Ar	yes: parsimony	yes
Bub3	h, y	MCC	WD40 [59]	O, Am, E, S, Ar	yes: parsimony	yes
Cdc20	h, y	MCC	WD40 [60]	O, Am, E, S, Ar	yes: parsimony	yes
MadBub	h, y	MCC	TPR, kinase [61]	O, Am, E, S, Ar	yes: parsimony	yes
Mps1	h, y		TPR, kinase [62–64]	O, Am, E, S, Ar	yes: parsimony	yes
p31	h	Mad2-Mad2, MCC	HORMA [65]	O, Am, E, S, Ar	yes: parsimony	yes
Trip13	h	Mad2-Mad2, MCC	AAA+ ATPase [66–68]	O, Am, E, S, Ar	yes: parsimony	yes
Knl1	h, y	Knl1-Zwint-1	RWD (2x) [43]	O, Am, E, S, Ar	yes: parsimony	yes
Zwint-1	h, y	Knl1-Zwint-1	RWD (2x) [30, 43]	O, Am, E, S, Ar	yes: parsimony	yes
Dsn1	h, y	Mis12-C	Mis12 NANO [40, 42, 69]	O, Am, E, S, Ar	yes: parsimony	yes
Nsl1	h, y	Mis12-C	Mis12 NANO [40, 42, 69]	O, Am, S, Ar	yes: parsimony	yes

Nnf1	h, y	Mis12-C	Mis12 NANO [40, 42, 69]	O, Am, E, S, Ar	yes: parsimony	yes
Mis12	h, y	Mis12-C	Mis12 NANO [40, 42, 69]	O, Am, E, S, Ar	yes: parsimony	yes
CEP57	h			O, E, S	yes: parsimony	yes
ARHGEF 17	h		RhoGEF [70]	O	no: parsimony	no
Ndc80	h, y	Ndc80-C	CH [55, 71]	O, Am, E, S, Ar	yes: parsimony	yes
Spc24	h, y	Ndc80-C	RWD [72]	O, Am, E, S, Ar	yes: parsimony	yes
Spc25	h, y	Ndc80-C	RWD [72]	O, Am, E, S, Ar	yes: parsimony	yes
Nuf2	h, y	Ndc80-C	CH [55, 71]	O, Am, E, S, Ar	yes: parsimony	yes
Ska1	h	Ska-C	Ska-like [31, 73]	O, Am, S, Ar	yes: parsimony	yes
Ska2	h	Ska-C	Ska-like [31, 73]	O, Am, E, S, Ar	yes: parsimony	yes
Ska3	h	Ska-C	Ska-like [31, 73]	O, Am, S, Ar	yes: parsimony	yes
Ska1	H	Ska-C	winged-helix [74]	O, Am, S, Ar	yes: parsimony	yes
Dam1	y	Dam1-C	all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no
Duo1	y	Dam1-C	Duo2-Dad2-like [31], all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no
Dad2	y	Dam1-C	Duo2-Dad2-like [31], all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no

Dad1	y	Dam1-C	Dad1/4-Ask1-like [31], all Dam1-C subunits [75]	O, Am, S, Ar	no: HGT [31]	no
Dad3	y	Dam1-C	all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no
Dad4	y	Dam1-C	Dad1/4-Ask1-like [31], all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no
Hsk3	y	Dam1-C	all Dam1-C subunits [75]	O, S	no: HGT [31]	no
Ask1	y	Dam1-C	Dad1/4-Ask1-like [31], all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no
Spc19	y	Dam1-C	all Dam1-C subunits [75]	O	no: parsimony	no
Spc34	y	Dam1-C	all Dam1-C subunits [75]	O, S, Ar	no: HGT [31]	no
SKAP	h	SKAP-Astrin		O	no: parsimony	no
Astrin	h	SKAP-Astrin		O	no: parsimony	no
Spindly	h	RZZS	BICD(R), TRAK/HAP, CDR2 [76]	O	no: parsimony	no
Rod	h	RZZS	NRH, Sec39, WD40 [77]	O, E, S	yes: parsimony	yes, but without Spindly, it is unclear how it was recruited to the KT

