Evolution of the Kinetochore Network in Eukaryotes

Eelco Tromer

ISBN 978-90-393-6905-0

Printed by Ridderprint BV
Cover Artistic interpretation of a plane fractal termed the pythagoras tree, which was first visualized by the Dutch mathematician Albert E. Bosman in 1942.

Evolution of the Kinetochore Network in Eukaryotes

Evolutie van het kinetochoornetwerk in eukaryoten

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 12 december 2017 des middags te 4.15 uur

door

Elbert Cornelis Tromer

geboren op 23 april te Soest

Promoteren
Prof. dr. G.J.P.L. Kops
Prof. dr. B. Snel

‘Not all those who wander are lost’
- **J.R.R. Tolkien**-

Contents

CHAPTER 1

General Introduction
Thesis Outline

General Introduction

DNA is the physical code of cellular life

Evolution of life on earth emerges from the propagation and stable maintenance of long double-stranded helical polymer chains consisting of 2 pairs of complementary deoxyribonucleic acid nucleotide bases: adenine (A) – thymine (T) and cytosine (C) – guanine (G), widely known as 'the DNA'. Encapsulated in dividing cells, the total complement of this deceptively simple and negatively charged biopolymer – here called the genome, reveals its striking capacity to encode, express and transfer most heritable traits of an organism and as such is the primary substrate of evolutionary forces. Hidden within the sequence of ATCGs lay islands called genes that are the major vehicles of heredity and harbor the code for all subunits of intra- and extracellular machineries such as proteins and various structural and regulatory RNA molecules.

In eukaryotes the genome is compartmentalized in a double-membraned nucleus and its genes lay dispersed over multiple large gene bodies, called chromosomes. The chromosomal location of genes is subject to strong evolutionary pressure, often creating boundaries between species. The DNA of chromosomes is wrapped around cylindrical structures that consist of 8 histone proteins called nucleosomes; forming a structural state known as chromatin. Compared to prokaryotes, the evolution of the chromatin state in eukaryotes enabled the increase in temporal and spatial (epigenetic) regulation of the genome and provided additional protection to the DNA against excessive mutations. This likely sparked the simultaneous increase in genome size and extensive genome rearrangements, resulting in low gene density observed for many eukaryotic genomes. While the blueprint of cells is thus hidden within their DNA, knowing the sequence of the genomes of all eukaryotes and the chromosomal locations of all their genes has the premise to unweave the very fabric of eukaryotic life. Simultaneously, such information would allow us to peek into evolution's past and reconstruct the events that have driven the large cellular diversity of eukaryotes over the past ~2.0 billion years.

Comparative genomics of eukaryotes

For a long time the endeavor of determining the nucleotide sequence of a single gene was a slow and laborious process and could well be the main topic of one's thesis. The last three decades however, a tremendous effort to improve technologies, catalyzed by the sequencing of the ~3 billion base pair long human genome in 2001 [1,2], has resulted in the timely and cost-effective determination of whole genome sequences, including large eukaryotic genomes. The rapid and continuous influx of new sequence data is completely revolutionizing molecular and evolutionary biology, promising to provide cell

biologists with an ever broadening view on the molecular and cellular diversity of eukaryotic life, reaching far beyond the realm of the intensively researched and cherished model organisms (Figure 1).

The wealth of sequencing data ignited the development of dedicated bioinformatics tools, pipelines and models to determine the shared ancestry (i.e. homology) of genes and genomes and reconstruct the processes that governed their respective (co-)evolution. Observations such as the widespread duplication of genes and even whole genomes [3] in combination with extensive gene loss [4,5] and horizontal gene transfer (HGT, especially rampant in prokaryotes) [6], revealed the remarkable evolutionary flexibility of genomes, providing ample opportunities to detect co-evolving genes and other genomic features using phylogenetic profiling methods (see chapter 2). In addition, since homologous sequences are often functionally similar, the establishment of homology in general allows for the (partial) transfer of functional properties such as substrate specificity, binding partners and subcellular localization. Accurate evolutionary reconstruction can therefore drive the rapid expansion of functional knowledge for homologous cellular systems and allow for the global characterization of gene functions in organisms that have yet to be molecularly examined. The power of comparing genomes and genes thus lies in the correlation of functional data and evolutionary reconstructions.

Targeted sequencing of key eukaryotic genomes substantially altered the view on eukaryotic phylogeny. Although currently still debated and in continuous flux, the consensus is that there are 5 eukaryotic supergroups: Opisthokonta (Metazoa and Fungi), Amoebozoa, Excavata, SAR (Stramenopila-Alveolata-Rhizaria), and Archeaplastida [7] (Figure 1). In contrast to the large eukaryotic diversity in terms of genomic content, evolutionary reconstructions imply that the Last Eukaryotic Common Ancestor (LECA) likely possessed a complex genome [8], harboring cellular systems such as flagella and cilia [9] and complex membrane coat/tether systems that organize vesicle trafficking between membranous organelles such as the nucleus, the endoplasmic reticulum and the Golgi complex [10,11]. The findings of these studies suggest that, contrary to some popular vocations of evolutionary theory echoing in the minds of cell biologists, cellular complexity in eukaryotes is not a gradual feature that increased over time per se, but is likely ancient and many genes or genomic properties of present-day eukaryotes are therefore derived. For example, many differences in the gene repertoire between the simple model organism baker's yeast *Saccharomyces cerevisiae* and human are the result of loss in the lineage leading to yeast rather than inventions in human [12].

Altogether, many of these new ideas and concepts have precipitated over the last 20 years and are further developed in the research discipline of evolution-

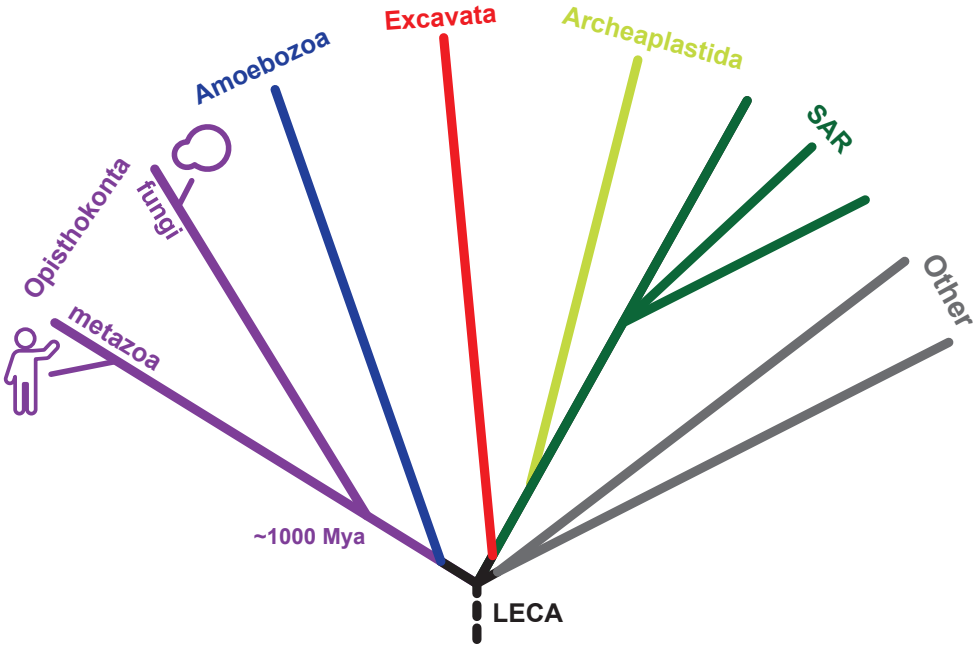


Figure 1 The eukaryotic tree of life. Depicted are the five supergroups (colors) derived from the Last Eukaryotic Common Ancestor (LECA), with several major branches [7,159]. The SAR supergroup consists of Stramenopila, Alveolata and Rhizaria. 'Other' includes species which exact phylogenetic associations are unclear, such as 'CCTH' species (Cryptista, Centrohelida, Telonemia and Haptophyta). The intensively studied metazoan and fungal lineages only represent a limited fraction of eukaryotic cellular and evolutionary diversity (see cartoon of human and budding yeast) and belong to the Opisthokont supergroup that was estimated to have diverged ~1000 Mya [160].

ary or comparative genomics. In the end evolution presents us with the results of a gigantic molecular experiment with millions of different outcomes. The reconstruction of the trajectories leading to these outcomes, using comparative genomics methods and the ever growing number of sequenced genomes, provides an exciting new outlook for understanding the principles that govern molecular systems in eukaryotic cells.

Genome replication and segregation is organized by the cell cycle

Central to the continuation of life and the process of evolution is cell multiplication. Since the genome resides within cells, the processes that involve its propagation are deeply entangled with the cell division machinery, and entail robust replication and subsequent accurate segregation into two forming daughter cells. While mutations during these processes such as replication slippage, unequal crossing over and chromosome missegregations in essence drive genome evolution, a high incidence results in loss of genomic integrity

and is incompatible with healthy cellular life, as seen for instance in cancer. To reduce these mutational events to a bare minimum and ensure overall genomic integrity, eukaryotes strictly coordinate cell division with genome replication and segregation by various complex cyclic regulatory systems, together known as the cell cycle.

In eukaryotes, the cell cycle is driven by the cyclic expression of an elaborate network of kinases (Cyclin-dependent kinases; Cdk) that self organizes into 4 distinct phases (G1-S-G2-M) [13]. G1 and G2 are two gap phases that allow the cells to grow, acquire sufficient nutrients and delay the cell cycle in case the DNA is severely damaged. During S-phase, the chromatin is replicated and the resulting sister chromatids are physically linked by cohesin rings (cohesion). Upon full replication of the genome, cells progress into the final phase of the cell cycle: mitosis (M-phase), during which the sister chromatids are physically separated and equally distributed over the two forming daughter cells.

Chromosome segregation: centromeres and microtubules

At the start of mitosis (prophase) the chromatin is hyper condensed and this is when the characteristic X-shaped chromosome becomes apparent, as sister chromatid cohesion is only maintained at a primary constriction of the chromosome termed the centromere. Intriguingly, while being the main and thus essential attachment site for the chromosome segregation machinery, centromeres consist of a myriad of different non-coding repeat regions that are characterized by a remarkable fast evolution [14,15]. Simultaneously, the centrosomes move to opposite poles of the cell to form the basis for chromosome segregation machinery called the spindle apparatus. The basic building block of this large bipolar structure is a highly dynamic protein-based biopolymer that forms long hollow tubular structures composed of α/β tubulin heterodimers, also known as microtubules. This apparent self-organizing microtubule-based spindle structure is formed through the actions of a large diversity of motor proteins, crosslinkers and other microtubule-associated proteins (MAPs) that essentially modulate the intrinsically instable (de)polymerization state of microtubules [16,17]. Upon loss of nuclear envelope integrity cells enter the second stage of mitosis: prometaphase. The sister chromatids spill into the cytoplasm and meet the highly dynamic network of growing and shrinking microtubules that emanate from both poles of the spindle.

The kinetochore – a bridge to divide

While both the DNA and microtubules are negatively charged biopolymers, they do not have the intrinsic capacity to interact. Therefore, large proteinaceous structures are built on top of centromeric chromatin to facilitate and regulate the interaction of microtubules and chromosomes and drive their seg-

regation during the remaining phases of mitosis [18]. These complex structures are known as kinetochores, a name derived from its propensity to enable chromosome movement (kinetics) in during mitosis and meiosis. Being at the heart of genome partitioning during cell division, kinetochores are important in the evolutionary process. As such, insights into kinetochore function across species, and evolution of kinetochores themselves, can inform on evolutionary mechanisms. The aim of the work described in this thesis was therefore to perform comparative genomics analyses on the ~70 protein subunits of the kinetochore, to examine its diversity amongst eukaryotes and to reconstruct its evolutionary history. In the rest of this introduction I will give a short overview of the protein composition and molecular functions of the kinetochore, highlighting some of the striking differences between eukaryotic species.

Kinetochore function and composition

The kinetochore orchestrates chromosome segregation

As the spindle itself does not have the intrinsic capacity to organize sister chromatids in a bioriented fashion and align them in the middle of the cell, kinetochores unveil their central role during prometaphase and carefully regulate and promote the progression towards a state termed metaphase; in multiple ways (see for the functions of the kinetochore, Figure 2): (1) the initial encounter of kinetochores with microtubules is often at the microtubule lattice, also known as lateral attachment. The differential actions of opposing motor proteins that localize at the kinetochore and the continuous growth and shrinkage of microtubules promotes the conversion of lateral to stable end-on attachments facilitating congression towards the 'metaphase plate' [19]. (2) Initially formed end-on attachments are often incorrect and do not result in biorientation, as sister kinetochores have no intrinsic systems to a priori favor microtubules that emanate from either side of the bipolar spindle. To correct these erroneous attachments, kinetochores act as mechanosensors, through a complex feedback system that somehow factors in the distance or tension between sister kinetochores, known as error correction. (3) To provide sufficient time for the progressive correct attachment and subsequent alignment of sister chromatids, the kinetochore integrates attachment status with the generation of a signal that halts the progression of the cell cycle and prevents sister chromatid separation, a phenomenon known as the spindle assembly checkpoint (SAC) [20]. When all sister chromatids are properly aligned and under tension, the SAC signal is silenced, the cohesin complexes that hold together sister chromatids around the centromere are cleaved and chromosome segregation commences. The last task of kinetochores is to track depolymerizing microtubules during anaphase so that chromosomes can be pulled apart and end up in the two forming daughter cells.

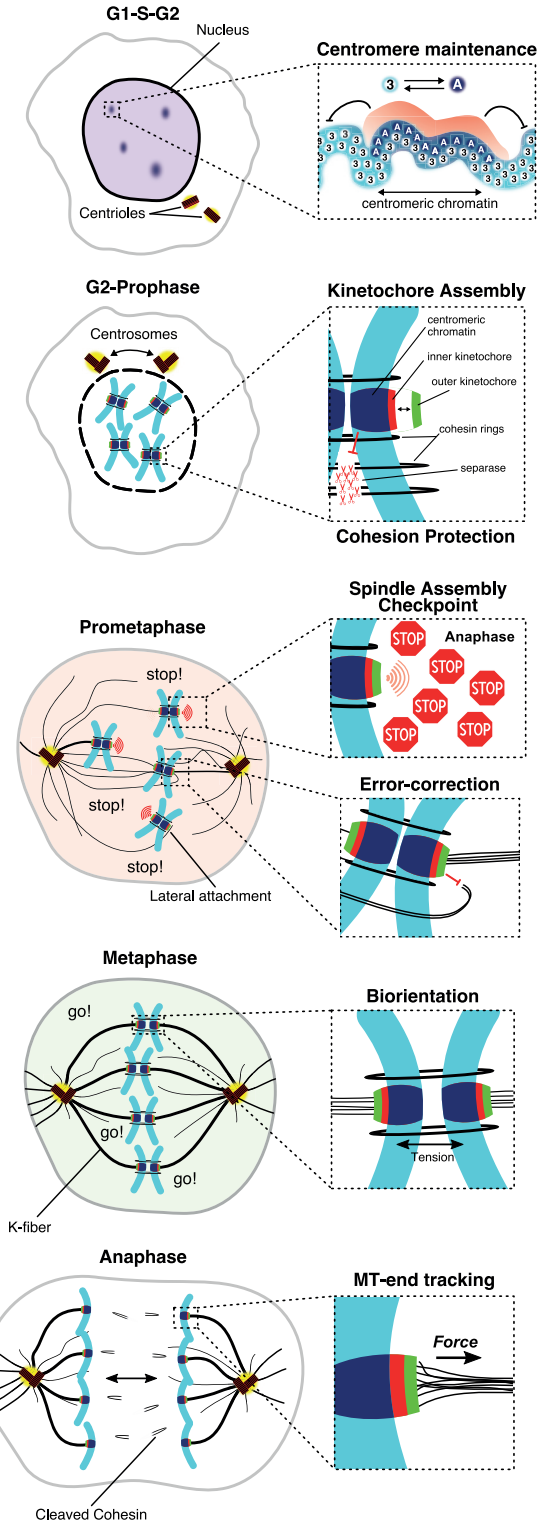


Figure 2 Functions of the kinetochore during the cell cycle. Cartoons of phases of the mitotic cell cycle including specific zoom ins of the kinetochore. **(G1-S-G2)** During interphase of the cell cycle, a part of the kinetochore complex (red: inner kinetochore) resides at the centromere and is involved in the maintenance of centromeric chromatin and the loading of cohesin complexes. **(Prophase)** Just before mitosis, sister chromatids hyper condense and the centrosomes start to move apart along a disintegrating nuclear envelope to form the basis for a bipolar spindle. Kinetochores mature and inner kinetochore complexes (red) direct the recruitment of microtubule-binding complexes of the outer kinetochore (green). In addition kinetochores specifically protect cohesin rings at the centromeric chromatin against cleavage by the enzyme separase, resulting in the characteristic X-shape of chromosomes. **(Prometaphase)** When the nuclear envelope is completely broken down, sister chromatids meet the dynamic microtubules of the spindle. Unattached kinetochores catalyze the formation of a stop signal that prevents the progression of mitosis into anaphase. While the microtubules of the spindle have no intrinsic capacity to distinguish between any of the two sister chromatids, kinetochores prevent the formation of improper attachment via a mechanism known as error-correction. **(Metaphase)** When all sister chromatids are properly attached, they are directed to the 'metaphase plate' and tension builds up as microtubules from opposing poles of the spindle promote kinetochore biorientation. While the SAC is silenced and all cohesin complexes are cleaved, cells can progress into anaphase and chromosome segregation can commence. **(Anaphase)** The last task of kinetochores is to hold on to the depolymerizing microtubules of the spindle, which ensures the delivery of two packs of identical chromosomes in the two forming daughter cells.

Kinetochore assembly onto centromeric chromatin

CenpA is the basis for kinetochore assembly centromere identity

Centromeres come in all shapes and sizes, and are highly unstable throughout eukaryotic evolution [14,15,21,22]. While in *Saccharomyces cerevisiae* the centromere is a single short sequence named CEN consisting of 3 recognizable regions (CDEI-III; point centromeres), the centromeres of most eukaryotic species consist of large arrays of higher order repeat regions (up to ~1.10^6 bp in humans; regional centromeres) and some even cover the entirety of a chromosome, a phenomenon also known as holocentrism (i.e in nematodes, various winged insects and lower plants; holocentromeres) [23–25]. Interestingly, it was found in multiple model organisms that the repetitiveness of the centromeric DNA is not an absolute requirement for centromere identity. Rather the loading and maintenance of a specific Histone H3 variant CenpA (Centromeric Protein A) marks the position of the centromere and constitutes the specific chromatin environment needed for building a kinetochore (Figure 3a, see for discussion [18,26]). Strikingly, CenpA loading occurs through different histone complexes in diverse eukaryotic species i.e. Scm3 (fungi), Cal1 (flies) and HJURP/MIS18BP (mammals) [18,26]. Remarkably, both CenpA and its specific histone loading complexes evolve under positive selection in a number of eukaryotes [27,28], indicating that centromeres and kinetochores are in a continuous (epi) genetic conflict reminiscent of what has been observed for viruses and hosts, a phenomenon known as centromere drive [29,30]. Notable species-specific exceptions to the sequence-independent centromere paradigm exist, such as the CEN directed recruitment of Cbf1, the 4-subunit CBF3 complex and Cse4 (scCenpA) in budding yeasts [31] and the transposon-like protein CenpB that binds to a 17-bp CenpB box present in mammalian centromeres [32].

CenpC and CenpT constitute the main centromere-microtubule axis

CenpA-containing nucleosomes direct the recruitment of a 16-subunit complex of which a subset is constitutively localized at the centromere throughout the cell cycle, hence its name: Constitutive Centromere-Associated Network (CCAN) [33–35]. The CCAN can be subdivided into 4 distinct sub complexes: CenpTWSX, CenpOPQRU, CenpLN, CenpHIKM [18] of which the latter two are organized through their scaffolding protein CenpC to form a dimer that selectively binds and stabilizes CenpA over canonical Histon 3-containing nucleosomes (Figure 3a) [36]. The four members of the CenpTWSX complex all contain histone-fold domains and were suggested to form a bona fide nucleosome structure [37]. Clear evidence however is lacking and it remains to be seen how the DNA-binding capacity of this complex adds to the chromatin environment at centromeres [18]. CenpOPQRU recruitment is dependent on CenpCHIKMLN [38] and has been shown to be involved in chromosome con-

gression through the microtubule-binding sites in CenpQ and association with the motor protein CenpE [39]. CenpU has been implicated in the recruitment of the mitotic kinase Plk1 [40–42]. In addition, various species-specific CCAN subunits were reported, such as Nkp1 and Nkp2 in various fungi and Fta6 in fission yeast [43–45]. Although it remains to be seen how all the different sub complexes of the CCAN exactly contribute to maintenance of centromeric stability, kinetochore assembly and the regulation of kinetochore-microtubule interactions, a number of recent studies point to specific (phospho) motifs in the disordered N-termini of CenpC and CenpT for the recruitment [46–49] of the 10-subunit Knl1-C/Mis12-C/Ndc80-C (KMN) network, which constitutes the main microtubule-binding module of the kinetochore (Figure 3a) [50]. The striking finding that, except for CenpC, the largest part of the CCAN has been recurrently lost throughout eukaryotic evolution seems to further underline the crucial role of CenpC in establishing the centromere-microtubule axis [51,52] (see chapter 2 and 3).

Binding microtubules: the KMN network

The KMN network connects to microtubules and is the most intensively studied of all kinetochore complexes for which 3D structures have been recently resolved in high detail [46,47,53–55].

The Mis12 complex connects the CCAN to microtubule-binding proteins

The 4-subunit Mis12 complex is most proximal to the CCAN and constitutes a Y-shaped structure consisting of two heads that are formed by the heterodimers Mis12:Nnf1 and Dsn1:Nsl1 and a stalk that is formed through the tetramerization of long C-terminal coiled-coil segments of all 4 subunits [46,47]. The Mis12:Nnf1 head interacts with a rapidly evolving hydrophobic patch in the N-terminus of CenpC [46,47,49,54]. While no explicit function could be assigned to the Dsn1:Nsl1 head, it was hypothesized that a head-to-head orientation of ~5-6 Y-shaped Mis12 complexes could explain the circular superstructure that was observed for the budding yeast kinetochore complex [18,47,56]. The large, disordered N-terminus of Dsn1 contains numerous linear (phospho) motifs that potentially modulate the function of the Mis12 complex and might well drive the recruitment of different temporary kinetochore components such as the monopolin component Csm1 [46,47,57] (see chapter 3). The stalk of the Y-shaped Mis12 complex points away from the centromere and linear (phospho) motifs in the C-termini provide a landing platform the RWD domain-containing subunits that reside within the Knl1 complex and Ndc80 complex [54,58]. Interestingly, a parallel pathway via Cyclin-dependent kinase 1 (Cdk1)-mediated phosphorylation of CenpT competes for the (additional) recruitment of both Mis12 and Ndc80 complexes [48,58], creating many opportunities for the differential localization of microtubule-binding complexes. How these dynamics

may play a role in kinetochore-microtubule interactions is not yet understood.