Zwilch	h	RZZS		O, S	yes: parsimony	yes, but without Spindly, it is unclear how it was recruited to the KT
Zw10	h	RZZS	Vsp51 [78]	O, Am, E, S, Ar	yes: parsimony	yes, but without Spindly, it is unclear how it was recruited to the KT
Aurora	h, y	CPC	Kinase [79]	O, Am, E, S, Ar	yes: parsimony	yes
Incenp	h, y	CPC		O, Am, E, S, Ar	yes: parsimony	yes
Survivin	h, y	CPC	Baculoviral IAP repeat (BIR) [80]	O, E, S	yes: parsimony	yes
Borealin	h	CPC		O, Am, E, S, Ar	yes: parsimony	yes
Sgo	h, y	CPC		O, Am, S, Ar	yes: parsimony	yes
BugZ	h		Zinc finger	O, Am, E, S, Ar	yes: parsimony	yes
Plk	h, y		Kinase, polo box [81, 82]	O, Am, E, S, Ar	yes: parsimony	yes
CenpA	h, y	Histone	Histone [83]	O, Am, E, S, Ar	yes: parsimony and phylogeny [1]	yes
CenpB	h	CCAN	TC5-DDE (TE) [84]	O	no: parsimony	no
CenpC	h, y	CCAN	Cupin [85]	O, Am, E, S, Ar	yes: parsimony	yes
CenpF	h			O	no: parsimony	no
CenpE	h		Kinesin [86]	O, Am, E, S, Ar	yes: parsimony	yes

CenpH	h, y	CCAN-HIKM	[42]	O, Am, S	yes: parsimony	yes
CenpI	h, y	CCAN-HIKM	HEAT-like repeats [87]	O, Am, S	yes: parsimony	yes
CenpK	h, y	CCAN-HIKM	[42]	O, Am, S	yes: parsimony	yes
CenpM	h	CCAN-HIKM	GTPase [87]	O, Am, S	yes: parsimony + pre-LECA duplicate	yes
CenpL	h, y	CCAN-LN	TBP-like [42, 88]	O, Am, S	yes: parsimony	yes
CenpN	h, y	CCAN-LN	TBP-like [42, 88]	O, Am, S	yes: parsimony	yes
CenpO	h, y	CCAN-OPQRU	RWD (2x) [42, 89]	O, Am, S, Ar	yes: parsimony + pre-LECA duplicate	yes
CenpP	h, y	CCAN-OPQRU	RWD (2x) [42, 89]	O, Am, S	yes: parsimony + pre-LECA duplicate	yes
CenpQ	h, y	CCAN-OPQRU	Mis12 NANO [42]	O, Am, S	yes: parsimony	yes
CenpR	h	CCAN-OPQRU		O	no: parsimony	no
CenpU	h, y	CCAN-OPQRU	Mis12 NANO [42]	O, Am, S	yes: parsimony	yes
Nkp1	y	CCAN-Nkp	Mis12 NANO [42]	O	yes: predicted pre-LECA duplicate	yes
Nkp2	y	CCAN-Nkp	Mis12 NANO [42]	O, Am	yes: predicted pre-LECA duplicate	yes
CenpT	h, y	CCAN-TSWX	Histone [90, 91]	O, Am, S	yes: parsimony + pre-LECA duplicate	yes

CenpW	h, y	CCAN-TSWX	Histone [90, 91]	O, Am, S	yes: parsimony + pre-LECA duplicate	yes
CenpS	h, y	CCAN-TSWX	Histone [91]	O, Am, E, S, Ar	yes: parsimony + pre-LECA duplicate	yes
CenpX	h, y	CCAN-TSWX	Histone [91]	O, Am, E, S, Ar	yes: parsimony + pre-LECA duplicate	yes
Ndc10	y	CBF3	GCR1_C, Crypton F-like [92]	O	no: parsimony	no
Ctf13	y	CBF3	F-box [93]	O	no: parsimony	no
Cep3	y	CBF3	Zinc finger, HEAT repeat [94]	O	no: parsimony	no
Skp1	y	CBF3	[95]	O, Am, E, S, Ar	yes: parsimony	No: likely operated in SCF ubiquitin ligase complex and was recruited to the KT with the CBF3 complex
Csm1	y	Monopolin	RWD [2, 96]	O, Am, S, Ar	Yes: parsimony	yes
Lsr4	y	Monopolin	[96, 97]	O	No: parsimony	no
Mam1	y	Monopolin	[97]	O	No: parsimony	no
Hrr25	y	Monopolin	Kinase [97]	O, Am, E, S, Ar	Yes: parsimony	No: likely KT function since the origin of Lsr4, Mam1 (Saccharomyce tales)

Table S3. Eukaryotic species used for phylogenetic analysis.

Table containing all eukaryotic species information from which sequences are used to construct the phylogenetic trees in "SI Appendix, Fig. S1A-I".