The Knl1 and Ndc80 complex interact with microtubules

The Knl1 complex consists of Knl1 and Zwint-1 (Zw10-interacting 1). Knl1 consists of a long disordered region that is involved in the recruitment of proteins involved in error-correction and SAC signalling (see sections on 'Error correction' and 'Spindle Assembly Checkpoint') and a C-terminal coiled-coil and RWD domain drives recruitment towards the kinetochore [59]. The positively charged N-terminus has low affinity for microtubules and harbors phosphatase-recruitment motifs that play a role in biorientation (see section on 'Error correction') [60–62]. Although the name of Zwint-1 suggests an interaction with ZW10 [63], recent experiments suggest that not Zwint-1 itself, but rather its stabilizing effect on Knl1 aids the recruitment to the Zw10-containing Rod-Zwilch-Zw10 (RZZ) complex [52,64] (see chapter 2,3). The primary microtubule-binding capacity of the KMN network resides within heterotetrameric Ndc80 complex, which seems to be the most conserved protein in the kinetochore network [52] (see chapter 2) and has been shown to bind microtubules in a number of species [50,65–69]. The Ndc80 complex forms a ~65 nm long coiled-coil rod consisting of two heterodimeric complexes of which the globular RWD domains of the Spc24:Spc25 dimer face the kinetochore while the positively charged Calponin-homology (CH) domains of the Ndc80:Nuf2 dimer project into the cytoplasm [70] and can interact with the lattice of microtubules [50,71]. In addition, the differential phosphorylation of the highly positively charged disordered N-terminal tail of Ndc80 allows for the modulation of the affinity of the Ndc80 complex for microtubules [50,53,65,71–73].

Accessory proteins modulate microtubule dynamics at the kinetochores

Although the Ndc80 complex is the major microtubule-binding complex, a host of proteins dynamically localize at the kinetochore to adapt to and drive the different modes of microtubule dynamics that can be observed at the kinetochore throughout mitosis (see for review [19]). We here mention only a few. In prometaphase when sister chromatids are not yet fully exposed to microtubules, EM studies have revealed that a currently understudied 'fibrous corona' enlarges the surface of kinetochores [74,75] and thereby likely increases the chance of lateral kinetochore attachment to microtubules. A number of proteins have been reported to associate to this layer of the kinetochore i.e. the microtubule plus-end directed motor protein CenpE [76] and several factors that recruit the microtubule minus-end directed motor protein dynein, such as CenpF [77] and the RZZ-Spindly complex [78,79]. As kinetochores have to specifically stabilize correct end-on attachments a number of microtubule-plus end tracking proteins modulate microtubule behaviour i.e. Astrin, SKAP [80], the ch-TOG homolog in budding yeast, Stu2 [81] and the Mitotic cen-

tromere-associated kinesin (MCAK), a protein with microtubule depolymerizing activity [82]. To track depolymerizing microtubules, budding yeast kinetochores specifically localize the DASH or Dam1 complex which forms a ring around the microtubules and docks onto the Ndc80 complex utilizing a loop in Ndc80 (Figure 3a) [83,84]. Mammalian kinetochores, which do not possess the Dam1 complex, localize the W-shaped SKA complex that also binds the Ndc80 complex to perform a similar function by interacting with the curved ends of depolymerizing microtubules (Figure 3a) [85,86]. Interestingly, during eukaryotic evolution the Dam1 and SKA complex evolve in a mutually exclusive manner, suggesting that somehow these complexes interfere at a functional level and cannot co-exist [87].

Bioriented sister chromatids: error correction

While initial kinetochore-microtubule attachments are often incorrect and do not result in the biorientation of sister chromatids [88], a number of kinetochores have to go through multiple cycles of microtubule destabilization and reengagement in a process termed error-correction [89]. Error-correction emerges from the opposing actions of the inner centromere-localized Chromosomal Passenger Complex (CPC) and several phosphatases that localize to the outer kinetochore upon the progressive stabilization of correct microtubule attachments [89]. The CPC consists of 3 non-catalytic subunits (Incenp, Borealin Survivin) that localize to the inner centromere during prometaphase to activate and specifically localize its catalytic subunit, the kinase Aurora B. While sister kinetochores are under low tension (a measure for incorrect attachments), Aurora B kinase drives error-correction through the phosphorylation of a great number of sites in positively charged regions of proteins that engage microtubules such as the N-terminal tails of Ndc80 and Knl1 [61,65], the motor protein CenpE [90] and proteins of the SKA complex [91,92]. The addition of negative charges effectively neutralizes the positively charged regions, likely resulting in the repulsion of the negatively charged tubulin and subsequent attachment destabilization. As kinetochore-microtubule attachments become increasingly correct, sister kinetochores are put under tension, as they are pulled apart by forces exerted from both ends of the spindle. Previous models proposed that the spatial restriction of Aurora B at the inner centromere would prevent the phosphorylation of its outer kinetochore substrates, allowing the dephosphorylation at progressive stabilization of bioriented attachments [93]. Recent work in both budding yeast and human cells however, suggests that inner centromere localization of Aurora B is not essential for error-correction [94,95] and could potentially point to an active role for the recruitment of phosphatases at the outer kinetochore to dampen excessive aurora B-mediated phosphorylation. Indeed, Shugoshin-1 and Knl1-bound BubR1 [96–98], localize PP2A-B56 to the inner centromere and outer kinetochore, respectively, to modulate Aurora

B driven destabilization of attachments. Subsequent recruitment of the phosphatase PP1 by the outer kinetochore protein Knl1 [60–62] and members of the SKA complex [99,100], results in even more dephosphorylation and the progressive stabilization of kinetochore-microtubule attachments.

Spindle assembly checkpoint

As all kinetochore-microtubules are stabilized and sister chromatids are aligned at the metaphase plate, cells progress into anaphase and chromosomes are physically separated, destined to end up in the forming daughter cells. The transition of cells into anaphase is tightly regulated by an elaborate feedback system known as the spindle assembly checkpoint (SAC), which integrates the microtubule-attachment status of kinetochores with the activity of the cell cycle progression machinery in the form of the Anaphase Promoting Complex/Cyclosome (APC/C) [101]. Strikingly, several Alveolate species (e.g. *Tetrahymena thermophila* and *Toxoplasma gondii*) lack the molecular components of the SAC [52], and the excavate *Giardia intestinales* [102,103] is even devoid of the APC/C altogether (see chapter 2). The loss of the APC/C and various SAC components in these species indicates the absence of a conventional checkpoint system and suggest a high degree of evolutionary flexibility for these molecular systems in eukaryotes.

The APC/C targets Cyclin B1 and Securin for degregation

The APC/C is a large 16-subunit E3 ubiquitin-ligase that consists of three subcomplexes: the catalytic core (Apc2 and Apc11), a scaffolding platform (Apc1) and an 'arc' that contains multiple copies of different TPR domain-containing proteins (Apc3, Apc6-8) [104–107]. Together these subcomplexes enclose a cavity that forms the basis for the association of multiple E2 ubiquitin-conjugating enzymes (Ube2S and Ube2C) [108,109] and the specific recruitment of substrate-bound co-activators (Cdc20 and Cdh1). The recognition of substrates by the WD-40 propeller domains of Cdc20 and Cdh1 relies on the interaction of a number of low-affinity short linear motif (SLiMs) or degrons that can be divided into 3 classes (see for review [110]): (1) the Destruction-box (D-box; Arginine-x-x-Leucine) sandwiches in between Apc10 and Cdc20, (2) the KEN-box (Lysine-Glutamate-Asparagine) that binds on the top of the Cdc20 propeller and (3) the ABBA-motif (Acm1-Bub1-BubR1-cyclinA) that binds a hydrophobic pocket at the side of the Cdc20 molecule [111–114] (see chapter 6). The sequential recruitment of Cdc20 and Cdh1 to the APC/C provides temporal regulation of the degradation of specific substrates, which is a main driver of the cell cycle [107,115]. To promote the metaphase-to-anaphase transition, APC/C^{Cdc20} targets two essential mitotic regulators for proteasomal degradation: (1) Cyclin B1, a co-factor of Cyclin-dependent kinase 1 (Cdk1), which sustains mitotic activities as long as it is present [116], and (2) Securin, an inhibitor of

the enzyme Separase, which cleaves the cohesin rings that hold together the sister chromatids during mitosis [117,118]. Upon its activation, the APC/C thus drives cells out of mitosis and allows for the loss of sister chromatid cohesion and their subsequent separation [20,119].

The SAC signal is generated at unattached kinetochores

To prevent the progression of mitosis and the premature loss of sister chromatid cohesion before all chromosomes are stably attached to the mitotic spindle, unattached kinetochores catalyze the formation of a soluble complex (the Mitotic Checkpoint Complex, MCC) that sequesters Cdc20, thereby forming a pseudo-substrate of APC/C^{Cdc20} that clogs its substrate-binding cavity [120,121]. Early experiments using laser-ablation of both microtubules and kinetochores suggested that the SAC operates like an 'on/off switch' (hence the name checkpoint) that could be generated and maintained by as little as one kinetochore [122]. Recent studies using various drugs that modulate spindle-microtubule dynamics, however, indicate that the SAC response correlates linearly with the amount of unattached kinetochores and operates more like a 'rheostat' [123,124]. Furthermore, a long standing discussion in the field on how the SAC is satisfied was recently resolved by some elegant experiments that showed that the SAC does not sense forces that are generated upon sister kinetochore biorientation, but that kinetochore-microtubule attachments alone are sufficient for the initiation of anaphase [125,126]. Over the past two decades many molecular aspects of SAC signaling have been elucidated [101]. The generation and maintenance of a SAC signal mainly involves the following 3 classes of proteins (Figure 3b): (1) subunits of the APC/C inhibitory complex MCC (BubR1 (Mad3 in budding –and fission yeast), Bub3, Mad2 and Cdc20), (2) activators that initiate and catalyze MCC production at the kinetochore (Mps1, Bub1 and Mad1) and (3) scaffolds that localize all checkpoint proteins to outer kinetochore (Knl1, Ndc80 and the RZZ complex).

Mps1 directs all steps of MCC production at kinetochores

The production of the MCC at kinetochores entails the localization and formation of a 4-subunit complex (BubR1, Bub3, c-Mad2 and Cdc20), in a multi-step cascade that is primarily catalyzed by the activity of Mps1 kinase (Figure 3b) [127,128]. Although this model fits with the largest part of the literature, formally it has to be discerned if the MCC is formed at the kinetochore or whether this occurs in the cytosol. (1) During mitosis, the promiscuous kinase Mps1 is localized to unattached Ndc80 complexes through the interaction with the CH-domains of the Ndc80:Nuf2 dimer [129,130]. In vertebrates, a large disordered region in Knl1 contains repeated units that harbor multiple different phosphomodules (TxxΩ, MELT and SHT) that are sequentially phosphorylated by Mps1 and interact with Bub3:Bub1 dimer via positively charged patch-

es on the flanks of the WD40 structure of Bub3 [131,132] (see also chapter 4). While Bub3 is one of the most conserved proteins at the kinetochore, the repeated units in Knl1 show remarkable widespread divergence of both its number and sequence, reminiscent of virus-host type evolution (see chapter 5) [133]. (2) Subsequently, Bub1 recruits the heterotetramer Mad12:Mad22 through the phosphorylation of several residues in a region called CDI or CMI (Conserved Motif I) [127,128,134]. Binding of Mad2 to a short linear motif in Mad1 (Mad2-binding motif) induces a conformation switch from the open (O-Mad2) to a closed form (C-Mad2) [135,136]. The tetramer, containing two molecules of C-Mad2, catalyzes the conversion of soluble O-Mad2 to C-Mad2 in a prion-like manner, thereby priming these molecules for the incorporation into the MCC through interaction with a Mad2-binding motif in the disordered N-terminus of Cdc20 [137–139]. While in budding yeast Bub1 stably recruits Mad1 [140], Bub1 in humans is only involved in initial recruitment of Mad1 as Bub1 depletion does not affect the final levels of Mad1 at kinetochores upon nocodazole treatment [141]. Multiple experiments point in the direction of the RZZ (Rod-Zwilch-Zw10) complex for the stable recruitment of Mad1 at the kinetochore [63,142,143], but more work is needed to understand how this works. (3) As C-Mad2, BubR1 and Bub3 are localized at the kinetochore, the last step is the recruitment and incorporation of Cdc20. How this exactly works is not yet fully understood. Both Bub1 and BubR1 harbor multiple low-affinity binding ABBA-motifs that engage the lateral side of the Cdc20 propeller [111,113,114,141,144]. In addition, Mps1 phosphorylates the N-terminal tail of Mad1, increasing its affinity for Cdc20 [127].

The MCC is a pseudo-substrate inhibitor of APC/C^{Cdc20}

Upon its formation by the kinetochore, the MCC diffuses into the cytosol to engage the APC/C. While former models always assumed that either sequestration of Cdc20 in the MCC or the inhibition of a 3-subunit MCC (BubR1:Bub3:C-Mad2) of APC/C^{Cdc20} were the basis for MCC-mediated inhibition of the APC/C, the Pines lab elegantly showed that a 4-subunit MCC (including Cdc20) locks APC/C^{Cdc20} that is primed for activation [145]. Recent cryoEM (cryogenic Electron Microscopy) studies confirmed these findings and revealed that BubR1 acts as a pseudo-substrate of the APC/C through the simultaneous engagement of two Cdc20 molecules (Cdc20^{MCC}, Cdc20$^{APC/C}$) [108,121]. While a number of APC degrons were previously mapped to BubR1, such as a D-box [120] and two KEN-boxes [120,146], recent analyses revealed two ABBA-motifs that flank the second KEN-box and provide a symmetrical platform to alter the position of both Cdc20 molecules in such a way that APC/C activity is abrogated [113,114,121] (see also chapter 6). The evolution of BubR1/Mad3-like and Bub1 present an interesting case. Previous analyses from our lab found that these genes independently duplicated nine times during eukaryotic evolution.

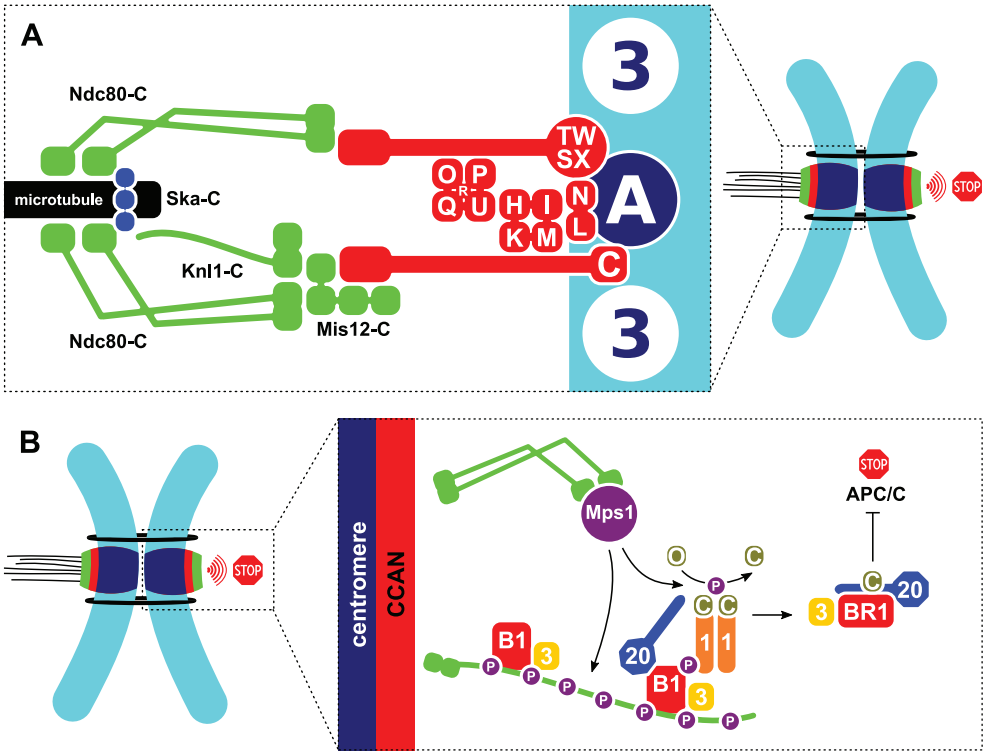


Figure 3 Molecular organization of the kinetochore and the spindle assembly checkpoint (SAC). (A) Schematic model of the kinetochore. The 16 subunits of the CCAN (red) connect the centromere-specific histone CenpA (depicted as 'A') with the microtubule-binding complexes of the KMN network (green; Knl1-C, Nkdc80-C and Mis12-C). The Ska or Dam1 complex is involved in the tracking of depolymerizing microtubules. **(B)** Unattached kinetochores recruit the promiscuous kinase Mps1, which upon phosphorylation of the kinetochore scaffold Knl1 and various SAC pathway members directs the rapid activation of the spindle assembly checkpoint. This signaling cascade involves the recruitment of Bub1 (B1) and Mad1 (1), which ultimately results in the conversion of open (O) to closed (C) Mad2. Closed Mad2 facilitates the binding of BubR1 (BR1) and Bub3 (3) to Cdc20 (20) that together form the soluble inhibitor of the APC/C.

While many eukaryotic species still harbor the ancestral gene we termed Mad-Bub, subfunctionalization of both Bub1-like and BubR1/Mad3-like occurred in a remarkable similar fashion: losing either the N-terminal KEN-box or the C-terminal kinase domain [147] (see also chapter 6).

SAC silencing

When the last kinetochore is attached, the SAC is rapidly silenced and the APC/C is activated to promote the entry into anaphase. SAC silencing is thought to occur through 3 pathways: (1) Upon stabilization of attachments, the dynein-dynactin-adaptor Spindly promotes the processivity of dynein and

strips the kinetochore of its main cargo, the RZZ complex [78,79,148,149]. While the Mad1:Mad2 tetramer is mostly localized to the kinetochore by the RZZ complex, dynein activation leads to the removal Mad1 and Mad2 from the kinetochores [148,149]. (2) The MCC is specifically disassembled at the APC/C. Apc15, a core subunit of the APC/C, drives the ubiquitination of Cdc20, which was found to destabilize the MCC [150,151]. Furthermore, the divergent HORMA domain protein p31^{comet} promotes MCC disassembly [152–154]. p31^{comet} dimerizes with C-Mad2 and is an adapter for the AAA+ ATPase Trip13, which upon binding, converts C-Mad2 back to the catalytically inactive O-Mad2 and thereby drives MCC disassembly [155–157]. (3) The influx of PP1 phosphatase counters both Mps1 and Aurora B-mediated phosphorylations and promotes SAC silencing through the complete shutdown of MCC production at the kinetochore [158].

Scope of this thesis

In this thesis we describe comparative genomics studies of protein subunits of the kinetochore network in a large diversity of eukaryotic genomes. We utilize this information to reconstruct the kinetochore network of the Last Eukaryotic Common Ancestor (LECA) and to track how 2 billion years of evolution have shaped the kinetochores of present-day eukaryotic species. In addition we develop an approach to trace the sequence (co-)evolution of kinetochore protein families in high detail. Functional hypotheses emanating from these analyses are tested by RNAi-based protein knockdown and reconstitution in human cells. The effects of targeted mutations are assayed by quantitative immunofluorescence microscopy, time-lapse imaging and immunoprecipitation. We look into the evolutionary dynamics of a large set of kinetochore-associated proteins and discuss the patterns that we observe in light of their respective functions and the diversity of chromosome segregation mechanisms in eukaryotes.

Outline

In **chapter 2**, we perform a large-scale phylogenomic analysis of proteins associated to the kinetochore network in a set of 90 genomes that are representative of the extensive genomic diversity throughout the eukaryotic tree of life. Our reconstructions imply that LECA possessed an elaborate kinetochore network that subsequently diverged by rapid sequence evolution, extensive gene loss and duplication, and in some cases invention and replacement, as seen in many current-day eukaryotic lineages. In **chapter 3**, we assess the quality of our established ortholog sets and develop a de novo sequence discovery workflow to track the eukaryote-wide (co-)evolution of short linear motifs, domains and proteins within the kinetochore network. In **chapter 4**, we discover an

array of 19 phosphomotif-containing repeat units in the disordered N-terminal half of the human outer kinetochore scaffold Knl1. We model these units in common tissue culture cell lines and find that they operate as independent docking sites for BUB proteins, of which only a limited number is sufficient for accurate chromosome segregation. In **chapter 5**, we trace the evolution of Knl1 repeat arrays in a wide variety of eukaryotes. Our comparative analyses reveal that these arrays diverged extensively both in terms of amino acid sequence composition as well as the number of repeats. Extensive species-specific array reorganization in combination with modular repeat evolution points to widespread recurrent episodes of concerted Knl1 repeat evolution. In **chapter 6**, we present an elaborate subfunctionalization analysis of the Bub1/BubR1 gene family, which independently duplicated at least 15 times during eukaryotic evolution. Using our workflow established in chapter 3, we trace the distribution of ancestral sequence features to extant paralogs and discover a conserved cassette of short linear motifs that is essential for the SAC. Our research has laid the groundwork for examining kinetochore function in diverse eukaryotes. The implications of our work and our views on the future of comparative molecular cell biology of kinetochores are discussed in **chapter 7**.

CHAPTER 2

Evolutionary dynamics of the kinetochore network as revealed by comparative genomics

Jolien JE van Hooff, Eelco Tromer, Leny M van Wijk,
Berend Snel & Geert JPL Kops

EMBO reports, 2017

Abstract

During eukaryotic cell division, the sister chromatids of duplicated chromosomes are pulled apart by microtubules, which connect via kinetochores. The kinetochore is a multiprotein structure that links centromeres to microtubules, and that emits molecular signals in order to safeguard the equal distribution of duplicated chromosomes over daughter cells. Although microtubule-mediated chromosome segregation is evolutionary conserved, kinetochore compositions seem to have diverged. To systematically inventory kinetochore diversity and to reconstruct its evolution, we determined orthologs of 70 kinetochore proteins in 90 phylogenetically diverse eukaryotes. The resulting ortholog sets imply that the last eukaryotic common ancestor (LECA) possessed a complex kinetochore and highlight that current-day kinetochores differ substantially. These kinetochores diverged through gene loss, duplication, and, less frequently, invention and displacement. Various kinetochore components co-evolved with one another, albeit in different manners. These co-evolutionary patterns improve our understanding of kinetochore function and evolution, which we illustrated with the RZZ complex, Trip13, the MCC, and some nuclear pore proteins. The extensive diversity of kinetochore compositions in eukaryotes poses numerous questions regarding evolutionary flexibility of essential cellular functions.

Introduction

During mitotic cell division, eukaryotes physically separate duplicated sister chromatids using microtubules within a bipolar spindle. These microtubules pull the sister chromatids in opposite directions, toward the spindle poles from which they emanate [161]. Current knowledge indicates that all eukaryotes use microtubules for chromosome separation, suggesting that the last eukaryotic common ancestor (LECA) also did. Microtubules and chromatids are connected by the kinetochore, a multiprotein structure that is assembled on the centromeric chromatin [162,163]. Functionally, the kinetochore proteins can be subdivided into three main categories: proteins that connect to the centromeric DNA (inner kinetochore), proteins that connect to the spindle microtubules (outer kinetochore), and proteins that perform signaling functions at the kinetochore in order to regulate chromosome segregation. These signaling functions consist of the spindle assembly checkpoint (SAC), which prevents sister chromatids from separating before all have stably attached to spindle microtubules, and attachment error correction, which ensures that these sister chromatids are attached by microtubules that emanate from opposite poles. Together, the SAC and error correction machineries ensure that both daughter cells acquire a complete set of chromosomes.

Although microtubule-mediated chromosome segregation is conserved across eukaryotes, their mitotic mechanisms differ. For example, some species, such as those in animal lineages, disassemble the nuclear envelope during mitosis ("open mitosis"), while others, such as yeasts, completely or partially maintain it ("(semi-) closed mitosis") [164]. Species differ also in their kinetochore composition, both in the inner and in the outer kinetochore. For example, *Drosophila melanogaster* and *Caenorhabditis elegans* lack most components of the constitutive centromere-associated network (CCAN), a protein network in the inner kinetochore. In the outer kinetochore, diverse species employ either the Dam1 (e.g., various Fungi, Stramenopiles, and unicellular relatives of Metazoa) or the Ska complex (most Metazoa and Viridiplantae and some Fungi) for tracking depolymerizing microtubules [87]. The kinetochore of the excavate species Trypanosoma brucei mostly consists of proteins that do not seem homologous to the "canonical" kinetochore proteins [165,166]. Studying the evolution of kinetochore proteins revealed how kinetochore diversity was shaped by different modes of genome evolution: The inner kinetochore CenpB-like proteins were recurrently domesticated from transposable elements [51], the outer kinetochore protein Knl1 displays recurrent repeat evolution [133], the SAC proteins Bub1/BubR1/Mad3 (MadBub) duplicated and subfunctionalized multiple times in eukaryotic evolution [114,147], and the SAC protein p31^{comet} was recurrently lost [167].

Prior comparative genomics studies reported on kinetochore compositions in eukaryotes [167,168]. These studies raised various questions: Are kinetochores in general indeed highly diverse? How often do kinetochore proteins evolve in a recurrent manner in different lineages? How frequent is loss of kinetochore proteins? Does the kinetochore consist of different evolutionary modules? To address these and other questions, we studied the eukaryotic diversity of the kinetochore by scanning a large and diverse set (90) of eukaryotic genomes for the presence of 70 kinetochore proteins. We deduced the kinetochore composition of LECA and shed light on how, after LECA, eukaryotic kinetochores diversified. To understand this evolution functionally, we detected co-evolution among kinetochore complexes, proteins and sequence motifs: Co-evolving kinetochore components are likely functionally interdependent. Furthermore, we found that certain species contain yet inexplicable kinetochore compositions, such as absences of proteins that are crucial in model organisms. We nominate such species for further investigation into their mitotic machineries.

Results

Eukaryotic diversity in the kinetochore network

We selected 70 proteins that compose the kinetochore (see Materials and Methods). For comparison, we also included proteins that constitute the anaphase-promoting complex/cyclosome (APC/C), which is targeted by kinetochore signaling. We identified orthologous sequences of these kinetochore and APC/C proteins in 90 diverse eukaryotic lineages by performing in-depth homology searches. Our methods were aimed at maximizing detection of a protein's orthologs even if it evolves rapidly, which is the case for many kinetochore proteins (as we discuss below). The resulting sets of orthologous sequences are available (Sequence File S1). We projected the presences and absences of proteins ("phylogenetic profiles") across eukaryotes (Figure 1, Materials and Methods). In spite of our thorough homology searches, for some proteins the ortholog in a given species might have diverged too extensively to recognize it, resulting in a "false" absence. We however think that, globally, our analysis gives an accurate representation of kinetochore proteins in eukaryotes (Discussion).

We inferred the evolutionary histories of the proteins by applying Dollo parsimony, which allows only for a single invention and infers subsequent losses based on maximum parsimony. Of the 70 kinetochore proteins, 49 (70%) were inferred to have been present in LECA (Figure 1, Figure 2a,c). CenpF, Spindly and three subunits of the CenpOPQRU complex probably originated more recently. The Dam1 complex likely originated in early fungal evolution and may have propagated to non-fungal lineages via horizontal gene transfer [87].

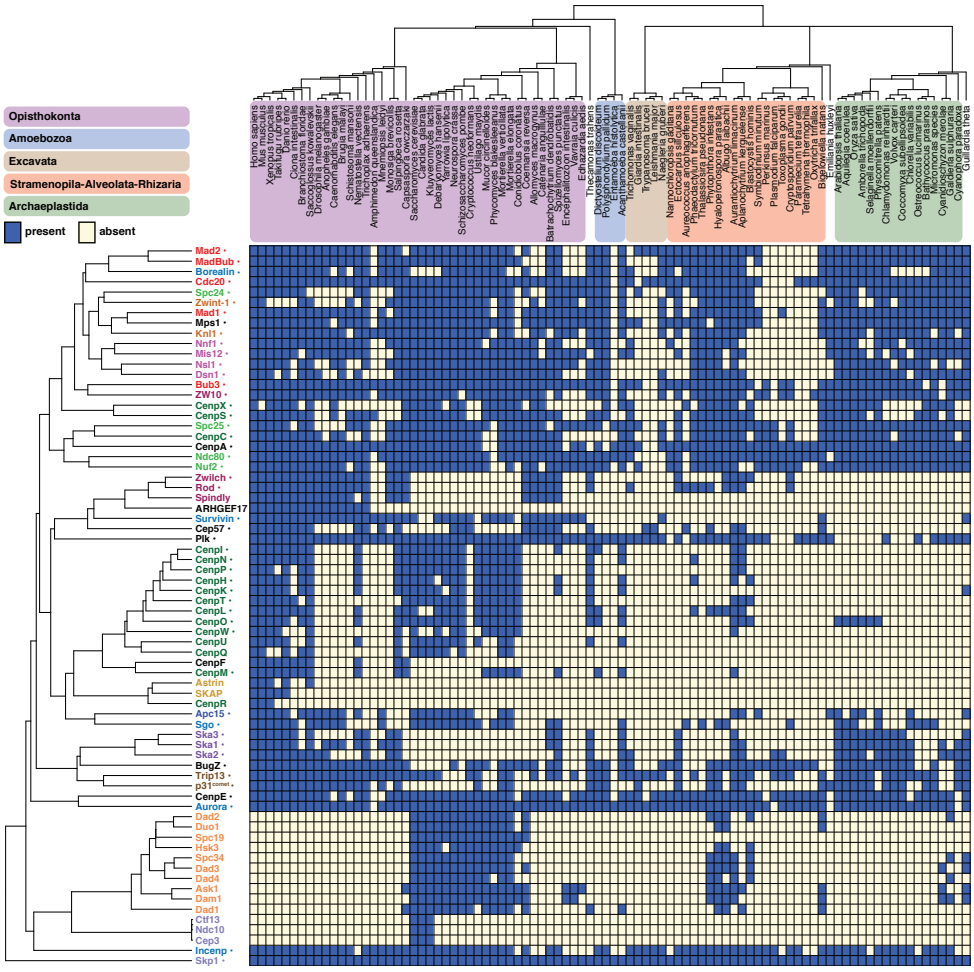


Figure 1 The kinetochore network across 90 eukaryotic lineages. Presences and absences ("phylogenetic profiles") of 70 kinetochore proteins in 90 eukaryotic species. **Top:** Phylogenetic tree of the species in the proteome set, with colored areas for the eukaryotic supergroups. **Left side:** Kinetochore proteins clustered by average linkage based on the pairwise Pearson correlation coefficients of their phylogenetic profiles. Protein names have the same colors if they are members of the same complex. Proteins inferred to have been present in LECA are indicated (*). The orthologous sequences (including sets of APC/C subunits, Nag, Rint1, HORMAD, Nup106, Nup133, Nup160) are available as fasta files in Sequence File S1, allowing full usage of our data for further evolutionary cell biology investigations.

Kinetochore proteins are less conserved than APC/C subunits (Figure S1, Table S1 [169]). Species on average possess 48% of the kinetochore proteins, compared to 70% of the APC/C subunits. Species that we predict to contain relatively few kinetochore proteins include *Tetrahymena thermophila* (Figure 2b) and *Cryptococcus neoformans* (Figure 2d). Some kinetochore proteins are absent from many different lineages, likely resulting from multiple indepen-

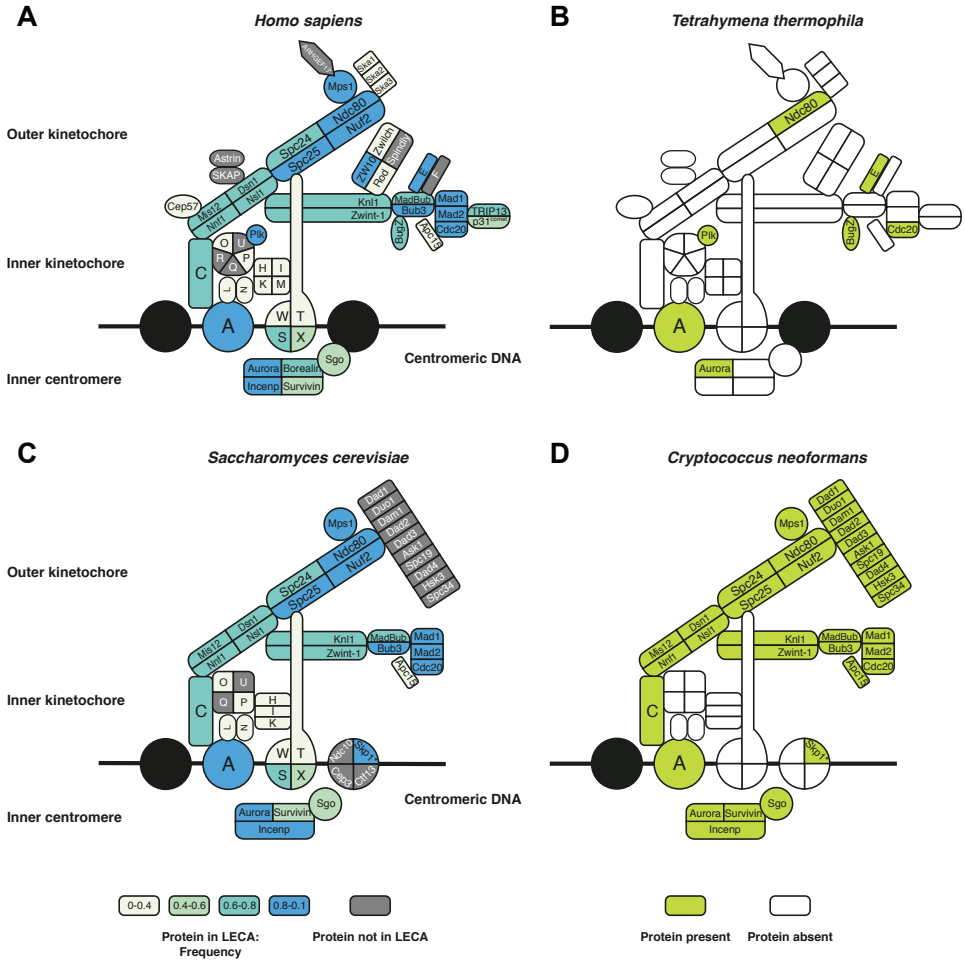


Figure 2 Kinetochores of model and non-model species. (A) The human kinetochore. The colors of the proteins indicate if they were inferred to be present in LECA and their occurrence frequency across eukaryotes (see Materials and Methods). **(B)** The predicted kinetochore of *Tetrahymena thermopila* projected onto the human kinetochore. **(C)** The budding yeast kinetochore; similar to panel B. **(D)** The predicted kinetochore of *Cryptococcus neoformans* projected onto the budding yeast kinetochore.

dent gene loss events. We counted losses of kinetochore and APC/C proteins during post-LECA evolution using Dollo parsimony. On average, kinetochore proteins were lost 16.5 times since LECA, while APC/C proteins were lost 13.1 times (not significantly different for kinetochore versus APC/C). Our homology searches hinted at some kinetochore proteins evolving also rapidly on the sequence level. The kinetochore proteins indeed have relatively high dN/dS values, a common measure for sequence evolution: When comparing mouse and human gene sequences, kinetochore proteins scored an average dN/dS of

0.24, compared to 0.06 for the APC/C proteins (P = 0.0016) and 0.15 for all human proteins (P = 4.8e-5). The loss frequency and sequence evolution seem to be correlated, suggesting a common underlying cause for poor conservation (Figure S2, Discussion). Overall, the kinetochore seems to evolve more flexibly than the APC/C.

We not only mapped the presences and absences of kinetochore proteins, but also counted their copy number in each genome (Figure S3). As observed before, MadBub and Cdc20 are often present in multiple copies. These proteins likely duplicated in different lineages and subsequently the resulting paralogs subfunctionalized [114,147,167]. CenpE, Rod, Survivin, Sgo and the mitotic kinases Aurora and Plk also have elevated copy numbers. Possibly, these proteins also underwent (recurrent) duplication and subfunctionalization, as, for example, suggested for Sgo: In the lineages of *Schizosaccharomyces pombe*, *Arabidopsis thaliana* and mammals, Sgo duplicated and likely subsequently subfunctionalized in a recurrent manner [170–172].

Co-evolution within protein complexes of the kinetochore

Subunits of a single kinetochore complex tend to co-occur across genomes: They have similar patterns of presences and absences ("phylogenetic profiles", Figure 1a). Such co-occurring subunits likely co-evolved as a functional unit [173]. To quantify how similar phylogenetic profiles are, we calculated the Pearson correlation coefficient (r) for each kinetochore protein pair. We defined a threshold of r = 0.477 for protein pairs likely to be interacting, based on the scores among established interacting kinetochore pairs (S). All pairwise scores were used to cluster the proteins (Figure 1 including Sequence File 1 and Tree S1) and to visualize the proteins using *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE, Figure S7) [174]. Many established interacting proteins correlate well and, as a result, cluster together and are in close proximity in our *t*-SNE map. Examples include the SAC proteins Mad2 and MadBub, centromere proteins (CENPs) located in the inner kinetochore (discussed below), the Ska complex and the Dam1 complex. Such complexes, with subunits having highly similar phylogenetic profiles, evolved as a functional unit.

While some kinetochore proteins have highly similar phylogenetic profiles, others lack similarity, pointing to a more complex interplay between evolution and function. First, two proteins might have strongly dissimilar, or inverse, phylogenetic profiles, potentially because they are functional analogs [175]. In the kinetochore network, phylogenetic dissimilarity is observed for proteins of the Dam1 complex and of the Ska complex, which are indeed analogous complexes [84,85,87]. Second, proteins that do interact in a complex might nevertheless have little similarity in their phylogenetic profiles. Either such a complex

did not evolve as a functional unit because its subunits started to interact only recently [176], or because one of its subunits serves a non-kinetochore function and thus also co-evolves with non-kinetochore proteins [177]. An example of a potentially recently emerged interaction is BugZ-Bub3, that form a kinetochore complex in human [178,179], but have little similarity in their phylogenetic profiles, measured by their low correlation ($r = 0.187$). In general, BugZ's phylogenetic profile is different from other kinetochore proteins; hence, this protein might be recently added to the kinetochore [180,181]. An example of a kinetochore protein that co-evolves with non-kinetochore proteins is Zw10, which joins Rod and Zwilch in the RZZ complex. The phylogenetic profile of Zw10 is dissimilar from those of Rod and Zwilch ($r = 0.218$ for Rod, $r = 0.236$ for Zwilch), while those are very similar to each other ($r = 0.859$, Figure 3), due to Zw10 being present in various species that lack Rod and Zwilch. In those species, Zw10 might not localize to the kinetochore but perform only in vesicular trafficking, in a complex with Nag and Rint1 (NRZ complex [182]. Indeed, the Zw10 phylogenetic profile is much more similar to that of Nag ($r = 0.644$) and Rint1 ($r = 0.512$) compared to Rod and Zwilch. Hence, Zw10 more strongly co evolves with Nag and Rint1. The Rod and Zwilch phylogenetic profiles are similar to that of Spindly ($r = 0.730$ for Rod, $r = 0.804$ for Zwilch), a confirmed RZZ-interacting partner [183–185]. These similarities argue for an evolutionary "Rod–Zwilch–Spindly" (RZS) module, rather than an RZZ module.

The phylogenetic profiles of kinetochore proteins shed new light on these proteins' (co-)evolution and on their function, examples of which are discussed in detail below.

The CCAN evolved as an evolutionary unit that was lost in many lineages

The kinetochore connects the centromeric DNA, mainly via CenpA, to the spindle microtubules, mainly via Ndc80. In human and yeast, CenpA and Ndc80 are physically linked via the constitutive centromere-associated network (CCAN, reviewed in [186]). Physically, the CCAN comprises multiple protein complexes (Figure 2). Evolutionarily, however, it comprises a single unit, as the majority of CCAN proteins have highly similar phylogenetic profiles (Figure 1, average $r = 0.513$). Four CCAN proteins are very different from the others: CenpC, CenpR, CenpX, and CenpS. CenpC is widely present and is sufficient to assemble at least part of the outer kinetochore in *D. melanogaster* and humans [34,187]. CenpR seems a recent gene invention in animals. CenpX and CenpS have a more ubiquitous distribution compared to other CCAN proteins, possibly due to their non-kinetochore role in DNA damage repair [188,189].

Our study confirmed that most CCAN proteins have no (detectable) homologs

in *C. elegans and D. melanogaster*. The CCAN is not only absent from these model species, but also from many other lineages, such as various animals and fungi, and all Archaeplastida. Because the CCAN is found in three out of five eukaryotic supergroups, it likely was present in LECA, and subsequently lost multiple times in diverse eukaryotic lineages. Alternatively, the CCAN was invented more recently and horizontally transferred among eukaryotic supergroups. However, under both scenarios the CCAN was recently lost in various lineages, for example in the basidiomycete fungi: while *Ustilago maydis* has retained the CCAN, its sister clade *Cryptococcus neoformans* eliminated it (Figure 2d). The finding that most of the CCAN (with the exception of CenpC) is absent in many eukaryotic lineages poses questions about kinetochore architectures in these species. Since they generally possess a protein binding to the centromeric DNA (CenpA, see Figure S4 for details on identifying the orthologs of CenpA) and a protein binding to the spindle microtubules (Ndc80), their kinetochore is not wholly unconventional. Is the bridging function of the CCAN simply dispensable, as proposed for *D. melanogaster* [190], or is it carried out by other, non-homologous protein complexes? In order to answer these questions, the kinetochores of diverse species that lack the CCAN should be experimentally examined in more detail.

Absence of co-evolution between RZS and its putative receptor Zwint-1

Various studies suggested that the RZZ/RZS complex is recruited to the kinetochore primarily by Zwint-1. Zwint-1 itself localizes to the kinetochore by binding to Knl1 [63,191]. We compared the phylogenetic profile of Zwint-1 to the profiles of these interaction partners: RZZ/RZS and Knl1 (Figure 3). While we searched for orthologs of Zwint-1, we concluded that Zwint-1, Kre28 (*Saccharomyces cerevisiae*), and Sos7 (*Schizosaccharomyces pombe*) belong to the same orthologous group [192,193], collectively referred to as "Zwint-1". Although these sequences are only weakly similar, they can be linked by multidirectional homology searches (Figure S8).

Our set of 90 species contains many species that possess a Zwint-1 ortholog (36 species) but lack RZS, and vice versa (11 species, $-0.065 < r < 0$). This lack of correlation strongly suggests that, at least in a substantial amount of lineages, RZZ/RZS is not recruited to kinetochores by Zwint-1, but by another, yet unidentified factor. Support for this inference was recently presented in studies using human HeLa cells [64,194]. Compared to RZS, the phylogenetic profile of Zwint-1 is more similar to that of Knl1 (Figure 3, $r = 0.506$), and of Spc24 and Spc25 (Figure 1, $r = 0.529$ for Spc24, $r = 0.499$ for Spc25), two subunits of the Ndc80 complex that are located in close proximity to Knl1-Zwint-1[54]. Perhaps Zwint-1 stabilizes the largely unstructured protein Knl1 [64], thereby indirectly affecting the recruitment of RZZ/RZS.

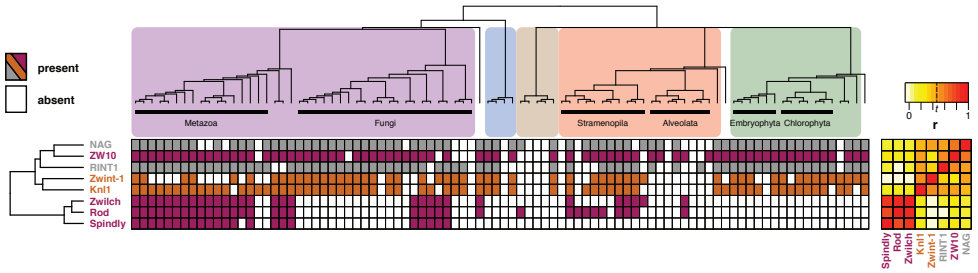


Figure 3 Phylogenetic profiles of the Rod–Zwilch–ZW10 (RZZ) complex, its mitotic interaction partners (Knl1, Zwint-1, and Spindly), and ZW10's interphase interaction partners in the NRZ (NAG and RINT1) complex. Presences and absences across eukaryotes of the RZZ subunits, Spindly, Zwint-1, and Knl1, and of the NRZ subunits, Nag and Rint1. Colored areas indicate eukaryotic supergroups as in Figure 1. **Right side:** Pairwise Pearson correlation coefficients (*r*) between the phylogenetic profiles including a heatmap. The indicated threshold *t* represents the value of *r* for which we found a sixfold enrichment of interacting protein pairs (see Figure S6). See also Figure S8 for the procedure by which homology between Zwint-1, Sos7, and Kre28 was detected.