Tax ID	Abbrv.	Scientific name	Common name	Eukaryotic supergroup	Group
595528	COWC	*Capsaspora owczarzaki ATCC 30864*	Amoeboid symbiont	Opisthokonta	Filasterea
431895	MBRE	*Monosiga brevicollis MX1 / ATCC 50154*	Choanoflagellate	Opisthokonta	Choanomonada
946362	SROS	*Salpingoeca rosetta*	Choanoflagellate	Opisthokonta	Choanomonada
400682	AQUE	*Amphimedon queenslandica*	Sponge	Opisthokonta	Metazoa
10228	TADH	*Trichoplax adhaerens*	Placozoan	Opisthokonta	Metazoa
45351	NVEC	*Nematostella vectensis*	Starlet sea anemone	Opisthokonta	Metazoa
27923	MLEI	*Mnemiopsis leidyi*	Sea walnut/Warty comb Jelly	Opisthokonta	Metazoa
6183	SMAN	*Schistosoma mansoni*	Flatworm	Opisthokonta	Metazoa
6239	CELE	*Caenorhabditis elegans*	Roundworm	Opisthokonta	Metazoa
6279	BMAL	*Brugia malayi*	Filarial nematode worm	Opisthokonta	Metazoa
7227	DMEL	*Drosophila melanogaster*	Fruitfly	Opisthokonta	Metazoa
180454	AGAM	*Anopheles gambiae str. PEST*	Mosquito	Opisthokonta	Metazoa
10224	SKOW	*Saccoglossus kowalevskii*	Acorn worm	Opisthokonta	Metazoa
7719	CINT	*Ciona intestinalis*	Transparent sea squirt	Opisthokonta	Metazoa
7739	BFLO	*Branchiostoma floridae*	Florida lancelet	Opisthokonta	Metazoa
7741	BBEL	*Branchiostoma belcheri*	Belcher's lancelet	Opisthokonta	Metazoa
7955	DRER	*Danio rerio*	Zebrafish	Opisthokonta	Metazoa
31033	TRUB	*Takifugu rubripes*	Japanese pufferfish	Opisthokonta	Metazoa
8364	XTRO	*Xenopus tropicalis*	Western clawed frog	Opisthokonta	Metazoa
10090	MMUS	*Mus musculus*	Mouse	Opisthokonta	Metazoa
9606	HSAP	*Homo sapiens*	Human	Opisthokonta	Metazoa
1993908	NUSP	*Parvularia atlantis*		Opisthokonta	Holomycota
948595	VCUL	*Vavraia culicis floridensis*	Microsporidian parasite	Opisthokonta	Microsporidia
876142	EINT	*Encephalitozoon intestinalis ATCC 50506*	Microsporidian parasite	Opisthokonta	Microsporidia
1003232	EAED	*Edhazardia aedis USNM 41457*	Microsporidian parasite	Opisthokonta	Microsporidia
684364	BDEN	*Batrachochytrium dendrobatidis JAM81*	Chytrid fungus	Opisthokonta	Chytridiomycota
645134	SPUN	*Spizellomyces punctatus DAOM BR117*	Chytrid fungus	Opisthokonta	Chytridiomycota

578462	AMAC	*Allomyces macrogynus ATCC 38327*	Chytrid fungus	Opisthokonta	Blastocladiales
765915	CANG	*Catenaria anguillulae PL171*	Chytrid fungus	Opisthokonta	Blastocladiales
1220926	MCIR	*Mucor circinelloides 1006PhL*	Zygomycete fungus	Opisthokonta	Mucoromycotina
763407	PBLA	*Phycomyces blakesleeanus NRRL1555*	Zygomycete fungus	Opisthokonta	Mucoromycotina
1069443	MVER	*Mortierella verticillata NRRL 6337*	Zygomycete fungus	Opisthokonta	Mortierellaceae
310910	MELO	*Mortierella elongata*	Zygomycete fungus	Opisthokonta	Mortierellaceae
1357683	BBER	*Basidiobolus meristosporus B9252*		Opisthokonta	Entomophtorales
796925	CCOR	*Conidiobolus coronatus NRRL28638*	Zygomycete fungus	Opisthokonta	Entomophtorales
763665	CREV	*Coemansia reversa NRRL 1564*	Zygomycete fungus	Opisthokonta	Kickxellomycotina
237631	UMAY	*Ustilago maydis*	Fungus	Opisthokonta	Dikarya
235443	CNEO	*Cryptococcus neoformans*	Fungus	Opisthokonta	Dikarya
4896	SPOM	*Schizosaccharomyces pombe (strain 972 / ATCC 24843)*	Fission yeast	Opisthokonta	Dikarya
367110	NCRA	*Neurospora crassa OR74A*	Filamentous fungus	Opisthokonta	Dikarya
284591	YLIP	*Yarrowia lipolytica CLIB 122*	Yeast	Opisthokonta	Dikarya
284592	DHAN	*Debaryomyces hansenii CBS767*	Yeast	Opisthokonta	Dikarya
284590	KLAC	*Kluyveromyces lactis NRRL Y-1140*	Yeast	Opisthokonta	Dikarya
284593	CGLA	*Candida glabrata CBS138*	Yeast	Opisthokonta	Dikarya
559292	SCER	*Saccharomyces cerevisiae S288C*	Baker's yeast	Opisthokonta	Dikarya
461836	TTRA	*Thecamonas trahens ATCC 50062*	Bacteriovorous flagellated protist	(sister to Opisthokonta)	Apusomonadida
1257118	ACAS	*Acanthamoebe castellanii str. Neff*	Amoeba	Amoebozoa	Discosea
294381	EHIS	*Entamoeba histolytica HM-1:IMSS*	Amoeba	Amoebozoa	Entamoebidae
352472	DDIS	*Dictyostelium discoideum AX4*	Slime mould	Amoebozoa	Dictyosteliida
670386	PPAL	*Polysphondylium pallidum PN500*	Cellular slime mold	Amoebozoa	Dictyosteliida
1410327	ASUB	*Acytostelium subglobosum LB1*		Amoebozoa	Dictyosteliida
280463	EHUX	*Emiliania huxleyi CCMP1516*	Haptophyte	(sister to SAR)	Haptophyta
1460289	CTOB	*Chrysochromulina tobin strain CCMP291*	Haptophyte	(sister to SAR)	Haptophyta
453998	MONO	*Monocercomonoides sp. PA203*		Excavata	Oxymonadida
412133	TVAG	*Trichomonas vaginalis G3*	Protozoan	Excavata	Parabasalia
941442	GINT	*Giardia intestinalis assemblage A*	Protozoan	Excavata	Fornicata
347515	LMAJ	*Leishmania major strain Friedlin*	Protozoan	Excavata	Discicristata
185431	TBRU	*Trypanosoma brucei TREU 927*	Protozoan	Excavata	Discicristata
3039	EGRA	*Euglena gracilis*		Excavata	Discicristata

744533	NGRU	*Naegleria gruberi strain NEG-M*	Protozoan	Excavata	Discicristata
12968	BHOM	*Blastocystis hominis*	Single-celled protozoan parasite	SAR	Stramenopiles
702273	AKER	*Aplanochytrium kerguelense PBS07*	Marine protist	SAR	Stramenopiles
717989	ALIM	*Aurantiochytrium limacinum ATCC MYA-1381*	Marine protist	SAR	Stramenopiles
	SPAR				
403677	PINF	*Phytophthora infestans T30-4*	Potato late blight fungus	SAR	Stramenopiles
890382	ALAI	*Albugo laibachii Nc14*	Arabidopsis pathogen	SAR	Stramenopiles
272952	HPAR	*Hyaloperonospora parasitica*	former name: Peronospora parasitica. Also called Hyaloperonospora arabidopsidis	SAR	Stramenopiles
1093141	NGAD	*Nannochloropsis gaditana CCMP526*	Stramenopile alga	SAR	Stramenopiles
44056	AANO	*Aureococcus anophagefferens CCMP1984*	Harmful bloom alga	SAR	Stramenopiles
2880	ESIL	*Ectocarpus siliculosus*	Brown alga	SAR	Stramenopiles
309737	COKA	*Cladosiphon okamuranus*	Brown alga	SAR	Stramenopiles
556484	PTRI	*Phaeodactylum tricornutum CCAP1055/1*	Diatome	SAR	Stremenopiles
296543	TPSE	*Thalassiosira pseudonana*	Marine diatom	SAR	Stremenopiles
423536	PMAR	*Perkinsus marinus ATCC 50983*		SAR	Alveolata
1202447	SMIN	*Symbiodinium minutum*	Dinoflagellate	SAR	Alveolata
36329	PFAL	*Plasmodium falciparum 3D7*	Protozoan	SAR	Alveolata
414452	CPAI	*Cryptosporidium parvum Iowa II*	Protozoan	SAR	Alveolata
508771	TGON	*Toxoplasma gondii ME49*	Protozoan	SAR	Alveolata
5888	PTET	*Paramecium tetraurelia*	Paramecium	SAR	Alveolata
5911	TTHE	*Tetrahymena thermophila*	Tetrahymena	SAR	Alveolata
1172189	OTRI	*Oxytricha trifallax*	Ciliate	SAR	Alveolata
5963	SCOE	*Stentor coeruleus*	Ciliate	SAR	Alveolata
753081	BNAT	*Bigelowiella natans CCMP2755*	Chlorarachniophyte alga	SAR	Rhizaria
905079	GTHE	*Guillardia theta CCMP2712*	Cryptomonad alga	Archaeplastida	Cryptophyceae
2762	CPAR	*Cyanophora paradoxa*	Glaucophyte	Archaeplastida	Glaucophyta
130081	GSUL	*Galdieria sulphuraria*	Red alga	Archaeplastida	Rhodophyceae
280699	CMER	*Cyanidioschyzon merolae 10D*	Red alga	Archaeplastida	Rhodophyceae
554065	CVAR	*Chlorella variabilis NC64A*	Green alga	Archaeplastida	Chloroplastida
574566	CSUB	*Coccomyxa subellipsoidea C-169*	Green alga	Archaeplastida	Chloroplastida