Higher-order co-evolution between the AAA+ ATPase Trip13 and HORMA domain proteins

SAC activation and SAC silencing are both promoted by the AAA+ ATPase Trip13. Trip13 operates by using the HORMA domain protein p31^{comet} to structurally inactivate the SAC protein Mad2, also a HORMA domain protein (Figure 4a). Since the SAC requires Mad2 to continuously cycle between inactive and active conformations, Trip13 enables SAC signaling in prometaphase. In metaphase, however, when no new active Mad2 is generated, Trip13 stimulates SAC silencing [155,157,195]. The Trip13 ortholog of budding yeast, Pch2, probably has a molecularly similar function in meiosis: Pch2 is proposed to bind oligomers of the HORMA domain protein Hop1 (Hormad1 and Hormad2 in mammals, hereafter referred to as "HORMAD") and to structurally rearrange one copy within the oligomer, resulting in its redistribution along the chromosome axis. HORMAD, p31^{comet} and Mad2 are homologous as they belong to the family of HORMA domain proteins that also includes Rev7 [196] and autophagy-related proteins Atg13 and Atg101 [197,198]. All of these proteins likely descend from an ancient HORMA domain protein that duplicated before LECA.

Although the Trip13 phylogenetic profile is relatively similar to both the profiles of p31^{comet} (r = 0.526) and HORMAD (r = 0.517), Trip13 does not co-occur with these proteins in multiple species (Figure 4b). These exceptions to the co-occurrences of Trip13/p31 and Trip13/HORMAD can be explained by the dual role of Trip13, which is to interact with both p31^{comet} and with HORMAD. If we combine profiles of p31^{comet} and HORMAD, the similarity with

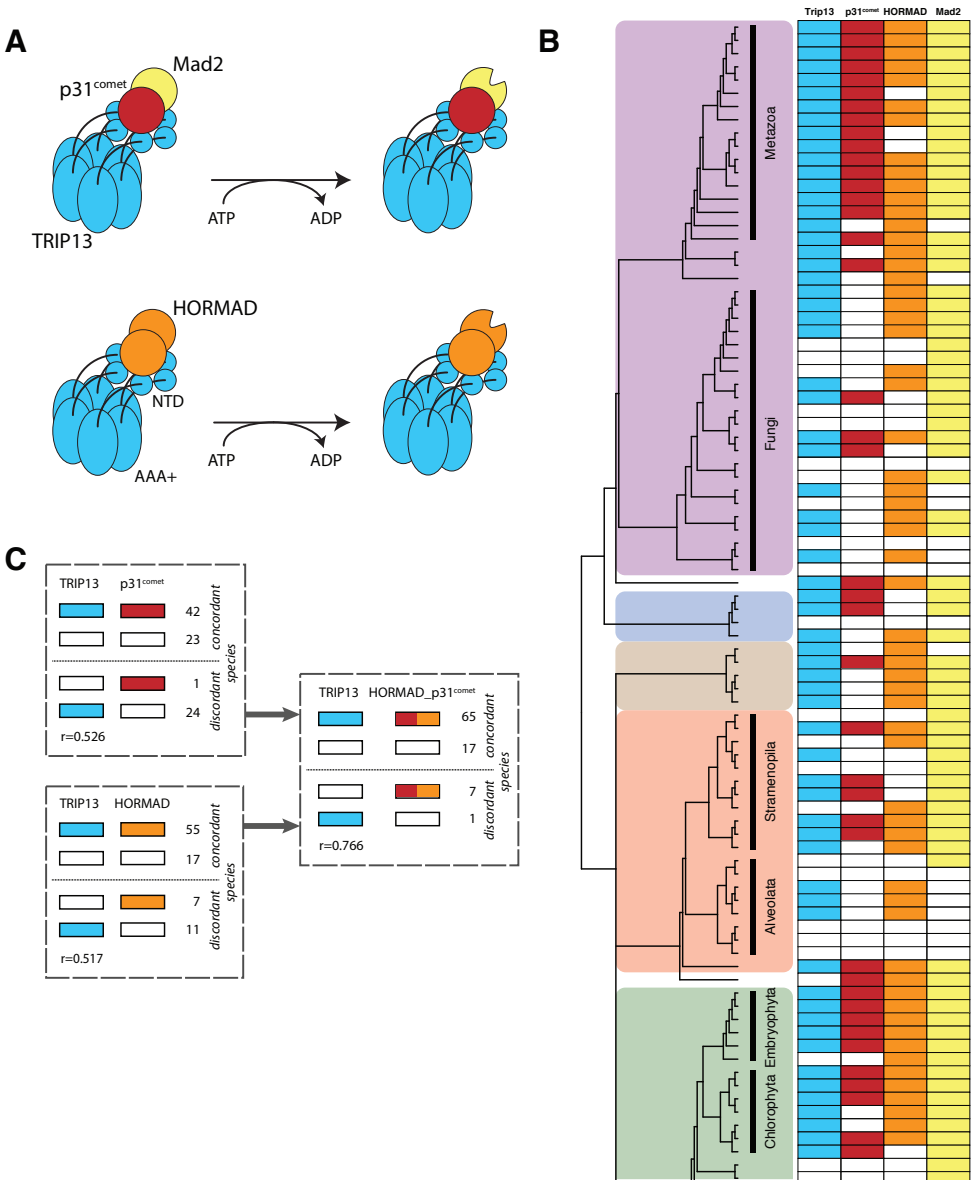


Figure 4 The co-evolutionary patterns of the multifunctional protein Trip13 (A) Model for the mode of action of Trip13 as recently suggested [222]. By hydrolyzing ATP, TRIP13 changes the conformation of HORMAD and Mad2 from closed to open, the latter via binding to co-factor p31comet, which forms a heterodimer with Mad2. TRIP13 has a C-terminal AAA+ ATPase domain (AAA+) and an N-terminal domain (NTD) and forms an hexamer [156]. **(B)** Presences and absences of Trip13 and of its interaction partners p31comet and HORMAD. Colored areas indicate eukaryotic supergroups as in Figure 1. **(C)** Numbers of lineages in which Trip13 is present or absent, compared to the presences of p31comet, HORMAD or their joint presences. Also the Pearson correlation coefficients of the phylogenetic profiles as in (B) are given.

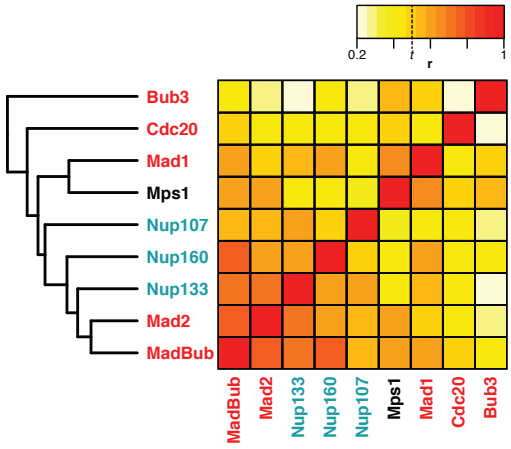


Figure 5 Correlations between proteins of the Nup107-160 complex and proteins of the SAC. Heatmap indicating the pairwise Pearson correlation coefficients (*r*) of the phylogenetic profiles of proteins of the Nup107-160 complex and of the SAC. The clustering (average linkage) on the left side of this heatmap was also based on these correlations. The indicated threshold *t* represents the Pearson correlation coefficient for which we found a six-fold enrichment of interacting protein pairs (see Figure S6).

Trip13 increases: the joint p31^{comet} and HORMAD profile strongly correlates with the Trip13 profile (r = 0.766, Figure 4c). Trip13 was indeed expected to co-evolve with both of its interaction partners, as has been demonstrated for other multifunctional proteins [177]. Based on the phylogenetic profiles, we conclude that Trip13 is only retained if at least p31^{comet} or HORMAD is present (with the exception of the diatom *Phaeodactylum tricornutum*). We predict that Trip13-containing species that lost p31^{comet} but retained HORMAD, such as *Saccharomyces cerevisiae and Acanthamoeba castellanii*, only use Trip13 during meiosis and not in mitosis.

The phylogenetic profiles of SAC proteins predict a role for nuclear pore proteins in the SAC

Because similar phylogenetic profiles reflect the functional interaction of proteins, similar phylogenetic profiles also predict such interactions. We applied this rationale by comparing the phylogenetic profiles of the kinetochore proteins (Figure 1) to those of proteins of the genome-wide PANTHER database in search of unidentified connections. PANTHER is a database of families of homologous proteins from complete genomes across the tree of life. We assigned all proteins present in our eukaryotic proteome database to these homologous families (see Materials and Methods). For each kinetochore protein in Figure 1, we listed the 30 best matching (with the highest Pearson correlation coefficient) families in PANTHER, and screened for PANTHER families that occur often in

these lists (Table S3). Of these families, we considered the nuclear pore protein Nup160 an interesting hit, because it is part of the Nup107-Nup160 nuclear pore complex that localizes to the kinetochore [199,200] The phylogenetic profile of Nup160 (as defined by PANTHER) was particularly similar to that of the SAC protein MadBub ($r = 0.718$). In order to improve the phylogenetic profile of Nup160, we manually determined the orthologous group of Nup160 in our own proteome dataset. We also determined those of Nup107 and Nup133, two other proteins of the Nup107-Nup160 complex. The Nup160, Nup133 and Nup107 phylogenetic profiles strongly correlated with those of SAC proteins MadBub ($0.541 < r < 0.738$) and Mad2 ($0.528 < r < 0.715$, Figure 5)even stronger than these three nuclear pore proteins correlated with one another ($0.475 < r < 0.601$). Furthermore, Nup160, Nup133 and Nup107 correlate better with MadBub and Mad2 than these SAC proteins do with the other SAC proteins (MadBub: average $r = 0.563$, Mad2: average $r = 0.511$) and far better than these SAC proteins do with all kinetochore proteins (MadBub: average r = 0.290, Mad2: average $r = 0.239$). While previous studies have shown that the Nup107-Nup160 complex localizes to the kinetochore in mitosis, our analysis in addition suggests that these proteins may function in the SAC and that they potentially interact with Mad2 and MadBub.

The Mad2-interacting motif (MIM) in Mad1 and Cdc20 is coupled with Mad2 presence

While interacting proteins are expected to co-evolve at the protein–protein level, as exemplified by many complexes within the kinetochore, interacting proteins might also co-evolve at different levels, such as protein-motif. Co-evolution between a protein and a sequence motif has been incidentally detected before, for example in case of CenpA and its interacting motif in CenpC [24] and in case of MOT1 and four critical phenylalanines in TBP [201]. We here explore potential co-evolution of Mad2 with the sequence motif it interacts with in Cdc20 and Mad1: the Mad2-interacting motif (MIM). Both the Mad2–Mad1 and the Mad2–Cdc20 interactions operate in the SAC [135,139]. We defined the phylogenetic profiles of the MIM in Mad1 and Cdc20 [202,203] (Figure 6a) by inspecting the multiple sequence alignments of Mad1 and Cdc20. These alignments revealed that the MIM is found at a similar position across the Mad1 and Cdc20 orthologs; hence, the motif likely predates LECA in both these proteins. Notable differences exist between the MIMs of Cdc20 and Mad1, which could reflect differences in binding strength to Mad2.

The phylogenetic profiles of Mad2 and of the MIM in Cdc20 or Mad1 orthologs correlated stronger than the full-length proteins (Figure 6b,c). In particular, species lacking Mad2, but having Mad1 and/or Cdc20, never contained the canonical MIM in either their Cdc20 or their Mad1 sequences (hypergeometric

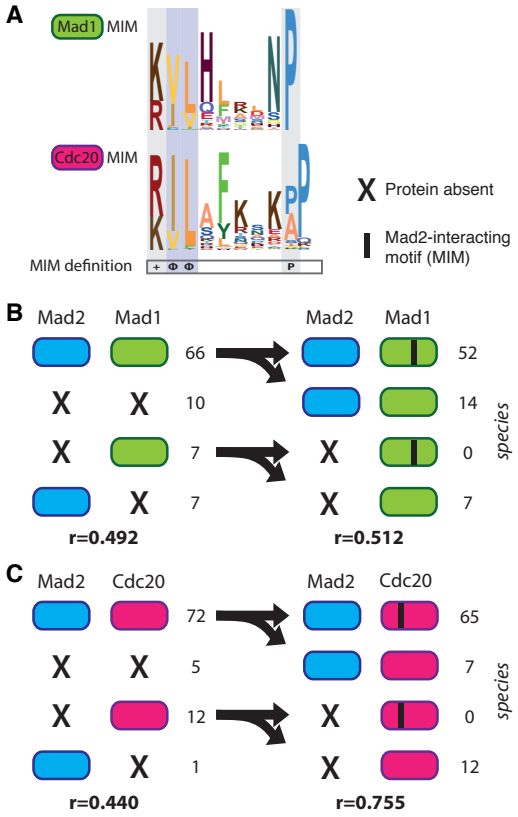


Figure 6 Phylogenetic co-occurrence of Mad2 with its interaction partners Mad1 and Cdc20 and their Mad2-interacting motifs (MIMs). (A) The sequence logos of the MIMs of Mad1 (upper panel) and Cdc20 (lower panel) based on the multiple sequence alignments of the motifs. Below is indicated the required amino acid sequence of the MIM (+: positive residue, Ω: hydrophobic residue, P: proline) which is restricted by the pattern [RK][ILV]{2}.{3,7}P. **(B,C)** Left side: Numbers of presences and absences of Mad2 in 90 eukaryotic species and its interaction partners Mad1 (B) and Cdc20 (C). Right side: Frequencies of Mad2 and canonical MIM occurrences in species having Mad1 (B) or Cdc20 (C), respectively. Also the Pearson correlation coefficients (*r*) for the corresponding phylogenetic profiles are shown.

test: $P < 10^{-4}$, $P < 10^{-9}$ for Mad1 and Cdc20, respectively). Such species hence likely lost Mad2 and subsequently lost the MIM in Mad1 and Cdc20, because it was nod longer functional. Moreover, absence of the MIM in Mad1/Cdc20 supports that in these species Mad2 is indeed absent. While we expected to only find a MIM in species that actually have Mad2, we also expected the reverse: that species that have Mad2 also have a MIM in their Mad1/Cdc20. This is however not the case, most notably for Mad1: Many lineages (14) have both Mad1 and Mad2 but lack the MIM in Mad1. A substantial fraction (6) of this group belongs to the land plant species that have a somewhat different motif in Mad1 that is conserved within this lineage (Figure S5a). This altered land plant motif might mediate the Mad1–Mad2 interaction, which has been reported in *A. thaliana* [204]. If we consider this plant motif to be a "valid" MIM, the Mad1-MIM and Mad2 correlate substantially better (Figure S5b-d). Overall, under both motif definitions the protein-motif correlations are higher than the protein–protein correlations. Hence, including sequence motifs can expose that interaction partners co-evolve, albeit at a different level, and may aid to predict functional interactions between proteins de novo.

Discussion

Our evolutionary analyses revealed that since LECA, the kinetochores of different lineages strongly diverged by different modes of genome evolution: kinetochore proteins were lost, duplicated and/or invented, or diversified on the sequence level. In addition to straightforward protein–protein co-evolution, we found alternative evolutionary relationships between proteins that hint at a more complex interplay between evolution and function. Some established interacting proteins have not co-evolved (Zwint-1 and RZS, Bub3 and BugZ), which has been previously shown for other interaction partners to reflect evolutionary flexibility [176]. Lack of co-evolution may also reflect that a protein has multiple different functions, for which it interacts with different partners. The phylogenetic profile of such a multifunctional protein differs from either of its interaction partners, and instead is similar to the combined profiles of its interaction partners [177], as we showed for Trip13 with HORMAD and p31^{comet}. Some co-evolutionary relationships predicted novel protein functions, such as nuclear pore proteins operating in the SAC, which should be confirmed with experiments. Finally, not only proteins, but also functional sequence motifs co-evolved with their interaction partner, as we found for the MIMs in Cdc20/Mad1 with Mad2. Probably, including more proteins and (known and de novo predicted) motifs/domains will not only improve the correlation between known interaction partners, but will also enhance predicting yet unknown interactions and functions.

While we carefully curated the orthologous groups of each of the kinetochore proteins, their phylogenetic profiles might contain some false positives and/or false negatives: incorrectly assigned presences (because a protein sequence in fact is not a real ortholog) and incorrectly assigned absences (because a species does contain an ortholog, but we did not detect it). For most kinetochore proteins, we estimate the chance of false negatives larger than of false positives, mainly because they likely are vulnerable to homology detection failure, given that their sequences evolve so rapidly (Table S1, Results). Such false negatives of a particular protein will result in falsely inferred gene loss events. A failure to detect homology might therefore also cause sequence divergence to correlate with loss frequency (Figure S2). Specific examples of suspicious absences (potential false negatives) include the inner centromere protein Borealin in *Saccharomyces cerevisiae* and the KMN network proteins Spc24, Spc25, Nsl1/Dsn1 in *Drosophila melanogaster* and *Caenorhabditis elegans*, and possibly Ndc80 in *Trypanosoma brucei*, since functional counterparts of these proteins have been characterized in these species [50,166,205–209]. Moreover, species that we predicted to have very limited kinetochore compositions, such as *Tetrahymena thermophila* (Figure 2b), might actually contain highly divergent orthologs that we could not detect. If such a species' kinetochore would be ex-

amined biochemically, its undetected orthologs might be uncovered. Although the phylogenetic profiles of the kinetochore proteins presented here might contain some of such errors, we think that our manual curation of the orthologs groups (see Materials and Methods) yields an accurate global representation of the presences and absences of these proteins among eukaryotes. We think this accuracy is supported by the high similarity of phylogenetic profiles of interacting proteins.

The set of kinetochore proteins we studied here is strongly biased toward yeast and animal lineages, lineages that are relatively closely related on the eukaryotic tree of life. This bias is due to the extensive experimental data available for these lineages. Highly different kinetochores might exist, such as the kinetochore of *Trypanosoma brucei* [165,166]. If in the future we know the experimentally validated kinetochore compositions of a wider range of eukaryotic species, we could sketch a more complete picture of kinetochore evolution and could potentially expand and improve our functional predictions.

Since the kinetochore seems highly diverse across species, several questions arise. Is the kinetochore less conserved than other core eukaryotic cellular systems/pathways, as comparing it to the APC/C suggested? And if so, why is it allowed to be less conserved, or are many of the alterations adaptive to the species? Why do certain lineages (such as multicellular animals and plants) contain a particular kinetochore submodule (such as the Ska complex) while others lack it, or have an alternative system (such as Dam1)? Do these genetic variations among species have functional consequences for kinetochore-related processes in their cells? To answer such questions, our dataset should be expanded with specific (cellular) features and lifestyles, when this information becomes available for the species in our genome dataset. Together with biological and biochemical analyses of processes in unexplored species, an expanded dataset may reveal the true flexibility of the kinetochore in eukaryotes and show how chromosome segregation is executed in diverse species. The comparative genomics analysis that we presented here provides a starting point for such an integrated approach into studying kinetochore diversity and evolution, since it allows for informed decisions about which species to study.

Materials and Methods

Constructing the proteome database

To study the occurrences of kinetochore genes across the eukaryotic tree of life, we constructed a database containing the protein sequences of 90 eukaryotic species. This size was chosen because we consider it to be sufficiently large to represent eukaryotic diversity, but also sufficiently small to allow for manual

detection of orthologous genes. We selected the species for this database based on four criteria. First, the species should have a unique position in the eukaryotic tree of life, in order to obtain a diverse set of species. Second, if available we selected two species per clade, which facilitates the detection of homologous sequences and the construction of gene phylogenies. Third, widely used model species were preferred over other species. Fourth, if multiple proteomes and/or proteomes of different strains of a species were available, the most complete one was selected. Completeness was measured as the percentage of core KOGs (248 core eukaryotic orthologous groups [210]) found in that proteome. If multiple splice variants of a gene were annotated, the longest protein was chosen. A unique protein identifier was assigned to each protein, consisting of four letters and six numbers. The letters combine the first letter of the genus name with the first three letters of the species name. The versions and sources of the selected proteomes can be found in Table S2.

Ortholog detection

The set of kinetochore proteins we studied were selected based on three criteria: (i) localizing to the kinetochore, (ii) being present in at least three lineages, and (iii) having an established role, supported by multiple studies, in the kinetochores and/or kinetochore signaling in human or in budding yeast. We applied a procedure comprising two different methods to find orthologs for the kinetochore proteins in our set within our database of 90 eukaryotic proteomes, and the same procedure was followed for determining orthologs of the APC/C proteins, Nag, Rint1, Nup107, Nup133, Nup160, and HORMAD. The method of choice depended on whether or not it was straightforward to find homologs across different lineages for a specific protein. In both methods, initial searches started with the human sequence, or, if the protein is not present in humans, with the budding yeast sequence.

Method 1. If many homologs were easily found, the challenge was to distinguish orthologs from outparalogs. Here, we defined an orthologous group as comprised of proteins that result from speciation events and that can be traced back to a single gene in LECA, whereas outparalogs are related proteins that resulted from a pre-LECA duplication. For example, Cdc20 and Cdh1 are homologous proteins, both having their own orthologous groups among the eukaryotes. They resulted from a duplication before LECA; therefore, members of the Cdc20 and Cdh1 group are outparalogs to each other. To find homologs, we used blastp online to search through the non-redundant protein sequences (nr) as a database [211]. We aligned the sequences found with MAFFT [212] (version v7.149b, option einsi, or linsi in case of expected different architectures) to make a profile HMM (www.hmmer.org, version HMMER 3.1b1). If the homologs are known to share only a certain domain, that domain was used for

the HMM; otherwise, we used the full-length alignment. This HMM was used as input for hmmsearch to detect homologs across our own database of 90 eukaryotic proteomes. From the hits in this database, we took a substantial number of the highest scoring hit sequences, up to several hundreds. We aligned the hit sequences using MAFFT and trimmed the alignment with trimAl [213] (version 1.2, option automated1). Subsequently, RAxML version 8.0.20 [214] was used to build a gene tree (settings: varying substitution matrices, GAMMA model of rate heterogeneity, rapid bootstrap analysis of 100 replicates). We interpreted the resulting gene tree by comparing it to the species tree and thereby determined which clusters form orthologous groups. These orthologous groups were identified by finding the cluster that contained sequences from a broad range of eukaryotic species and had a sister cluster that also has sequences from this broad range of species. The cluster that contained the initial human query sequence was the orthologous group of interest, while the sister cluster is the group of outparalogs. In our search of orthologs of CenpA, we applied this first method. CenpA is part of the large family of histone H3 proteins and has long been recognized to diverge rapidly, due to which it is a challenge to reconstruct CenpA's evolution [215]. We determined this orthologous cluster with help of experimentally identified centromeric histone H3 variants in a wide range of species, and we included two Toxoplasma gondii sequences that were not part of this cluster. For details, see Figure S4 and Tree S1. The tree in this figure was visualized using iTOL [216] **Method 2.** If homologs were not easily found, no outparalogs were obtained by these searches and hence the homologs defined the orthologous group. For these cases, we used a different strategy to find the orthologous group in our database. Iterative searching methods (jackhmmer and/or psi-blast) were applied to find homologs across the nr and UniProt database [217]. In specific cases, we cut the initial query sequence, for example to remove putative coiled-coil regions. If a protein returned very few hits, we tried to expand the set of putative homologous sequences by using some of the initially obtained hits as a query. If candidate orthologous proteins were reported in experimental studies in species other than human or budding yeast, but not found by initial searches, we specifically searched using those as a query. If this search yielded hits overlapping with previous searches, these candidate orthologous sequences were added to the set of hits. The sequences in this set were aligned to obtain a refined profile HMM. In addition, we searched for conserved motifs in the hit sequences using MEME [218] (version 4.9.0), which aided in recognizing conserved positions that could characterize the homologs. The obtained profile HMM was used to search for homologs across in local database. The resulting hits were checked for the motifs identified by MEME and applied to online (iterative) homology searches to check whether we retrieved sequences already identified as orthologous. Based on this evaluation of individual hits, we defined a scoring threshold for the hmmsearch with this profile HMM and searched our database until no new hits were found. The

resulting set of sequences was the orthologous group of interest. The sequences of the orthologous groups can be found in the Sequence File S1.

Calculating correlations between phylogenetic profiles

In order to study the co-evolution of the kinetochore proteins and to infer potential functional relationships of these genes based on co-evolution, we derived the phylogenetic profiles of these genes. The phylogenetic profile of a gene is a list of its presences and absences across our set of 90 eukaryotic genomes based on the composition of the orthologous groups. The phylogenetic profile consists of a string of 90 characters containing a "1" if the gene is present in a particular species (either single- or multicopy), and a "0" if it is absent. To reveal whether two genes often co-occur in species, we measured how similar their phylogenetic profiles were using the Pearson correlation coefficient [219]. All pairwise scores can be found in Matrix S1. To identify pairs of proteins that potentially have a functional association, we applied a threshold of $r = 0.477$. Figure S6 clarifies why the Pearson correlation coefficient was opted for and how the threshold was set. The Pearson correlation coefficients of all gene pairs were converted into distances ($d = 1 - r$), and the genes were clustered based on their phylogenetic profiles using average linkage. The Pearson correlation coefficients were also used to map the kinetochore proteins in 2D by Barnes-Hut *t*-SNE (Figure S7) [174].

Detecting the MIM in Mad1 and Cdc20 orthologs

We made multiple sequence alignments of the Cdc20 and Mad1 orthologous groups using MAFFT (option einsi). We used these alignments to search for the Mad2-interacting motif (MIM). The typical MIM is defined by [KR][IVL]{2}.{3,7}P for both Mad2 and Cdc20 [202,203], but we also used an alternative definition: [ILV]{2}.{3,7}P or [RK][ILV].{2}. We inferred that the location of the motif in the protein is conserved in Mad2 as well as in Cdc20, because the position of the MIM in the multiple sequence alignments was the same in highly divergent species (e.g., plants and animals). For all orthologous sequences, we checked whether the motif, either the typical MIM (Figure 6) or the alternative MIM (Figure S5) was present on these conserved positions.

Finding novel proteins functioning in the kinetochore

To find new proteins performing essential roles at the kinetochore by phylogenetic profiling, a reference protein set was needed. This reference set was based on the protein families present in PANTHER. More specifically, we assigned the proteins within our proteome database of 90 eukaryotic species to PANTHER (sub)families [220] (version 10). This assignment was done by applying hmmscan to the protein sequences of our database, using the complete set of PANTHER family and subfamily HMMs as a search database. Each protein was

assigned to the PANTHER (sub)family to which it had the highest hit, if significant. If a protein was assigned to a subfamily, it was also assigned to the full family to which that subfamily belongs. For each PANTHER (sub)family, a phylogenetic profile was constructed and compared to the phylogenetic profiles of the kinetochore proteins. For each kinetochore protein, the best 30 matches of PANTHER (sub)families were selected. The PANTHER protein families often occurring in these top lists can be found in Table S3.

Comparing diversity of kinetochore and APC/C proteins

For the kinetochore and APC/C proteins in this dataset, we calculated their occurrence frequencies and entropies across 90 eukaryotic species. The entropy reflects a protein's diversity of presences and absences across species: a protein that is present in half of the species has the highest entropy. We also calculated and compared all pairwise Pearson correlation coefficients of the phylogenetic profiles for both of these protein datasets. To assess how complete the kinetochores and APC/C complexes of the species in our dataset are, we calculated the percentage of present kinetochore proteins in species having Ndc80 and CenpA (because those species are expected to have a kinetochore), and we calculated the percentage of present APC/C proteins in species having the main APC/C enzyme Apc10. Loss frequencies were inferred from Dollo parsimony for all kinetochore and APC/C proteins inferred to have been present in LECA. Transitions (also a measure for the evolutionary dynamics of proteins) were measured for each protein by counting all changes in state (so from present to absent, or from absent to present) along a phylogenetic profile. Since the ordering of the species in the phylogenetic profile is an indication of their relatedness, these transitions are expected to reflect the evolutionary flexibility of proteins as well. dN/dS and percent identity scores for human and mouse sequences were derived from Ensembl [221] (downloaded via Ensembl BioMart on November 24, 2016). If multiple one-to-one orthologs for a single orthologous group/family exist, the average dN/dS or percent identity was taken. The results of these kinetochore-APC/C comparisons can be found in Table S1.

Contributions

BS and GJPLK designed the research. JJEH and ET performed the research. LMW contributed the eukaryotic genome database. JJEH, BS, and GJPLK analyzed the data and wrote the manuscript.

Acknowledgements

We thank the members of the Kops and Snel labs for critical reading and helpful discussion on the manuscript. We thank John van Dam for his contribution

to compiling the eukaryotic proteome database. This work was supported by the UMC Utrecht and the Netherlands Organisation for Scientific Research (NWO-Vici 865.12.004 to GK).

Electronic Supplementary Material

Supplementary materials for this chapter are made available online at the following link http://bioinformatics.bio.uu.nl/eelco/thesis/

Sequence File S1 contains separate fasta files with the orthologs of all proteins used in this study.

Matrix S1 contains pairwise correlation scores of the phylogenetic profiles of all proteins used in this study.

Tree S1 contains a newick file of the phylogenetic analysis of Histone 3-like genes of Figure S4.

Figure S1 shows the phylogenetic profiles for subunits of the anaphase-promoting complex/cyclosome (APC/C) across 90 eukaryotic lineages.

Figure S2 shows plots of the loss frequencies (eukaryote-wide) and sequence evolution (human-mouse) of kinetochore and APC/C proteins.

Figure S3 visualizes the copy numbers of kinetochore proteins per species.

Figure S4 shows the gene phylogeny of histone H3 homologs; CENPA is a single orthologous group in LECA.

Figure S5 shows Mad2-interacting motif evolution in green plants and a similar analysis as in Figure 5 under a less strict Mad2-interacting motif definition.

Figure S6 shows the performance of various measures that compare phylogenetic profiles in predicting physically interacting proteins within the kinetochore network.

Figure S7 contains a *t*-Distributed Stochastic Neighbor Embedding plot (*t*-SNE) based on their pairwise distances measured by the Pearson correlation coefficient of the phylogenetic profiles.

Figure S8 shows the procedures and sequences that were used to determine that Zwint-1, Sos7 and Kre28 belong to the same orthologous protein family in LECA.

Table S1 shows a comparison of different measures for protein diversity of the set of kinetochore and APC/C proteins.

Table S2 contains the sources of the proteomes that are used in this study.

Table S3 shows similarity of phylogenetic profiles of kinetochore proteins with PANTHER (sub)families.

CHAPTER 3

Kinetochore evolution deconstructed: mapping motifs and domains

Eelco Tromer, Carlos Sacristan, Jolien JE van Hooff
Berend Snel & Geert JPL Kops

Manuscript in preparation

Abstract

The kinetochore is an advanced molecular apparatus built on top of centromeric DNA that functions at the heart of chromosome segregation during mitosis and meiosis. A complement of ~100 proteins at 1-to many stoichiometry forms and regulates a landing platform for the bioriented attachment of sister chromatids to microtubules of the spindle. We recently determined the orthologs of 70 kinetochore network components in 90 phylogenetically diverse eukaryotes and found that kinetochore compositions diverge extensively through various modes of gene evolution such as loss, duplication, invention and even displacement. To shed light on these evolutionary dynamics we have started to map conserved protein features to orthologs using a de novo sequence motif discovery pipeline, and initiated co-evolution analysis of such motifs. Using our approach we dissect the (co-)evolution of the Mis12 complex and its reported interacting partners CenpC, CenpU and other members of the KMN network. We find that Csm1, a member of the Monopolin Complex in budding yeast, is a eukaryote-wide kinetochore protein that is specifically lost in metazoan lineages. Strikingly, the phylogenetic profile of Csm1 strongly correlates with that of a highly charged motif in the disordered N-terminus of Dsn1, indicating both their co-evolution and physical interaction. In addition, we show that two conserved motifs in the N-terminal coiled-coil of Spindly are involved in the recruitment of the dynein-dynactin complex to kinetochores. Finally, we uncover the evolutionary origin of Zwint-like orthologs and show the striking recurrent loss of ancestral RWD domains in Zwint-1, Kre28 and Sos7.

Introduction

Kinetochores are at the heart of the propagation of genetic information across generations. In essence, they link the DNA to the cell division machinery. As such, they consist of DNA-binding and microtubule-binding proteins, supplemented with proteins involved in linking those functions, and proteins involved in signal transduction, chromosome movement and microtubule dynamics. In opisthokhont species, kinetochores consist of 50-100 proteins, making it one of the largest known protein complexes [18,162].

A 16-subunit assembly known as the Constitutive Centromere-associated Network (CCAN) constitutes the centromeric chromatin environment and provides a docking platform for the microtubule-binding complexes of the Knl1-C/Mis12-C/Ndc80-C (KMN) network. This core complex forms the basis for the recruitment of various auxiliary modules that monitor and modulate kinetochore-microtubule attachments e.g. proteins of the Spindle Assembly Checkpoint (SAC), regulators of the dynein motor complex (Rod-Zwilch-Zw10 (RZZ)-Spindly), microtubule plus-end tracking complexes (Dam1 and SKA) and the meiosis-specific Monopolin complex in budding yeast [18,162].

We recently performed detailed analysis of the conservation of kinetochore proteins across eukaryotic species. Although the core functions (DNA- and microtubule-binding) are present in most species, we found a high degree of putative divergent kinetochore architectures. This suggests that kinetochore composition may be evolutionary flexible without loss of kinetochore function [52]. Our comparative genomics analyses provided a wealth of sequence information on kinetochore protein orthologs. We reasoned that detailed comparisons of orthologous sequences is likely to reveal conserved functional sequences and may inform on potential functional interactions between proteins and sequence motifs by virtue of a high degree of co-evolution. Similar analyses are rare in literature, but motif-protein co-evolution has been shown to be successful in showing, for example, co-evolution between CenpA and its interacting motif in CenpC [24] and between the transcription factor MOT1 and four critical phenylalanines in TBP [201].

We here describe a pipeline for *de novo* discovery of conserved protein features in sets of orthologous sequences. We use this pipeline to interrogate conserved regions in the Mis12 complex, in the dynein-dynactin adaptor Spindly and the Knl1 complex member Zwint-1. Co-evolutionary analyses and limited experimental analyses inform on potential functions of the uncovered motifs and domains.

Results and Discussion

Supplementary gene prediction to improve phylogenetic profile completeness and orthologous sequence quality

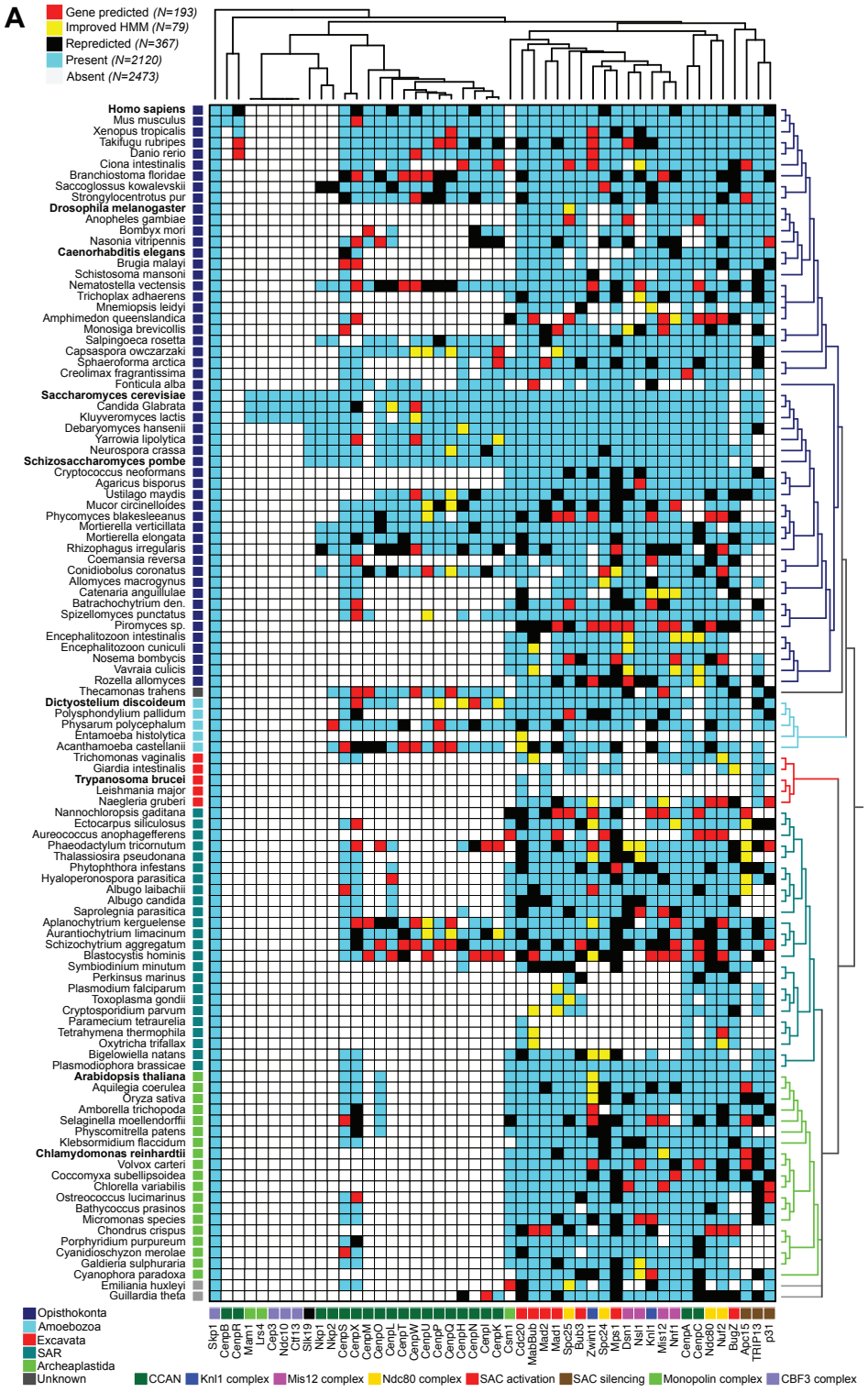
For the successful detection of conserved features and inference of correct co-evolutionary scenarios, both the quality of orthologous sequences and the completeness of phylogenetic profiles is key. Since we observed conspicuous absences of some kinetochore subunits in various species [52] and found problems with sequence quality to be widespread amongst eukaryotic genomes [114,133] we wondered to what extend faulty gene prediction might be a factor. We therefore established phylogenetic profiles for 48 kinetochore proteins in 109 phylogenetically diverse eukaryotic proteomes (largely based on our previous analyses [3]) and manually assessed the quality of gene prediction by meticulously checking the validity of all absences and incomplete sequences of orthologous proteins (see Material and Methods, Sequence File S1, Table S1). We were able to detect 193 orthologs that were previously called absent using six-frame-translated genomes (Figure 1a, category 'Gene prediction') and another 79 using improved Hidden Markov Models (HMM, Figure 1a, category 'Improved HMM') based on inclusion of the 193 new orthologs and the addition of phylogenetically informative genomes (see Methods for more details). Based on multiple sequence alignments of orthologous protein families we flagged incomplete genes and re-predicted them in 367 cases (Figure 1a, category 'Repredicted'). All in all our detailed survey resulted in a 10,9% increase of called presences (272 out of 2487 total present) and indicate the need for gene reprediction in 14,8% (367 out of 2487 total present) of all genes.

Of note, we found various patterns in this 'false negative' set of orthologs that may be of general interest (see for number of gene prediction issues per gene and genome Table S1). First, we find that short genes (aa<100) are often found to be absent (e.g. CenpX (N=16), CenpW (N=12)), likely indicating too strict gene length cut-offs of gene prediction algorithms used to annotate a number of the genomes. Second, orthologous protein families that contain multiple domains and/or multiple strongly conserved motifs (e.g. Mps1 (N=40) and Knl1 (N=16)) frequently require gene reprediction, mostly because of incorrectly called gene fissions. Last, a number of genomes (e.g. *Blastocystis hominis* (N=20), *Branchiostoma floridae* (N=16) and *Selaginella moellendorffii* (N=13)) reveal consistent problems with correct gene prediction, which strongly advocates strict quality criteria for the inclusion of genomes in phylogenetic profiling analyses. Further analyses are needed to assess whether supplemental gene prediction can have a profound effect on the inference of protein complex co-evolution and de novo motif discovery.

Co-evolution of kinetochore protein complexes

We proceeded with the manually corrected orthologous protein families of 48 kinetochore complex members (see for subunit topology Figure 1b). To assess the evolutionary dynamics of these kinetochore proteins, we adopted our previously used approach [52] and clustered the Pearson correlation scores of the phylogenetic profiles of each protein pair (Figure 1a, Matrix S1). For a more intuitive interpretation of the pairwise correlation scores we visualized them in two dimensions using a Barnes-Hut implementation of *t*-distributed Stochastic Neighbor Embedding (*t*-SNE, Figure 1b) [174]. On average, the pairwise correlation scores are similar to our former analysis (table S2) and the *t*-SNE map indicates strong co-evolution for CCAN subcomplexes (CenpAC, SX, OPQUTW, HIKMLN), KMN network members and SAC subunits, proteins involved in SAC silencing (Apc15, p31^{comet} and Trip13) and the budding yeast-specific centromere CBF3 complex (see clusters in Figure 2b).

We also performed phylogenetic profiling for seven reported lineage-specific kinetochore proteins that we did not include in our previous analysis: the transposon-derived protein CenpB (mammals) [51], three members of the core Monopolin complex (Csm1, Lrs4 and Mam1 in fungi) [57], the microtubule plus end-binding protein Slk19 (ascomycetes) [223] and two non-essential CCAN-related proteins Nkp1/Fta4 and Nkp2 (ascomycetes) [43,44]. While the narrow phylogenetic profiles of CenpB, Slk19, Lrs4 and Mam1 indeed indicate their lineage-specific roles in the kinetochore (Figure 1a,b), we found more extended presence-absence profiles for Nkp1, Nkp2 and Csm1. This means that these genes are not as lineage-specific as previously reported. In addition, the phylogenetic profiles of Nkp1 and Nkp2 show high similarity (r = 0,848) and these proteins were only found in lineages that harbor the CCAN complex (e.g. *Salpingoeca rosetta* and *Rhizophagus irregularis*), confirming their previously described association with the Ctf19 complex in budding yeast and the Mal2-Sim4 complex in fission yeast [43–45]. Strikingly however, we could also detect orthologs in animals (Nkp1-2: *Saccoglossus kowalevski* and *Nematostella vectensis*) and amoebozoa (Nkp2: *Acanthamoeba castellanii* and *Physarum polycephalum*), suggesting that Nkp1 and Nkp2 may have been part of the kinetochore of the last common ancestor of opisthokonts and were subsequently lost in many lineages. The phylogenetic profiles of subunits of the monopolin complex (Csm1, Lrs4 and Mam1), which holds together homologous kinetochores during meiosis I [224], show no similarity. In fact we here report the novel discovery of orthologs of the RWD domain-containing protein Csm1 in all eukaryotic supergroups, indicating that it was a likely subunit of the kinetochore in LECA and that its function in the Monopolin complex is possibly derived. See Box1 for discussion of the implications of this finding.