3055	CREI	*Chlamydomonas reinhardtii*	Green alga	Archaeplastida	Chloroplastida
3068	VCAR	*Volvox carteri*	Green alga	Archaeplastida	Chloroplastida
296587	MSPE	*Micromonas species RCC299*	Picoplanktonic green alga	Archaeplastida	Chloroplastida
436017	OLUC	*Ostreococcus lucimarinus CCE9901*	Green alga	Archaeplastida	Chloroplastida
1075084	BPRA	*Bathycoccus prasinos RCC1005*	Green alga	Archaeplastida	Chloroplastida
3175	KFLA	*Klebsormidium flaccidum*		Archaeplastida	Klebsormidiaceae
145481	PPAT	*Physcomitrella patens subsp. patens*	Moss	Archaeplastida	Chloroplastida
88036	SMOE	*Selaginella moellendorffii*	Spikemoss	Archaeplastida	Chloroplastida
13333	ATRI	*Amborella trichopoda*		Archaeplastida	Chloroplastida
39947	OSAT	*Oryza sativa japonica*	Rice	Archaeplastida	Chloroplastida
3702	ATHA	*Arabidopsis thaliana*	Thale cress	Archaeplastida	Chloroplastida
218851	ACOE	*Aquilegia coerulea Goldsmith*	Rocky mountain columbine	Archaeplastida	Chloroplastida

References

1. van Hooff JJ, Tromer E, van Wijk LM, et al (2017) Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep* 18:1559–1571.
2. R. Plowman, N. Singh, E. C. Tromer, A. Payan, E. Duro, C. Spanos, J. Rappsilber, B. Snel, G. J. P. L. Kops, K. D. Corbett ALM (2019) The Molecular Basis of Monopolin Recruitment to the Kinetochore. *Chromosoma* (*in press:* please contact the corresponding authors for the relevant data).
3. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780.
4. Mi H, Huang X, Muruganujan A, et al (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45:183–189.
5. Madera M (2008) Profile Comparer: a program for scoring and aligning profile hidden Markov models. *Bioinformatics* 24:2630–2631.
6. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960.
7. Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.
8. Velankar S, Dana JM, Jacobsen J, et al (2013) SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res* 41:483–489.
9. van Wijk LM, Berend S (2019) Phylogenomics reveals ancestral kinase relations, repertoire, and fate in present-day eukaryotes. (*in preparation:* please contact the corresponding authors for more information)
10. Burroughs AM, Zhang D, Schäffer DE, et al (2015) Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* 43:10633–10654.
11. Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066.
12. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
13. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
14. 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.
15. Rambaut A (2012) FigTree v1. 4. Mol Evol phylogenetics Epidemiol Edinburgh, UK Univ Edinburgh, Inst Evol Biol
16. Finn RD, Bateman A, Clements J, et al (2014) Pfam: The protein families database. *Nucleic Acids Res.* 42:222–230
17. Cheng H, Schaeffer RD, Liao Y, et al (2014) ECOD: An Evolutionary Classification of Protein Domains. *PLoS Comput Biol* 10:e1003926.
18. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
19. Dawson NL, Lewis TE, Das S, et al (2017) CATH: An expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* 45:289–295.
20. Schrödinger L (2010) PyMOL The PyMOL Molecular Graphics System
21. Holm L, Sander C (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci* 20:478–480.
22. Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44:W351–W355.
23. Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43:389–394.
24. Waterhouse AM, Procter JB, Martin DMA, et al (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.
25. Makarova KS, Wolf YI, Mekhedov SL, et al (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* 33:4626–4638.
26. Mistry J, Finn RD, Eddy SR, et al (2013) Challenges in homology search: HMMER3 and convergent