A
Gene predicted (N=193)
Improved HMM (N=79)
Repredicted (N=367)
Present (N=2120)
Absent (N=2473)
Homo sapiens
Mus musculus
Xenopus tropicalis
Takifugu rubripes
Danio rerio
Ciona intestinalis
Branchiostoma floridae
Saccoglossus kowalevskii
Strongylocentrotus pur
Drosophila melanogaster
Anopheles gambiae
Bombyx mori
Nasonia vitripennis
Caenorhabditis elegans
Brugia malayi
Schistosoma mansoni
Nematostella vectensis
Trichoplax adhaerens
Mnemiopsis leidyi
Amphimedon queenslandica
Monosiga brevicollis
Salpingoeca rosetta
Capsaspora owczarzaki
Sphaeroforma arctica
Creolimax fragrantissima
Fonticula alba
Saccharomyces cerevisiae
Candida Glabrata
Kluyveromyces lactis
Debaryomyces hansenii
Yarrowia lipolytica
Neurospora crassa
Schizosaccharomyces pombe
Cryptococcus neoformans
Agaricus bisporus
Ustilago maydis
Mucor circinelloides
Phycomyces blakesleeanus
Mortierella verticillata
Mortierella elongata
Rhizophagus irregularis
Coemansia reversa
Conidiobolus coronatus
Allomyces macrogynus
Catenaria anguillulae
Batrachochytrium den.
Spizellomyces punctatus
Piromyces sp.
Encephalitozoon intestinalis
Encephalitozoon cuniculi
Nosema bombycis
Vavraia culicis
Rozella allomyces
Thecamonas trahens
Dictyostelium discoideum
Polysphondylium pallidum
Physarum polycephalum
Entamoeba histolytica
Acanthamoeba castellanii
Trichomonas vaginalis
Giardia intestinalis
Trypanosoma brucei
Leishmania major
Naegleria gruberi
Nannochloropsis gaditana
Ectocarpus siliculosus
Aureococcus anophagefferens
Phaeodactylum tricornutum
Thalassiosira pseudonana
Phytophthora infestans
Hyaloperonospora parasitica
Albugo laibachii
Albugo candida
Saprolegnia parasitica
Aplanochytrium kerguelense
Aurantiochytrium limacinum
Schizochytrium aggregatum
Blastocystis hominis
Symbiodinium minutum
Perkinsus marinus
Plasmodium falciparum
Toxoplasma gondii
Cryptosporidium parvum
Paramecium tetraurelia
Tetrahymena thermophila
Oxytricha trifallax
Bigelowiella natans
Plasmodiophora brassicae
Arabidopsis thaliana
Aquilegia coerulea
Oryza sativa
Amborella trichopoda
Selaginella moellendorffii
Physcomitrella patens
Klebsormidium flaccidum
Chlamydomonas reinhardtii
Volvox carteri
Coccomyxa subellipsoidea
Chlorella variabilis
Ostreococcus lucimarinus
Bathycoccus prasinos
Micromonas species
Chondrus crispus
Porphyridium purpureum
Cyanidioschyzon merolae
Galdieria sulphuraria
Cyanophora paradoxa
Emiliania huxleyi
Guillardia theta
Shp1
CenpB
CenpR
Mam1
Lrs4
Cep3
Ndc10
Ctf13
Skp1
Nkp1
Nkp2
CenpS
CenpX
CenpM
CenpO
CenpL
CenpT
CenpW
CenpU
CenpP
CenpQ
CenpH
CenpI
CenpK
Csm1
Cdc20
MadBub
Mad2
Mad1
Spc25
Bub3
Zwint1
Spc24
Mps1
Dsn1
Nsl1
Knl1
Mis12
Nnf1
CenpA
CenpC
Ndc80
Nuf2
Bug2
Apc15
TRIP13
p31
Opisthokonta
Amoebozoa
Excavata
SAR
Archeaplastida
Unknown
CCAN
Knl1 complex
Mis12 complex
Ndc80 complex
SAC activation
SAC silencing
Monopolin complex
CBF3 complex

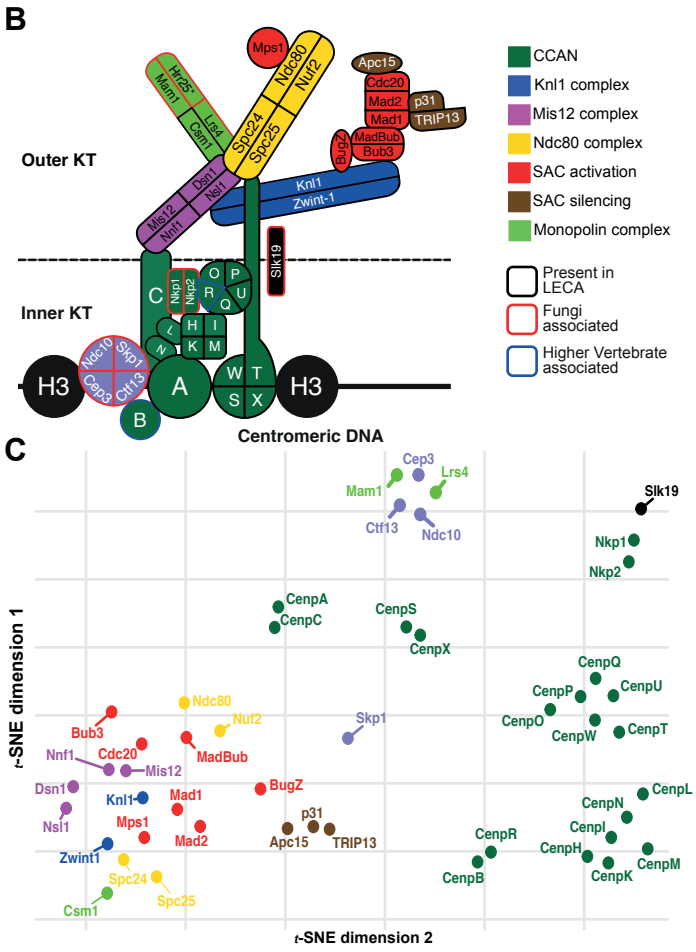


Figure 1 Improving phylogenetic profile completeness and orthologous sequence quality for 48 kinetochore proteins using ad hoc gene prediction. (A) Presence -absence profiles of 48 kinetochore proteins in 109 eukaryotic species. **Right side:** phylogenetic tree of the species in our genome set, with colored branches indicating different supergroups. **Top side:** kinetochore proteins are clustered by average linkage based on the pairwise Pearson correlation coefficients of the phylogenetic profiles. **Bottom side:** Annotation of to which protein complex they belong. **Left side:** species name sand additional coloring to indicate the supergroup these belong to. The differential coloring of the phylogenetic profiles indicates whether a particular protein was (1) previously called absent but in this study found using six frame-translated genome contigs and predicted [red], (2) newly discovered using improved HMM model searches [yellow] and (3) repredicted because of an incomplete protein sequence [black]. **(B)** The cartoon illustrates the approximate position and composition of all 48 kinetochore proteins studied in this figure. Black lines indicate that a protein was likely present in the Last Eukaryotic Common Ancestor (LECA). Blue and red lines point to lineage-specific kinetochore components associated with higher vertebrates and fungi, respectively. Protein colors correspond to those in the t-SNE map in Figure 1c and the matrix in Figure 1a (see legend at the bottom). **(C)** The co-evolution of the 48 kinetochore proteins is visualized using a Barnes-Hut implementation of t-Distributed Stochastic Neighbor Embedding (t-SNE) [8] based on their pairwise Pearson correlation coefficient of the phylogenetic profiles. The protein (names) are colored according to their complex memberships, identical to Figure 1a. Separate clustering of various proteins indicates strong co-evolution. For instance proteins of the CCAN are divided into six clusters indicating their (1) biochemical composition (CenpOPQUTW, HIKMLN) (2) lineage-specific presence (CenpBR, Nkp1-2, (3) widespread conservation in the kinetochores of all eukaryotes (CA) and (4) dual roles in the kinetochore and other protein complexes (SX).

BOX1: presence-absence profile of Csm1 in eukaryotes

The monopolin member Csm1 is widely present in species that lack its monopolin-binding partners. This striking pattern poses the question what the function of Csm1 at kinetochore might be in other eukaryotic lineages. Studies of Pcs1, a Csm1 ortholog in fission yeast, suggest that it may be involved in the prevention of merotelic attachments [225,226]. In budding yeast, a second function of Csm1 is to shuttle in and out of the nucleolus to regulate the silencing of the rDNA locus [226,227]. However, this function is dependent on the budding yeast-specific subunit Lrs4 and therefore likely not widespread. Knockout of the *Arabidopsis thaliana* homolog of Csm1 (Titan-9) results in the classic titan phenotype leading to problems in endosperm development and thus suggesting a role in meiosis [228,229]. Last the highly divergent Csm1 ortholog in *Dictyostelium discoideum*, named Cenp-68, was found to be involved in associating clustered kinetochores to the spindle pole body-like structure during interphase [230]. Altogether these studies indicate that Csm1 is in some way involved in bridging kinetochores during mitosis or meiosis. Strikingly however, Csm1 is lost in most metazoans, except for early-branching species *Nematostella vectensis* and *Amphimedon queenslandica* (Figure 1a). Clearly both the functional and evolutionary aspects of Csm1 remain to be explored and it would be interesting to assess its function at kinetochores outside fungal lineages. Our analysis also illustrates that it may be worthwhile to explore lineage-specific kinetochore subunits because it may well be that these are conserved in a broader range of eukaryotes than presently reported.

A pipeline to study the evolutionary dynamics of conserved features of protein families in distantly-related eukaryotes

While we observe that many kinetochore proteins co-evolve at a protein level, kinetochore proteins might also co-evolve at different levels, such as protein-motif or even domain-motif or motif-motif. To capture the evolutionary dynamics of domains and motifs of kinetochore proteins in diverse eukaryotes, we developed the partially manual pipeline ConFeaX (Conserved Feature eXtraction, see Figure 2). Simpler versions of the pipeline were used for our characterization of the evolution of Knl1 and the MadBub kinase family (Chapters 5 and 6) [114,133]. We specifically developed our ConFeaX approach in favor of currently available methods [231,232], because these heavily rely on high-quality full-length alignment of protein sequences and therefore do not allow for the detection of repeated or dynamic non-syntenic conserved features (which is a common characteristic of rapidly evolving features such as SLiMs). ConFeaX is therefore better tuned to finding conserved features in a eukaryote-wide sequence dataset in general and specifically in cases of recurrent loss or rearrangement, which commonly frustrate full-length multiple sequence alignment methods.

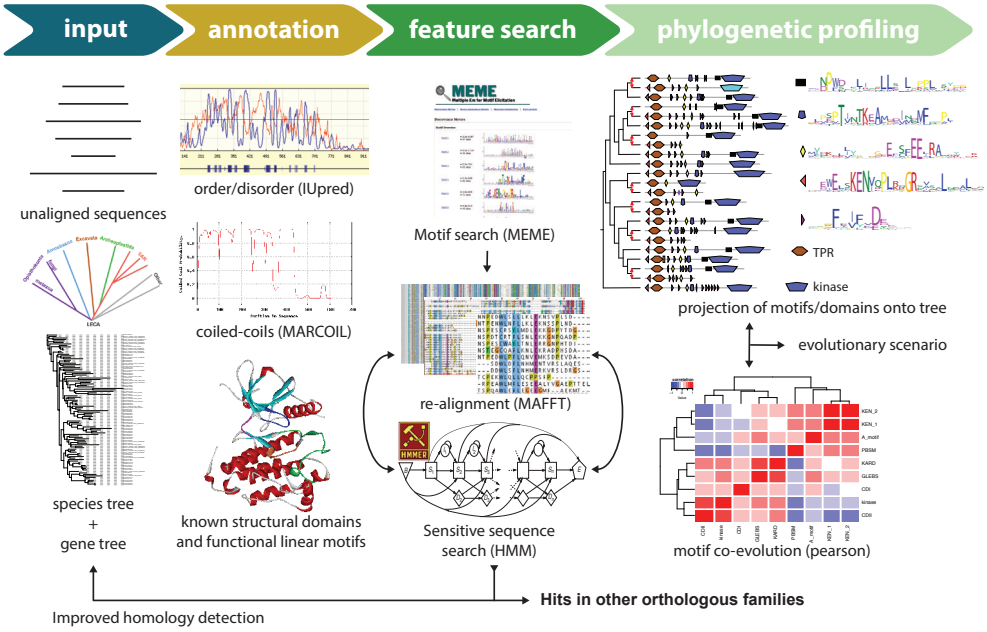


Figure 2 Overview of the ConFeaX workflow. The workflow is discussed in the results and the methods. Briefly it works as follows: **(1)** The main input for the ConFeaX pipeline are the sequences of orthologous protein families. The gene tree and species tree can be used for interpretation in certain cases of the co-evolutionary analyses for example when potential horizontal gene transfers or the presence of paralogous sequences is likely to play a role. **(2)** Sequences are annotated and masked based on their order/disorder state, tendency to form coiled-coils and known structural domains or functional linear motifs. **(3)** The MEME algorithm searches the (un) masked sequences for gapless conserved regions. Hits are aligned and subsequent HMM profile searches are iterated until no new features are uncovered. HMM profiles of the resulting conserved features are search against the complete to either find their presence in other orthologous protein families or to aid in the detection of previously unrecognized orthologous sequences. **(4)** The conserved domains or motifs are projected onto the tree and presence-absence profiles (0=absence, 1=presence) of each feature are established. Clustering of the Pearson correlation scores of the phylogenetic profiles allows for the detection of co-evolving features.

Our ConFeaX workflow encompasses the following procedures: **(1)** Orthologous sequences are annotated and masked based on predicted disorder/order regions (IUpred) [233], the tendency to form coiled-coils (MARCOIL) [234] and known structural domains and functional short linear motifs (literature-based curation and HMM-profile searches). Sequences are specifically masked to optimize for the detection of short linear motifs (SLiMs). **(2)** Both masked and unmasked sequences are searched for significantly similar gapless amino acid stretches of variable length (6-100) using the MEME algorithm [218]. Hits are extended on both sides and aligned using MAFFT to introduce gaps [212]. The alignments are modeled using the HMMER package (jackhmmer) [235] and sensitive profile HMM searches were iterated and specifically optimized using

permissive E-values/bit-scores until convergence. Due to oversensitivity issues, HMM searches are manually scrutinized for incorrectly identified features and supplemented with known instances, when applicable. Hits with a limited phylogenetic distribution (N < 5) are discarded. Subsequently, HMM models of the final hits are searched against the complete proteome set to detect additional conserved feature occurences that may indicate similar functions in other protein families or aid in the detection of previously unrecognized homologs. **(3)** To obtain an evolutionary scenario, we project presences and absences (phylogenetic profiles) of the discovered domains and motifs onto the gene tree and assess potential co-evolution by quantifying the similarity of the phylogenetic profiles of each motif/domain/protein pair using the Pearson correlation coefficient (*r)*. Finally, we manually reconcile patterns emerging from the former two steps with the known species tree.

Deconstructing the evolution of the Mis12 complex in eukaryotes

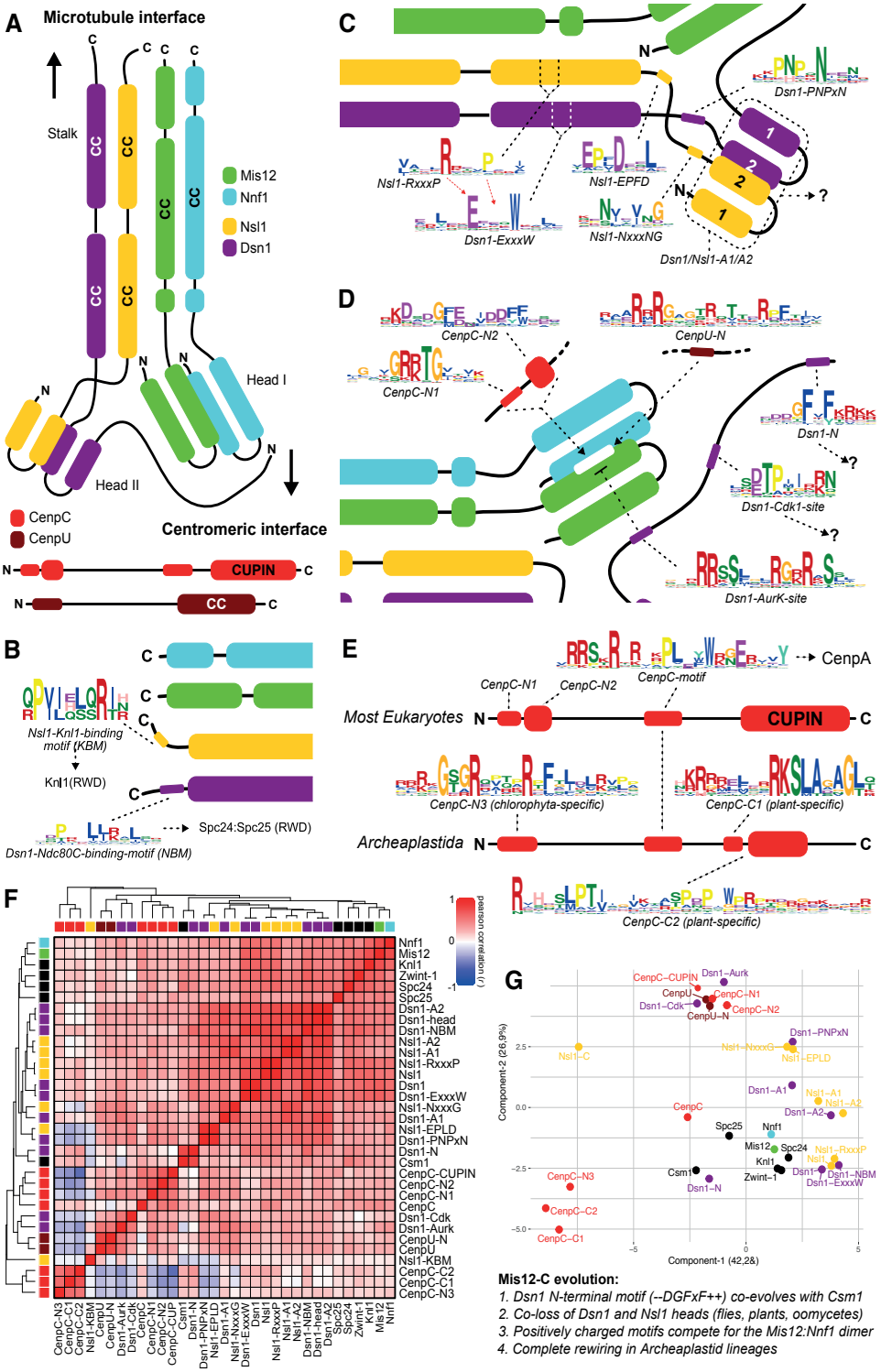
We first chose to interrogate the outer kinetochore Mis12 complex for 3 reasons. First, the structure of both the human and budding yeast *(Kluyveromyces lactis*) Mis12 complex was recently solved in great detail [46,47] and would allow us to interpret specific patterns of co-evolution and create better hypotheses for the functions of the novel motifs we uncover. Second, the Mis12 complex functions as a hub of the kinetochore and has many interactors since it bridges the CCAN and the Knl1 and Ndc80 complexes. Hub proteins cannot co-evolve with all of their interactors at a protein-protein level and the Mis12 complex would therefore be a likely candidate for the detection of motif-protein co-evolution. And lastly, Mis12 complex members were found to evolve rapidly [51,52]. While rapid sequence evolution usually hampers full-length alignment based methods, we designed ConFeaX to perform better in the context of divergent sequences.

The Mis12 complex has a Y-shaped structure, harboring two head domains that are formed by the heterodimers Mis12:Nnf1 (head I) and Dsn1:Nsl1 (head II), and a stalk that is assembled upon the tetramerization of long C-terminal coiled-coil segments of all 4 subunits [46,47] (Figure 3a). The head domains project towards the centromere and connect to the inner kinetochore via its main scaffolding protein CenpC and CenpU, while the stalk points in the opposite direction of the microtubule-binding interface and binds RWD domain-containing subunits of the Knl1 complex (Knl1:Zwint-1) and the Ndc80 complex (via Spc24:Spc25). The long disordered N-terminal tail of Dsn1 is thought to regulate the binding of the Mis12 complex with its various interaction partners and harbors a number of short linear motifs [18]. We ran our ConFeaX pipeline on the orthologs of the four Mis12 complex members, CenpC and CenpU and discovered both known and novel motifs (Figure 3, Matrix

S2, see for the sequences of the ConFeaX hits Sequence File S2). We discuss the (co-)evolution of these motifs in light of both known functional interactions and structural features that pertain to three parts of the Mis12 complex (Figure 3).

At the tip of the stalk, specific motifs in the C-termini of Nsl1 and Dsn1 were reported to interact with the RWD domains of Knl1 [54] and the Spc24:Spc25 dimer [29], respectively (Figure 3b). ConFeaX detects a single conserved motif in the C-terminus of Dsn1 (Dsn1-Ndc80 complex-binding motif (NBM), see Figure 3b). Dsn1-NBM is only lost in lineages with already known divergent Mis12 complexes (flies and worms) and therefore indicates that Spc24:Spc25 binding is a conserved function of Dsn1 in most eukaryotes. The C-terminus of Nsl1 changes so rapidly that only very limited lineage-specific motifs could be found, which do not meet our N=5 criterion. We have therefore added the 3 occurences (human, mouse and frog) of a short motif that was found in human Nsl1 to indicate the lineage-specific changes that occur at its C-terminus (Nsl1-Knl1-binding motif, see Figure 3b). Likely the rapid sequence evolution of Nsl1 mirrors that of its interactor Knl1 [133]. Furthermore, we could not detect any specific conserved motif in the C-termini of Mis12 and Nnf1, indicating that Nsl1 and Dsn1 form the primary docking site for the Knl1 -and Ndc80 complex.

The Dsn1:Nsl1 dimer (head II) diverges extensively, but two structural motifs define all their orthologs: an ExxxW motif in Dsn1 and a RxxxP motif in Nsl1. Interestingly, Dsn1-ExxxW and Nsl1-RxxxP interact and cause a kink in the first coiled-coils of Nsl1:Dsn1 near the hinge region of Mis12:Nnf1 [46,47] (Figure 3c), suggesting a central and conserved role for these motifs in the formation of the Mis12 complex. Furthermore, ConFeaX reported 3 motifs in the flexible hinge region of Dsn1:Nsl1, with a limited phylogenetic distribution: Dsn1-PNPxN, Nsl1-NxxxNG and Nsl1-EPFD (Figure 3c). While these motifs show highly similar phylogenetic patterns (Matrix S2), the extent of sequence variation in the hinge region in eukaryotes is of such magnitude that additional sequences are likely needed for our pipeline to detect conserved motifs in other eukaryotic lineages. The phylogenetic profiles of the N-terminal alpha-helices of Dsn1 (A1-A2) and Nsl1 (A1-A2) that constitute head II, are highly similar ($0,7<r<0,8$) (Figure 3f,g) suggesting strong co-evolution. They also reveal a striking recurrent loss of either parts or the full head II structure in phylogenetically distant lineages such as the protochordate *Branchiostoma floridae*, winged insects (e.g. *Drosophila melanogaster* and *Bombyx mori*), land plants (e.g. *Physcomitrella patens and Arabidopsis thaliana*) and oomycetes (e.g. *Albugo candida* and *Phytophthora infestans*). While it is currently unclear what the function of head II is, it was recently suggested that a head I-to-head II orientation of 5-6 Mis12 complexes forms the circular basis for the budding yeast kinetochore complex [18,47,56]. In light of this hypothesis we speculate



Mis12-C evolution:

1. Dsn1 N-terminal motif (--DGFxF++) co-evolves with Csm1
2. Co-loss of Dsn1 and Nsl1 heads (flies, plants, oomycetes)
3. Positively charged motifs compete for the Mis12:Nnf1 dimer
4. Complete rewiring in Archeaplastid lineages

Figure 3 Molecular evolution of the Mis12 complex. (A) The Mis12 complex is Y-shaped structure that consists of 4 subunits. The Mis12:Nnf1 heterodimer forms head I and the Dsn1:Nsl1 heterodimer forms head II. The stalk is formed through the tetramerization of C-terminal coiled-coils (CC) of each subunit. Head I and II project towards the centromere and the stalk points into the opposite direction, towards the microtubule-binding interface. This cartoon is based on two recently published structures of human Mis12 (pdb:5LSK) [46] and *Kluveromyces lactis* MIND complex (pdb:5T58) [47]. The motif and domain topology of the Centromeric proteins CenpC and CenpU is depicted, which is further described in panel D and E. In general for this figure: sequence logos represent the motifs and domains that are found by our ConFeaX pipeline; the height of the sequence logo indicates the conservation i.e. higher is more conserved. The coloring scheme is based on an adaptation of the Clustal used in MEME. Panel B-D contain zoom ins of various parts of the Mis12 complex as depicted in panel A. Arrows indicate interactions; questions marks point to unknown interactors. Panel F and G reveal the co-evolutionary patterns of the motifs that are shown in panel B-E. **(B)** Zoom in on the C-terminal end of the stalk. Both Dsn1 and Nsl1 contain one C-terminal motif that is involved in the recruitment of the Ndc80 –and Knl1-complex, respectively. **(C)** Zoom in of the rapid evolving Dsn1:Nsl1 dimer (head II). The N-terminal alpha-helices of Dsn1 and Nsl1 that form head II co-evolve and are lost together in various lineages (helices are numbered 1 and 2). Strikingly, no explicit function has been described for the helix 1 and 2 of Dsn1 and Nsl1. Furthermore, ConFeaX reported various motifs in the C-terminal half of Dsn1 and Nsl1. See for elaborate discussion point 2 of the section describing the evolution of the Mis12 complex. **(D)** Zoom in on the Mis12:Nnf1 dimer (head I) and the disordered N-terminal tail of Dsn1. Highly charged motifs in the N-termini of Dsn1, CenpC and CenpU compete for a conserved binding pocket on head I. Note the presence of the Dsn1-N motif whose distribution across eukaryotes seems to correlate with the presence of Csm1 across species. **(E)** Top: topology of CenpC in most eukaryotes. Bottom: extensive divergent CenpC orthologs in plants (archeaplastids) replaced the CUPIN domain by a plant-specific C-terminal extension (CenpC-C1/2). In addition the characteristic CenpC-N1 motif, present in most eukaryotes is lost as well. Only in chlorophyte lineages (green algae) a motif similar to CenpU-N was reported by ConFeaX. **(F)** Average clustering of Pearson correlation coefficient (distances = $1 - r$) of any pair of motifs/domains and proteins that are shown in this figure. High correlation scores (red) indicate co-evolution. Colors on both sides of the matrix reflect the colors used for proteins used in this figure. Each feature that is associated to a protein has the same color. Black indicates proteins that are not depicted in this figure but are interactors of various motifs here discussed. **(G)** Principal components analysis of the Pearson correlation coefficient (distances = $1 - r$) of each motif, domain and protein pair. Explanation of the variance: prin. comp.1 = 42,2%, prin. comp. 2 = 26,9%. Colors are similar to panel F.

that the rapid evolution of the Dsn1:Nsl1 dimer likely reflects changes in the basic kinetochore architecture of species such as *Drosophila melanogaster* and *Arabidopsis thaliana*.

In contrast to Dsn1:Nsl1, the Mis12:Nnf1 dimer (head I) is highly conserved and ConFeaX did not assign any additional motifs outside of the structured parts. The first alpha-helix of Nnf1 (A1) and the second alpha-helix of Mis12 (A2) form a composite interaction surface on head I for the competitive binding of highly charged motifs in the N-termini of CenpC and Ame1 (CenpU ortholog in *Kluyveromyces lactis*) [46,47] (Figure 3d). Several binding assays suggest that a similar motif in the N-terminus of Dsn1 autoinhibits the localization of the

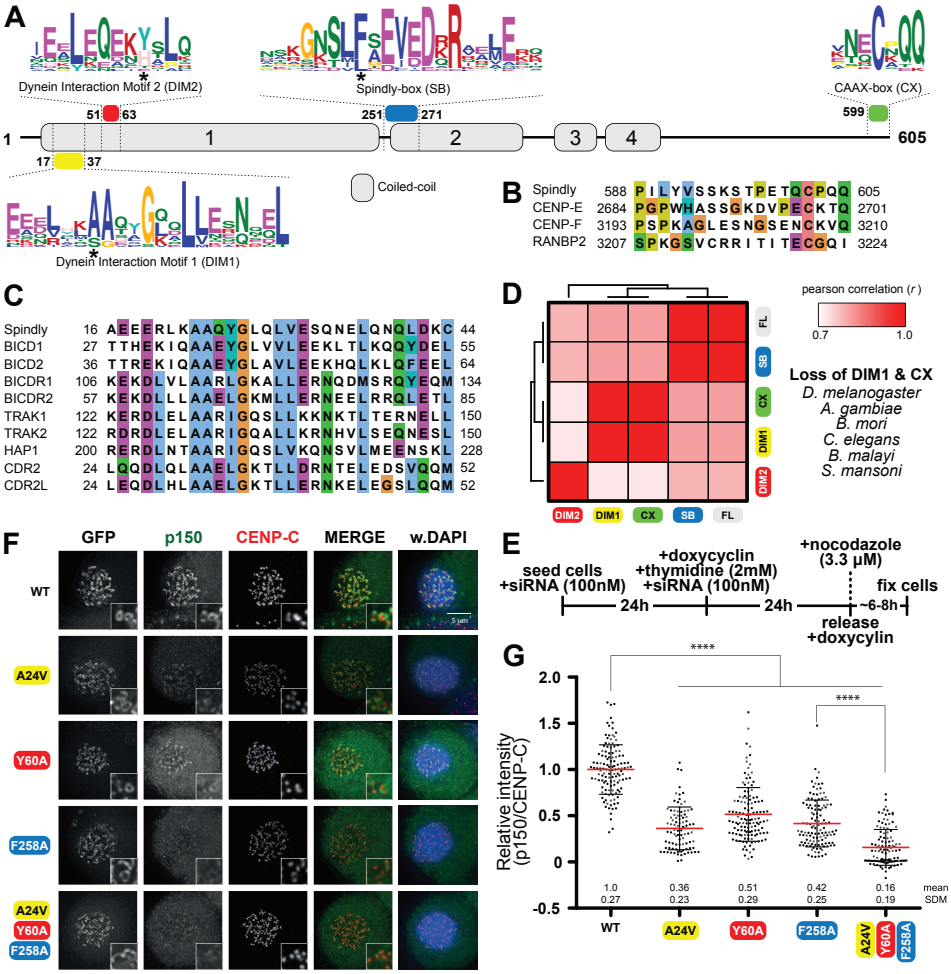


Figure 4 The N-terminus of hSpindly is homologous to dynein-dynactin adaptors and is involved in the recruitment of the dynactin subunit p150glued to the kinetochore. **(A)** Overview of the secondary structure of hSpindly with 4 sequence logos that represent the hits reported by the ConFeaX pipeline. The height of the sequence logos indicates the conservation. Colors of the motifs are used throughout the other panels. **(B)** Multiple sequence alignment of CAAX-box-like motifs in the C-termini of the human orthologs of Spindly, CenpE, CenpF and RanBP2. **(C)** Multiple sequence alignment of DIM1 in the coiled-coil regions of various human proteins involved in the recruitment of dynein and dynactin. **(D)** Average clustering based on the Pearson correlation coefficient of the phylogenetic profiles of each of the 4 motifs and full-length (FL) Spindly. **(E)** Protocol used to perform the experiments shown in panel F and G. **(F-G)** Representative images (F) and quantification (G) of GFP-Spindly-expressing HeLa Flp-in cells transfected with siRNA to Spindly and treated with high doses (3uM) of nocodazole. GFP-Spindly is shown in left the panel to indicate similar expression levels of the WT and mutants. p150 is shown in green, CENP-C in red. Merges of the p150 and CENP-C signals indicate co-localization and are represented with and without the DNA (DAPI). Quantification in G shows mean kinetochore signal intensity (+SD) of p150glued over CENP-C. Data are from >30 cells and representative of 3 experiments. Levels of p150/CENP-C in WT-Spindly-expressing cells are set to 1.

Mis12 complex to the kinetochore, which is relieved upon phosphorylation by the centromere-associated kinase Aurora B [46,47]. While the N-termini of CenpC, CenpU and Dsn1 evolve at a rapid pace, ConFeaX detected several motifs that are characterized by a high number of positive and negative charges, various conserved glycines, phenylalanines and potential phosphosites (CenpC-N1/-N2, CenpU-N, Dsn1-N-AurK-Cdk). The pattern of glycines and arginines in CenpC-N1, CenpU-N and Dsn1-Aurk are reminiscent of AT-hooks and may well indicate a dual role in binding both the Mis12 complex as well as DNA and/or microtubules. Although the phylogenetic profiles of CenpC-N1/N2, CenpU-N and Dsn1-Aurk/Cdk are all rather patchy, their correlation scores cluster them together (0,3<*r*<0,7). As this may indicate co-evolution, we predict that they have a combined role in the competitive binding of Mis12:Nnf1 (Figure 3f,g). Interestingly, all these motifs are absent in lineages of the Archeaplastida (Matrix S2, Figure 3e) and ConFeaX only finds a specific motif in chlorophyta (green algae) that resembles CenpU-N (green algae, see Figure 3e: CenpC-N3). In addition, CenpC in archeaplastid lineages replaced the C-terminal dimerization domain (Cupin) by a yet uncharacterized C-terminal extension, containing two conserved features (CenpC-C1/C2) (Figure 3e). Surprisingly, while CenpC diverges extensively in plant lineages we could not detect any obvious changes in the binding pocket of Mis12:Nnf1, suggesting that perhaps other proteins regulate the kinetochore localization of the Mis12 complex. Last, the most N-terminal motif in the disordered tail of Dsn1 (Dsn1-N) is characterized by two conserved phenylalanines flanked by stretches of either negatively charged (N-terminal) or positively charged (C-terminal) amino acids (Figure 3d). Being present in all eukaryotic supergroups but absent in most metazoans, the peculiar phylogenetic profile of Dsn1-N (Matrix S2) is strikingly similar to that of Csm1 (*r*=0.8) (Figure 3d,f,g). This suggests that Dsn1-N is the kinetochore-targeting motif of Csm1 in many eukaryotic lineages. In support of this, a recent yeast-2-hybrid screen for the binding of the Monopolin complex suggested that the first 110 amino acids of budding yeast Dsn1 interact with Csm1 [236]. Strikingly, in oomycetes the highly charged regions that flank the conserved phenylalanines in Dsn1-N flip, meaning that positively residues are now N-terminal and negatively charged residue C-terminal (Sequence File S2). We could however not detect any clear co-evolving features in the RWD domain that would indicate its interaction site.

Altogether, we find that our ConFeaX workflow recapitulates many of the recently discovered functional motifs in the Mis12 complex (e.g. CenpC-N1/N2, Dsn1-NBM). Using the structural information we could better interpret the logic of co-evolving features e.g. Dsn1-ExxxW - Nsl1-RxxxP and Dsn1-PNPxN – Nsl1-EPFD (Figure 3c). The strong co-evolution of Dsn1-N and Csm1 underlines the potential power of our approach and warrants similar analyses for other kinetochore complexes in the future. The evolution of the Mis12 com-

plex has two faces. The Mis12:Nnf1 dimer is mostly conserved. The Dsn1:Nsl1 dimer however, diverges extensively both in terms of its structure (loss of the head II in many lineages) and the number of its motifs (Dsn1 N-terminus). Strikingly, this rapid sequence evolution is also observed in other kinetochore proteins such as Knl1 [133]. What this exactly means is currently unclear.

Discovery of two novel motifs in Spindly that are involved in the recruitment of the dynein-dynactin complex at kinetochores

Spindly is a ~600 amino acid long coiled-coil protein that was shown to be involved in the recruitment of dynein to the kinetochore through a conserved motif called the Spindly box [148,183,184]. At the kinetochore, Spindly drives the silencing of the spindle assembly checkpoint through the activation of dynein-mediated stripping of its other interaction partner the RZZ (Rod-Zw10-Zwilch) complex [148,149]. In our previous survey of kinetochore evolution we showed that spindly co-evolved with the RZZ components Zwilch and Rod and was likely invented in the last common ancestor of opisthokonts [52]. At the time we started this analysis not much was know about the molecular evolution of Spindly beyond the conservation of the Spindly box motif. We therefore sought to use our ConFeaX approach to characterize its molecular evolution. We discuss our findings in light of recently discovered functional motifs [78,79,237,238].

ConFeaX was run on our improved set of Spindly orthologs (Sequence File S1). This revealed the presence of 3 novel conserved motifs and the known Spindly box (Figure 4a, Matrix S3, Sequence File S3). In the C-terminus of Spindly we find a conserved motif that is similar to a CAAX-box (CX). In proteins of the small GTPase family such as RAS, the reactive cysteine residue in a CAAX-box is modified with a hydrophobic farnesyl moiety, which drives its recruitment to the membrane [239]. Strikingly, a number of kinetochore-associated proteins such as CenpE, CenpF and RanBP2 (Figure 4b) contain a similar motif and it was recently shown that farnesylation is needed for the kinetochore localization of Spindly through its association with the RZZ complex [78,237,238]. Two additional conserved motifs are found in the first coiled-coil of Spindly with a consensus AAxxGxxLL and Qxx[HY] (Figure 4a). Interestingly, these motifs are also found in the N-terminus of the human dynein-dynactin adaptor BICD2 and a number of other related microtubule motor adaptors [240] (BICD1, BICDR1/2, TRAK1/2 and HAP1, see Figure 4c), suggesting that these proteins may share a common ancestor and have a similar function. Mutations of the AAxxGxxLL motif in BICD2 and BICDR1 abrogate the binding of the dynein-dynactin complex [241–243]; hence we termed these motifs, Dynein-Dynactin Interaction Motif 1 and 2 (DIM1/2). The presence of a DIM1-like motif in human CDR2 (Cerebellar degeneration related 2) and CDR2L,

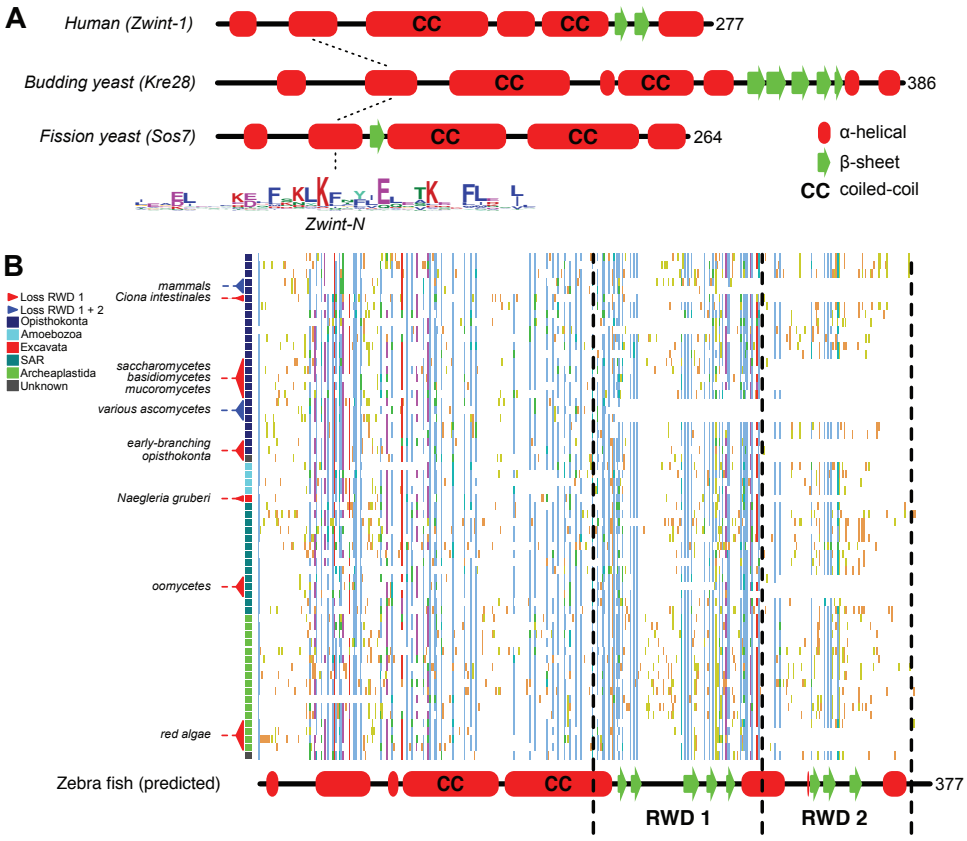


Figure 5 Recurrent loss of RWD domains during the evolution of Zwint-1. (A) Secondary structure prediction of three Zwint-like orthologs: Zwint-1 (human), Kre28 (budding yeast) and Sos7 (fission yeast). **(B)** Overview of a multiple sequence alignment of 83 Zwint-1 orthologs that was based on the manually predicted Zwint-like sequence in zebra fish. 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. On the left the small colored blocks indicate the supergroup to which each species belongs. The red and blue arrow-like structures indicate which of the orthologs lost the RWD-2 or both the RWD-1 and RWD-2 domain, respectively.

suggests that these proteins, for which no current function was known, are also involved in dynein-mediated transport (Figure 4c). Strikingly, we find that the phylogenetic profiles of DIM1 and the CAAX-box strongly correlate (r=1) and that DIM1 and the CAAX-box motif are simultaneously lost in flies, worms and trematodes (Figure 4d). What this means is currently unclear but it strongly advocates that these two motifs are somehow functionally linked.

To examine if our newly detected DIM1 and DIM2 are important for Spindly function as a Dynein-Dynactin adaptor at kinetochores, we mutated them in a GFP-Spindly construct that was resistant to Spindly RNAi. Of note: two recent

studies reported on the identification of a DIM1-like motif, and suggested it is a dynactin-binding motif [78,79]. Depletion of endogenous Spindly and expression of GFP-Spindly in nocodazole-arrested cells resulted in high levels of the dynactin subunit p150glued at kinetochores (Figure 4f,g). Strikingly, expressing Spindly mutants of DIM1 (A24V), DIM2 (Y60A) and the spindly box (F258A) markedly decreased p150glued levels at kinetochores. A triple mutant further lowers the levels of p150glued, indicating that DIM1, DIM2 and the spindly box cooperate in the stabilization of the dynein-dynactin complex at kinetochores (Figure 4g,f). Altogether, our findings suggest that Spindly functions as a dynein-dynactin adaptor at kinetochores using a similar conserved motif (DIM1) that was previously described for BICD2 and BICDR1 [240].

Recurrent loss of RWD domains in diverged Zwint-like orthologs

We previously showed that Zwint-1 in humans, Kre28 in budding yeast and Sos7 in fission yeast are part of the same divergent orthologous family of Zwint-like proteins, which was likely present in the Last Eukaryotic Common Ancestor (LECA) [52]. While our findings were confirmed by the fact that these three proteins interact with their respective Knl1 orthologs (Casc5/Knl1, Spc105 and Spc7) and that they were all reported to share a coiled-coil topology, we wondered whether a detailed analysis of their sequence evolution might give some hints on what the function of Zwint-like proteins would be.

ConFeaX reported a conserved N-terminal region (Zwint-N, Sequence File S4, Matrix S4) with predicted alpha-helical structure (Figure 5a) and many divergent lineage-specific motifs in the C-terminus (data not shown). While this did not provide us with any further insight we used a multiple sequence alignment (MSA) of all 83 orthologs to predict the secondary structure of Zwint-1, Kre28 and Sos7 (see Figure 5a and Materials and Methods). While we corroborate previous reports on large coiled-coil segments, we found some predicted beta-sheets, albeit with low confidence, in the C-terminal regions of Kre28 and Zwint-1 (see Figure S1). Upon further inspection of the alignment we observed that multiple Zwint-like orthologs have an extended C-terminal region compared to Zwint-1, Kre28 and Sos7. We therefore performed the same procedure on the Zwint-like ortholog in zebra fish, a sequence that was previously called absent and that we manually predicted in this study (see previous section on gene prediction of orthologs, Sequence File S1). Strikingly, we now observed two regions of consecutive beta-sheets, interleaved by alpha-helices, a structure which is reminiscent of the double RWD domain structure of Knl1 (Figure 5b, Sequence File S4, Matrix S4). Using the secondary structure of the zebra fish ortholog as an anchor for the Zwint-1 ortholog alignment, we find that most Zwint-like orthologs contain a double RWD structure (e.g. plants, Amoebozoa, early-branching stramenopiles, Rhizaria, various animals and verte-

brates). On the other hand we observe the recurrent loss of either the last RWD domain (*Ciona intestinales*, saccharomycetes, mucoromycota, basidiomycetes, early-branching opisthokonts, *Naegleria gruberi*, oomycetes and red algae) or both RWD domains (fission yeast, *Neurospora crassa*, *Yarrowia lipolytica* and in human and mouse) (Figure 5b). We tried to correlate the loss of a single and/or double RWD domain in Zwint-like orthologs but we could not detect any co-evolutionary patterns.

Altogether, we find that Zwint-1, Kre28 and Sos7 are descendant of an ancestor with a double RWD topology, which was likely part of the kinetochore in LECA. Although most RWD domain-containing kinetochore proteins diverge extensively, it is remarkable that Zwint-like orthologs are allowed to recurrently lose either single or double RWD domains. Possibly this behavior reflects the rapid evolution of the Knl1 RWD domains that we previously observed (see chapter 5) [133]. In addition we observe a similar pattern for the Spc24:Spc25 dimer, for which the RWD domain of Spc24 seems to evolve more rapidly then that of Spc25. Interestingly, with the addition of Zwint-like, the kinetochore of LECA contained four proteins with single (Spc24, Spc25, Mad1, Csm1) and four with tandem (CenpO, CenpP, Knl1 and Zwint-like) RWD domains. Given that 6 of the 8 RWD domain-containing proteins are part of a heterodimer (Spc24:Spc25, CenpO:CenpP and Knl1:Zwint-like), it is tempting to speculate that kinetochore complexity arose by internal duplication of RWD-containing proteins. Timing these duplications by making phylogenetic trees of the RWD domains would possibly allow us to uncover the order of RWD domain duplications and thereby draw an image of the origin of the kinetochore.