evolution of coiled-coil regions. *Nucleic Acids Res* 41:e121.
27. Derelle R, Torruella G, Klimeš V, et al (2015) Bacterial proteins pinpoint a single eukaryotic root. *Proc Natl Acad Sci U S A* 112:693–699.
28. Williams TA (2014) Evolution: Rooting the eukaryotic tree of life. *Curr. Biol.* 24:151–152
29. He D, Fiz-Palacios O, Fu C-J, et al (2014) An Alternative Root for the Eukaryote Tree of Life. *Curr Biol* 24:465–470.
30. Tromer EC (2017) Evolution of the Kinetochore Network in Eukaryotes. Utrecht Univerisity. https://doi.org/10.13140/RG.2.2.16846.56640
31. van Hooff JJE, Snel B, Kops GJPL (2017) Unique Phylogenetic Distributions of the Ska and Dam1 Complexes Support Functional Analogy and Suggest Multiple Parallel Displacements of Ska by Dam1. *Genome Biol Evol* 9:1295–1303.
32. Akiyoshi B, Gull K (2014) Discovery of Unconventional Kinetochores in Kinetoplastids. *Cell* 156:1247–1258.
33. D'Archivio S, Wickstead B (2017) Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J Cell Biol* 216:379–391.
34. Adl SM, Simpson AGB, Lane CE, et al (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* 59:429–493.
35. Archibald JM, Simpson AGB, Slamovits CH (2016) Handbook of the Protists. Springer International Publishing, Cham
36. Brown MW, Heiss AA, Kamikawa R, et al (2018) Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biol Evol* 10:427–433.
37. Akiyoshi B (2016) The unconventional kinetoplastid kinetochore: from discovery toward functional understanding. *Biochem Soc Trans* 44:1201–1217.
38. Cavalier-Smith T (2010) Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett* 6:342–345.
39. Zimmermann L, Stephens A, Nam S-Z, et al (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* 430:2237–2243.
40. Petrovic A, Keller J, Liu Y, et al (2016) Structure of the MIS12 Complex and Molecular Basis of Its Interaction with CENP-C at Human Kinetochores. *Cell* 167:1028–1040.
41. Zhou X, Zheng F, Wang C, et al (2017) Phosphorylation of CENP-C by Aurora B facilitates kinetochore attachment error correction in mitosis. *Proc Natl Acad Sci* 114:10677–10676.
42. Hinshaw SM, Harrison SC (2019) The structure of the Ctf19c/CCAN from budding yeast. *Elife* 8.
43. Petrovic A, Mosalaganti S, Keller J, et al (2014) Modular Assembly of RWD Domains on the Mis12 Complex Underlies Outer Kinetochore Organization. *Mol Cell* 53:591–605.
44. Burroughs AM, Jaffee M, Iyer LM, Aravind L (2008) Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J Struct Biol* 162:205–218.
45. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, et al (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541:353–358.
46. Nunoura T, Takaki Y, Kakuta J, et al (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* 39:3204–3223.
47. Hennell James R, Caceres EF, Escasinas A, et al (2017) Functional reconstruction of a eukaryotic-like E1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon. *Nat Commun* 8:1120.
48. Grau-Bove X, Sebe-Pedros A, Ruiz-Trillo I (2015) The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. *Mol Biol Evol* 32:726–739.
49. Sundquist WI, Schubert HL, Kelly BN, et al (2004) Ubiquitin Recognition by the Human TSG101 Protein. *Mol Cell* 13:783–789.
50. Xu L, Sowa ME, Chen J, et al (2008) An FTS/Hook/p107(FHIP) complex interacts with and promotes endosomal clustering by the homotypic vacuolar protein sorting complex. *Mol Biol Cell* 19:5059–5071.
51. Nameki N, Yoneyama M, Koshiba S, et al (2004) Solution structure of the RWD domain of the mouse GCN2 protein. *Protein Sci* 13:2089–2100.
52. Hodson C, Cole AR, Lewis LPC, et al (2011) Structural analysis of human FANCL, the E3 ligase in the

Fanconi anemia pathway. *J Biol Chem* 286:32628–32637.
53. Nozawa K, Schneider TR, Cramer P (2017) Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature* 545:248–251.
54. Mattiroli F, Bhattacharyya S, Dyer PN, et al (2017) Structure of histone-based chromatin in Archaea. *Science* 357:609–612.
55. 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.
56. Wickstead B, Gull K, Richards T a (2010) Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. *BMC Evol Biol* 10:110.
57. Kim S, Sun H, Tomchick DR, et al (2012) Structure of human Mad1 C-terminal domain reveals its involvement in kinetochore targeting. *Proc Natl Acad Sci* 109:6549–6554.
58. Aravind L, Koonin E V (1998) The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair. *Trends Biochem Sci* 23:284–286.
59. Wilson DK, Cerna D, Chew E (2005) The 1.1-Å Structure of the Spindle Checkpoint Protein Bub3p Reveals Functional Regions. *J Biol Chem* 280:13944–13951.
60. Tian W, Li B, Warrington R, et al (2012) Structural analysis of human Cdc20 supports multisite degron recognition by APC/C. *Proc Natl Acad Sci* 109:18419.
61. Bolanos-Garcia VM, Blundell TL (2011) BUB1 and BUBR1: multifaceted kinases of the cell cycle. *Trends Biochem Sci* 36:141–150.
62. Lee S, Thebault P, Freschi L, et al (2012) Characterization of spindle checkpoint kinase mps1 reveals domain with functional and structural similarities to tetratricopeptide repeat motifs of Bub1 and BubR1 checkpoint kinases. *J Biol Chem* 287:5988–6001.
63. Thebault P, Chirgadze D, Dou Z, et al (2012) Structural and functional insights into the role of the N-terminal Mps1 TPR domain in the spindle assembly checkpoint (SAC). *Biochem J* 448(3):321:328.
64. Nijenhuis W, von Castelmur E, Littler D, et al (2013) A TPR domain–containing N-terminal module of MPS1 is required for its kinetochore localization by Aurora B. *J Cell Biol* 201:217–231.
65. Yang M, Li B, Tomchick DR, et al (2007) p31comet Blocks Mad2 Activation through Structural Mimicry. *Cell* 131:744–755.
66. Ye Q, Kim DH, Dereli I, et al (2017) The AAA+ ATPase TRIP13 remodels HORMA domains through N-terminal engagement and unfolding. *EMBO J* 36:2419–2434.
67. Ye Q, Rosenberg SC, Moeller A, et al (2015) TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. *Elife* 4:e07367.
68. Brulotte ML, Jeong B-C, Li F, et al (2017) Mechanistic insight into TRIP13-catalyzed Mad2 structural transition and spindle checkpoint silencing. *Nat Commun* 8:1956.
69. Dimitrova YN, Jenni S, Valverde R, et al (2016) Structure of the MIND Complex Defines a Regulatory Focus for Yeast Kinetochore Assembly. *Cell* 167:1014–1027.
70. Rümenapp U, Freichel-Blomquist A, Wittinghofer B, et al (2002) A mammalian Rho-specific guanine-nucleotide exchange factor (p164-RhoGEF) without a pleckstrin homology domain. *Biochem J* 366:721–728.
71. Ciferri C, Pasqualato S, Screpanti E, et al (2008) Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell* 133:427–439
72. Wei RR, Schnell JR, Larsen NA, et al (2006) Structure of a central component of the yeast kinetochore: the Spc24p/Spc25p globular domain. *Structure* 14:1003–1009.
73. Jeyaprakash AA, Santamaria A, Jayachandran U, et al (2012) Structural and functional organization of the Ska complex, a key component of the kinetochore-microtubule interface. *Mol Cell* 46:274–286.
74. Abad MA, Medina B, Santamaria A, et al (2014) Structural basis for microtubule recognition by the human kinetochore Ska complex. *Nat Commun* 5:2964.
75. Jenni S, Harrison SC (2018) Structure of the DASH/Dam1 complex shows its role at the yeast kinetochore-microtubule interface. *Science* 360:552–558.
76. Sacristan C, Ahmad MUD, Keller J, et al (2018) Dynamic kinetochore size regulation promotes microtubule capture and chromosome biorientation in mitosis. *Nat Cell Biol* 20:.

77. Çivril F, Wehenkel A, Giorgi FM, et al (2010) Structural analysis of the RZZ complex reveals common ancestry with multisubunit vesicle tethering machinery. *Structure* 18:616–626.
78. Tripathi A, Ren Y, Jeffrey PD, Hughson FM (2009) Structural characterization of Tip20p and Dsl1p, subunits of the Dsl1p vesicle tethering complex. *Nat Struct &Amp; Mol Biol* 16:114.
79. Sessa F, Mapelli M, Ciferri C, et al (2005) Mechanism of Aurora B Activation by INCENP and Inhibition by Hesperadin. *Curr Biol* 18:379–391.
80. Jeyaprakash AA, Klein UR, Lindner D, et al (2007) Structure of a Survivin–Borealin–INCENP Core Complex Reveals How Chromosomal Passengers Travel Together. *Cell* 131:271–285.
81. Elling RA, Fucini R V, Romanowski MJ (2008) Structures of the wild-type and activated catalytic domains of Brachydanio rerio Polo-like kinase 1 (Plk1): changes in the active-site conformation and interactions with ligands. *Acta Crystallogr Sect D* 64:909–918.
82. Cheng K, Lowe ED, Sinclair J, et al (2003) The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. *EMBO J* 22:5757–5768.
83. Tachiwana H, Kagawa W, Shiga T, et al (2011) Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* 476:232–235
84. Kipling D, Warburton PE (1997) Centromeres, CENP-B and Tigger too. *Trends Genet* 13:141–145.
85. Cohen RL, Espelin CW, De Wulf P, et al (2008) Structural and functional dissection of Mif2p, a conserved DNA-binding kinetochore protein. *Mol Biol Cell* 19:4480–4491.
86. Garcia-Saez I, Yen T, Wade RH, Kozielski F (2004) Crystal Structure of the Motor Domain of the Human Kinetochore Protein CENP-E. *J Mol Biol* 340:1107–1116.
87. Basilico F, Maffini S, Weir JR, et al (2014) The pseudo GTPase CENP-M drives human kinetochore assembly. *Elife* 3:e02978.
88. Pentakota S, Zhou K, Smith C, et al (2017) Decoding the centromeric nucleosome through CENP-N. *Elife* 6.
89. Schmitzberger F, Harrison SC (2012) RWD domain: a recurring module in kinetochore architecture shown by a Ctf19–Mcm21 complex structure. *EMBO Rep* 13:216–222.
90. Hori T, Amano M, Suzuki A, et al (2008) CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. *Cell* 135:1039–1052.
91. Nishino T, Takeuchi K, Gascoigne KE, et al (2012) CENP-T-W-S-X forms a unique centromeric chromatin structure with a histone-like fold. *Cell* 148:487–501.
92. Cho US, Harrison SC (2012) Ndc10 is a platform for inner kinetochore assembly in budding yeast. *Nat Struct Mol Biol* 19:48–56.
93. Russell ID, Grancell AS, Sorger PK (1999) The Unstable F-box Protein p58-Ctf13 Forms the Structural Core of the CBF3 Kinetochore Complex. *J Cell Biol* 145:933–950.
94. Bellizzi JJ, Sorger PK, Harrison SC (2007) Crystal Structure of the Yeast Inner Kinetochore Subunit Cep3p. *Structure* 15:1422–1430.
95. Leber V, Nans A, Singleton MR (2018) Structural basis for assembly of the CBF3 kinetochore complex. *EMBO J* 37:269–281.
96. Corbett KD, Yip CK, Ee LS, et al (2010) The monopolin complex crosslinks kinetochore components to regulate chromosome-microtubule attachments. *Cell* 142:556–567.
97. Ye Q, Ur SN, Su TY, Corbett KD (2016) Structure of the Saccharomyces cerevisiae Hrr25:Mam1 monopolin subcomplex reveals a novel kinase regulator. *EMBO J* 35:2139–2151.

Datasets

Dataset S1 (Excel sheets)

HHsearch/PRC output.

Dataset S2 (Excel sheets)

Overview of all the structures analyzed in this study, including all all-vs-all structure similarity scores for RWD/UBC, histone, TBP-like, winged-helix and Mis12/NANO.

Dataset S3 (multiple sequence alignment)

Multiple Sequence alignment of RWD/UBC-like, associated to Figure 2D, "SI Appendix, Fig. S1E, Fig. S3".

Dataset S4 (multiple sequence alignment)

Multiple Sequence Alignment of histones, associated to Figure 3B and "SI Appendix, Fig. S1I".

Dataset S5 (Excel sheets)

Information on the operon structure of prokaryotes with HORMA-1/HORMA-2 and Trip13 proteins, associated to Figure 5 and "SI Appendix, Fig. S5".

External Data

Please find the datasets described below by clicking on this link: https://figshare.com/s/884e8848b032f02eb7c3

Hidden Markov Models (archive text files)

Archived text files (.zip) containing 147 HMM profiles (hmmer3.1b format) of kinetochore proteins and/or their domains used in the profile-versus-profile searches using HHsearch and PRC (for results see "SI Appendix, Dataset S1, External Data: HHsearch Network").

HHsearch Network (Cytoscape Network file)

Cytoscape files of the HHsearch network. Raw profile-versus-profile bit-scores and E-values, and the used HMM profiles can be found in "SI Appendix, Dataset S1, External Data: Hidden Markov Models". This file can be opened with Cytoscape (https://cytoscape.org/).

RWD/UBC structures (Pymol session file)

Pymol session file, containing all RWD/UBC structures used in this study, associated to the analysis shown in Figure 2C. Raw structural similarity scores can be found in the RWD table in "SI Appendix, Dataset S2". This file can be opened with Pymol (http://pymol.org/).

TBP-like structures (Pymol session file)

Pymol session file, containing all TBP-like structures used in this study, associated to the analysis shown in Figure 3C. Raw structural similarity scores can be found in the TBP-like table in "SI Appendix, Dataset S2". This file can be opened with Pymol (http://pymol.org/).

Histone structures (Pymol session file)

Pymol session file, containing all histone structures used in this study, associated to the analysis shown in Figure 3B. Raw structural similarity scores can be found in the histones table in "SI Appendix, Dataset S2". This file can be opened with Pymol (http://pymol.org/).

Mis12/NANO structures (Pymol session file)

Pymol session file, containing all Mis12/NANO structures used in this study, associated to the analysis shown in Figure 4. Raw structural similarity scores can be found in the Mis12/NANO table in "SI Appendix, Dataset S2". This file can be opened with Pymol (http://pymol.org/).