In conclusion, we show here the power of using orthologous sequences to detect conserved protein features that predict functionality. Our approach should be widely applicable to other protein networks and biological processes.

Materials and Methods

Proteome database

We compiled a database of 109 proteomes based on sets that our labs used in previous studies. For the versions and sources of the selected proteomes we therefore refer to two studies of van Hoof et al. 2017 [52,87]. Notable exceptions are the proteomes of *Bombyx mori*, *Nasonia vitripennis* and *Agaricus bisporus*, which we have downloaded on the 12th of January from the ensembl genomes database (http://ensemblgenomes.org/). In addition, we received the proteome of the amoebozoa *Physarum polycephalum* from the lab of Pauline Schaap (see for contigs http://www.physarum-blast.ovgu.de/). A unique protein identifier was assigned to each protein, consisting of four letters and six

numbers. The letters combine the first letter of the genus name with the first three letters of the species name. If multiple splice variants of a gene were annotated, the longest protein was chosen.

Orthologs

To create our set of orthologs we searched the 109 proteomes using our in-house established kinetochore protein HMM profiles of SAC-related, CCAN and KMN network proteins [52]. In cases where HMM profile searches where incomplete or inconclusive we manually searched for orthologs using previously established procedures and criteria [52] (see also chapter 2). In addition, we performed phylogenetic profiling of 7 lineage-specific kinetochore proteins that were not included in our previous analysis (Slk19, Csm1, Lrs4, Mam1, Nkp1, Nkp2 and CenpB). Orthologs used in this study can be found in Sequence File S1.

Gene search and gene prediction

To systematically search for genes that were absent in our previous analyses we adopted 3 strategies: **(1)** we used our custom made HMM models of either orthologous groups or specific features such as domains and motifs, to search for a gene of interest in six-frame translated genome contigs **(2)** we used an orthologous sequence of a closely-related species to query whole genome shotgun sequences using tblastn **(3)** we used an orthologous sequence of a closely-related species to query six-frame translated genome contigs using phmmer. To assess sequence quality issues we manually flagged incomplete proteins based on multiple sequence alignments of orthologous protein families. Proteins were deemed incomplete in case stretches of at least 15 amino acids were found missing. Common mistakes include incorrect gene fissions and fusions and wrongly omitted exons. Predicted or incomplete gene regions were extended with <50000 bp and used to predict a gene by GENESCAN [244] and AUGUSTUS [245] (using various species-trained models).

Conserved feature extraction workflow and co-evolutionary analyses

Orthologous sequences were masked using IUpred [246] (disorder/order threshold = 0.4) and MARCOIL [234] (coiled-coil threshold = 90). ConFeaX starts with a probabilistic search for short conserved regions (6-100 aa) using the MEME algorithm (option: any number of repeats) [218]. Significant motif hits are extended on both sides by five residues to compensate for the strict treatment of alignment information by the MEME algorithm and aligned using MAFFT-LINSI [212] to introduce gaps. The alignments were modeled using the HMMER packing [235] and sensitive profile HMM searches were iterated (jackhmmer-like approach; E-value =1) until convergence. In some cases we manually optimized the HMM profile searches using permissive bit-scores and

removed obvious false hits manually. Hits with a limited phylogenetic distribution (N < 5) were discarded. Subsequently, for each of the conserved features, a phylogenetic profile was derived (present is '1' and absent is '0'). For all possible pairs, we determined the correlation using Pearson correlation coefficient [219]. Average clustering based on Pearson distances (d=1-r) was used to indicate co-evolution. The Pearson distances were also used to map the kinetochore proteins in 2D using (1) Barnes-Hut *t*-SNE [174] (R-package 'Rtsne' [perplexity=4, dimensions=2 and theta=0], see Figure 1b) and (2) a principal component analysis (R-package 'prcomp', see Figure 3g). Sequence logos used in Figure 3-5 were obtained using weblogo2 [247].

Structure prediction of Zwint-1 homologs

The secondary structures of Zwint-1 (human), Kre28 (budding yeast), Sos7 (fission yeast) and drZwint-1 (zebra fish) were determined using the Jpred4 server [248]. As input we provided a multiple sequence alignment of all 83 Zwint-like orthologs (Sequence File 1) that we modified so that the query sequence would not contain any gaps (Figure S1).

Plasmids, cell culture and transfection

RNAi-resistant LAP (GFP)-SPINDLY was a gift of Reto Gassmann, Instituto de Biologia Molecular e Celular, Universidade do Porto. To acquire mutants, site-directed mutagenesis was performed using the quickchange strategy. HeLa-FlpIn TRex cells were grown in DMEM high glucose supplemented with 10% Tet-free FBS (Clontech), penicillin/streptomycin (50 mg ml^{-1}) and alanyl-glutamine (Sigma, 2 mM). pcDNA5-constructs were co-transfected with pOgg44 recombinase in a 10:1 ratio using FuGEHE HD (Roche) as a transfection reagent. After transfection, the medium was supplemented with hygromycin (200 µg ml^{-1}) and blasti-cidin (8 µg ml^{-1}) until cells were fully confluent in a 10 cm culture dish. siSPINDLY (5'-GAAAGGGUCUCAAACUGAA-3' custom Dharmacon) was transfected using Hiperfect (Qiagen) at 100 nM for 48 hours according to the manufacturer's guidelines. Constructs were expressed by addition of doxycycline (1 µg ml^{-1}) for 24 hours.

Immunofluorescence and antibodies

Thymidine-arrested cells were directly released in nocodazole (3,3 µM) and fixed 6-8 hours after the release. Cells plated on 12-mm coverslips were fixed and extracted (with icecold methanol) for 20min at -20°C. Coverslips were washed with PBS and blocked with 3% BSA in PBS for 1 h, incubated with primary antibodies for 16 h at 4°C, washed with PBS containing 0.1% Triton X-100, and incubated with secondary antibodies for an additional hour at room temperature. Coverslips were then washed, incubated with DAPI for 2

min, and mounted using ProLong Antifade (Molecular Probes). All images were acquired on a deconvolution system (DeltaVision Elite, Applied Precision) with a 100×/1.40 NA U Plan S Apochromat objective (Olympus) using softWoRx software (Applied Precision). Images are maximum intensity projections of deconvolved stacks. For quantification of immunostainings, all images of similarly stained experiments were acquired with identical illumination settings; cells expressing comparable levels of exogenous protein were selected for analysis and analyzed using ImageJ (National Institutes of Health). An ImageJ macro was used to threshold and select all centromeres and all chromosome areas (excluding centromeres) using the DAPI and anti-centromere channels as described previously [105]. Cells were stained using GFP-booster Atto-488 (ChromoTek, gba488, ChromoTek, 1:1000), anti-centromere antibody CENPC (MBL, PD030, 1:1000)and p150Glued BD, 612708, 1:1000). Secondary antibodies were goat anti–guinea pig Alexa Fluor 647 and goat anti–rabbit and anti–mouse Alexa Fluor 568 (Molecular Probes) for immunofluorescence experiments.

Contributions

ET performed the research. CS performed the Spindly experiments. JJEH helped to determine the orthologs. ET, BS and GJPLK conceived the project and wrote the manuscript.

Acknowledgements

We want to thanks the Kops lab and the Snel lab for discussions. We thank Pauline Schaap for kindly providing us with the predicted proteome of Physarum polycephalum.

Electronic Supplementary Material

Supplementary materials for this chapter are made available online at the following link http://bioinformatics.bio.uu.nl/eelco/thesis/

Sequence File S1 contains separate fasta files with the orthologs of all proteins used in this study.

Sequence File S2-S4 contain fasta files with the motifs and domains of all proteins used in Figure 3-5, respectively.

Matrix S1-S3 contain phylogenetic profiles and their pairwise correlation scores associated to in Figure 2a, Figure 3f and Figure 4b, respectively.

Table S1 contains the counts of repredicted, newly discovered and predicted

sequences for the phylogenetic profiles as visualized in Figure 2a.

Table S2 contains a comparison of the pairwise correlation scores for 41 kinetochore proteins of the current study and that of van Hooff et al. 2017 (see Chapter 2).

Figure S1 shows an overview of the structure prediction by the JPred server (embedded within the Jalview software package) of hZwint-1, kre28, sos7 and drZwint-1 using the sequence information of all Zwint-like orthologs. These predictions are used as a basis for the visualizations in Figure 5a,b.

CHAPTER 4

Arrayed BUB recruitment modules in the kinetochore scaffold KNL1 promote accurate chromosome segregation

Mathijs Vleugel, Eelco Tromer, Manja Omerzu, Vincent Groeneveld
Wilco Nijenhuis, Berend Snel, Geert JPL Kops

Journal of Cell Biology. 2013

Abstract

Fidelity of chromosome segregation relies on coordination of chromosome biorientation and the spindle checkpoint. Central to this is the kinetochore scaffold Knl1 that integrates the functions of various mitotic regulators including Bub1 and BubR1. We show that Knl1 contains an extensive array of short linear sequence modules that encompass TxxΩ and MELT motifs and that can independently localize Bub1. Engineered Knl1 variants with few modules recruit low levels of Bub1 to kinetochores but support a robust checkpoint. Increasing numbers of modules concomitantly increase kinetochore Bub1 levels and progressively enhance efficiency of chromosome biorientation. Remarkably, normal Knl1 function is maintained by replacing all modules with a short array of naturally occurring or identical artificially designed ones. A minimal array of generic BUB recruitment modules in Knl1 thus suffices for accurate chromosome segregation. Widespread divergence in the amount and sequence of these modules in Knl1 homologues may represent flexibility in adapting regulation of mitotic processes to altered requirements for chromosome segregation during evolution.

Introduction

Equal distribution of the replicated genome during mitosis is essential for accurate propagation of genetic information and the maintenance of healthy tissues. Large multiprotein complexes known as kinetochores perform several essential functions in this process [249,250]. These include generating and maintaining physical attachment between chromatids and microtubules of the mitotic spindle, and signaling to the spindle assembly checkpoint (SAC, also known as the mitotic checkpoint) when kinetochores are unbound by microtubules. Such checkpoint signaling involves production of a diffusible inhibitor of anaphase onset [167,251].

Chromosome biorientation as well as SAC activity critically rely on the kinetochore scaffold KNL1/CASC5/AF15q14/Blinkin (hereafter referred to as Knl1 [50,252]). This long, largely unstructured protein is a member of the Knl1/MIS12 complex/NDC80 complex (KMN) network that constitutes the microtubule-binding site of kinetochores [50]. Knl1 itself directly contributes to this through its N-terminal microtubule-binding region [61,62], but also by localizing the paralogs Bub1 and BubR1 to kinetochores. The pseudokinase BubR1 [147] is a component of the mitotic checkpoint complex [251] and additionally binds the PP2A-B56 phosphatase that is required for stabilizing kinetochore–microtubule interactions [96–98,253]. Bub1, in turn, promotes efficient chromosome biorientation by localizing the Aurora B kinase to inner centromere regions via phosphorylation of H2A-T120 [254,255]. Its contribution to checkpoint signaling, although important, is not entirely clear [256,257].

Although recruitment of Bub1 and BubR1 (the BUBs) to kinetochores is critical for error-free chromosome segregation, the mechanism by which Knl1 accomplishes this is unknown. Both BUBs directly interact via their conserved TPR domains with two so-called KI motifs in the N-terminal 250 amino acids of human Knl1 [258–260]. These interactions may, however, not be required for Bub1/BubR1 kinetochore localization [260], and the KI motifs are not apparent in nonvertebrate eukaryotic Knl1 homologs [167]. In contrast, kinetochore binding of at least Bub1 relies on Mps1-mediated phosphorylation of the threonine within MELT-like sequences of Knl1 in humans and yeasts [261–263]. Such MELT-like sequences can be identified in numerous Knl1 homologs [167].

In this study, we set out to investigate the mode of BUB recruitment to kinetochores, and show that Knl1 is an assembly of previously unrecognized repeating modules. These modules operate in a generic fashion to recruit sufficient BUB proteins to kinetochores to ensure high-fidelity chromosome segregation.

Results

The N-terminal MDLT-KI module in Knl1 independently recruits BUB proteins

Bub1 and BubR1 directly bind to KI motifs (KI1 and KI2) that are located near the N terminus of Knl1 [258–260]. Their localization to kinetochores additionally requires Mps1-dependent phosphorylation of MELT-like sequences [261–263], although it is unknown which of these sequences are phosphorylated and which ones are important for BUB recruitment and Knl1 function. Because one such MELT-like sequence (MDLT) is located close to the two KI motifs, we examined whether the N-terminal region (1–261) of Knl1 encompassing MDLT-KI1-KI2 is sufficient to bind Bub1 and BubR1. To this end, the Knl1 fragment was fused to LacI and tethered to an ectopic Lac operator (LacO) array that is stably integrated in the short arm of chromosome 1, distant to the centromere (1p36) in U2OS cells (Figure S1a [264]). LacI-LAP-Knl1$^{1–261}$ recruited endogenous Bub1 and BubR1 to the LacO array in mitotic cells. (Figure 1a, Figure S1b). This required the MDLT and KI1 sequences because mutation of these motifs (MDL**T** to MDL**A** [Knl1^{MDLT}] or KID**T**TS**F** to KID**A**TS**A** [Knl1^{KI1}] [260]) prevented both BUBs from localizing to the LacO array (Figure 1a, Figure S1b). In addition, BubR1 but not Bub1 localization was also lost after mutating the KI2 motif (KID**F**ND**F** to KID**A**ND**A** [Knl1^{KI2}] [258,260]. Thus, at least in the context of the ectopic Knl1 fragment, BubR1 recruitment to Knl1 is dependent on all three motifs (Figure 1a, Figure S1b).

The N-terminal MDLT-KI module in Knl1 is sufficient to support SAC activity but not chromosome biorientation

To next assess the contribution of the N-terminal module to Knl1's function in the SAC and chromosome biorientation, we generated a LAP-tagged Knl1 variant in which this region was directly fused to the C-terminal kinetochore localization domain of Knl1 (aa 1834–2342: generating Knl1-NC, Figure S1c). This ensured maintenance of proper KMN network integrity, Knl1 position on the outer kinetochore, and Zwint-1 and HP1 kinetochore localization [252,265]. Full-length Knl1 (Knl1-FL) and the C-terminal domain of Knl1 (Knl1-C) were used as controls. To ensure comparable genetic background and expression levels, siRNA-resistant Knl1 variants were expressed from a doxycycline-inducible promoter at a single integration site in HeLa cells [257]. All Knl1 variants efficiently incorporated into the outer kinetochore to similar levels, as judged by immunofluorescence (Figure 1b-d). Functionality of these proteins was assayed by their ability to restore Knl1 function upon siRNA-mediated depletion of endogenous Knl1. Depletion of Knl1 removed Bub1 and BubR1 from kinetochores (Figure 1b-d), and this was restored by expression of Knl1-FL and weakly by Knl1-NC, but not by Knl1-C (Figure 1b-d). In support of this, comparative proteomics analysis of LAP-Knl1 pull-downs showed strong reduction in BUB

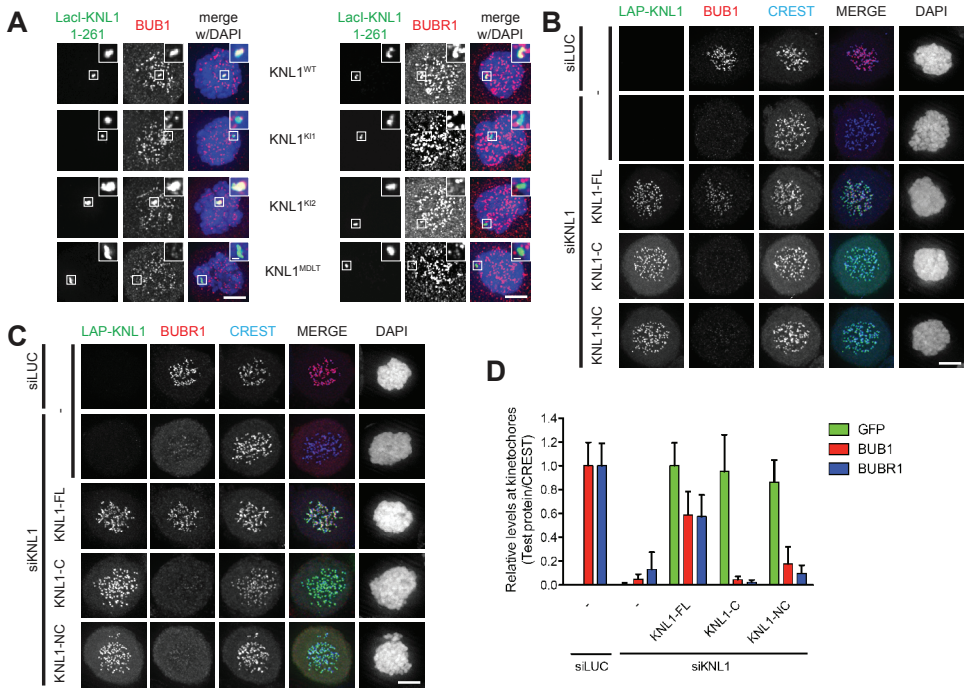


Figure 1 The N-terminal MDLT-KI module in Knl1 independently recruits BUB proteins. (A) Immunolocalization of Bub1 (left panels, red) and BubR1 (right panels, red) in LacI-LAP-Knl1$^{1-261}$-transfected, nocodazole-treated U2OS-LacO cells. LacI-LAP-Knl1$^{1–261}$ is shown in green and DNA (DAPI) is in blue. Insets show magnifications of the boxed regions. Knl1^{KI1} denotes LacI-LAP-Knl1$^{1–261}$ in which KIDTTSF is mutated to KIDATSA, Knl1^{KI2} is KIDFNDF mutated to KIDANDA, and Knl1^{MDLT} mutated is MDLT to MDLA. Bars, 5 µm (insets, 0.5 µm). **(B–D)** Representative images (B,C) and quantification (D) of LAP-Knl1–expressing Flp-in HeLa cells transfected with siRNAs to luciferase (siLUC) or to Knl1 (siKnl1) and treated with nocodazole. LAP-Knl1 is shown in green, BUB proteins in red, centromeres (CREST) in blue, and DNA (DAPI) in white. Bars, 5 µm. Quantification in D shows total kinetochore signal intensity (+SD) of LAP-Knl1,BUB proteins over CREST. Data are from >15 cells and representative of 3 experiments. Levels of kinetochore BUBs in control cells and of kinetochore LAP-Knl1 in Knl1-FL-expressing cells are set to 1.

co-precipitation with Knl1-NC compared with Knl1-FL (Figure S1d). The observation that KMN network members were present in roughly equal amounts in both pull-downs, and that Mps1 kinetochore localization was similar in cells expressing the Knl1 variants, further verified that KMN network integrity was unaffected in the various cell lines (Figure S1d-f [266]).

Knl1 depletion severely weakened the SAC: nocodazole-treated cells depleted of Knl1 rapidly exited mitosis when Mps1 kinase activity was slightly reduced with a low dose of reversine (250 nM [267,268]), whereas control cells maintained mitotic delays for many hours (Figure 2a, Figure S2a). Incomplete penetrance of RNAi or a nonessential role for Knl1 in the SAC can account for the

residual weak SAC response in Knl1-depleted cells (Figure S2b), and we were unable to distinguish between these possibilities because no residual kinetochore Knl1 or Bub1/BubR1 was detectable in siKNL cells. Regardless, the high sensitivity of nocodazole-treated, Knl1-depleted cells to low concentrations of reversine allowed us to examine functionality of Knl1 variants in the SAC. Somewhat unexpectedly, Knl1-NC was equally efficient as Knl1-FL in restoring SAC signaling to KNL-depleted cells (Figure 2a). In agreement with this, Knl1-NC was able to recruit significantly more Mad1 to kinetochores than Knl1-C (Figure S1g,h). Checkpoint activity of Knl1-NC depended on the MDLT and KI motifs (Figure S2c), indicating that Knl1-NC was able to recruit sufficient amounts of BUB proteins to perform SAC signaling. In support of this, SAC activity in Knl1-NC- but not Knl1-FL-expressing cells was highly sensitive to Bub1 levels (Figure 2b), and weak but detectable H2A-Thr120 phosphorylation (a mark that depends on Bub1 activity [254]) was restored on centromeric chromatin by Knl1-NC (Figure 2b). We thus conclude that Knl1-NC can support robust SAC function by recruiting low levels of BUB proteins to kinetochores.

Two observations indicated that unlike the SAC, chromosome biorientation was not efficiently restored in cells expressing Knl1-NC. First, Knl1-NC was unable to support chromosome alignment in cells that were prevented from exiting mitosis by addition of the proteasome inhibitor MG132 (Figure 2d). Second, Knl1-NC expression caused long mitotic delays, likely due to absence of proper kinetochore–microtubule attachment that prevents SAC silencing (Figure 2e, Figure S2d,e).

Together, these data indicate that the N-terminal MDLT-KI1-KI2 motifs in Knl1 function as an independent module that is capable of activating the SAC by recruiting low BUB levels to kinetochores, but is insufficient for proper chromosome biorientation.

Knl1 contains multiple independent BUB recruitment modules

Our analyses of KNL-NC function showed that the N-terminal MDLT-KI1-KI2 fragment of Knl1 recruited low amounts of Bub1 to kinetochores but was sufficient to maintain a robust SAC. To examine if the N-terminal fragment was also required for full-length Knl1 to promote SAC activity, we analyzed function of Knl1 carrying mutations in this fragment. Bub1 localization as well as SAC activity were indistinguishable in Knl1-depleted cells expressing Knl1 with mutations in the MDLT or KI motifs (Knl1-FL^{MDLA} or Knl1-FL^{KI1}) or lacking the module altogether (Knl1$^{\Delta 261}$, Figure S2f,g). In addition, cells expressing full-length Knl1 with mutations in either of the KI motifs (Knl1-FL^{KI1} or Knl1-FL^{KI2}) restored chromosome alignment as efficiently as wild-type Knl1 and progressed through an unperturbed mitosis with similar kinetics, even in a sensitized situation (Figure

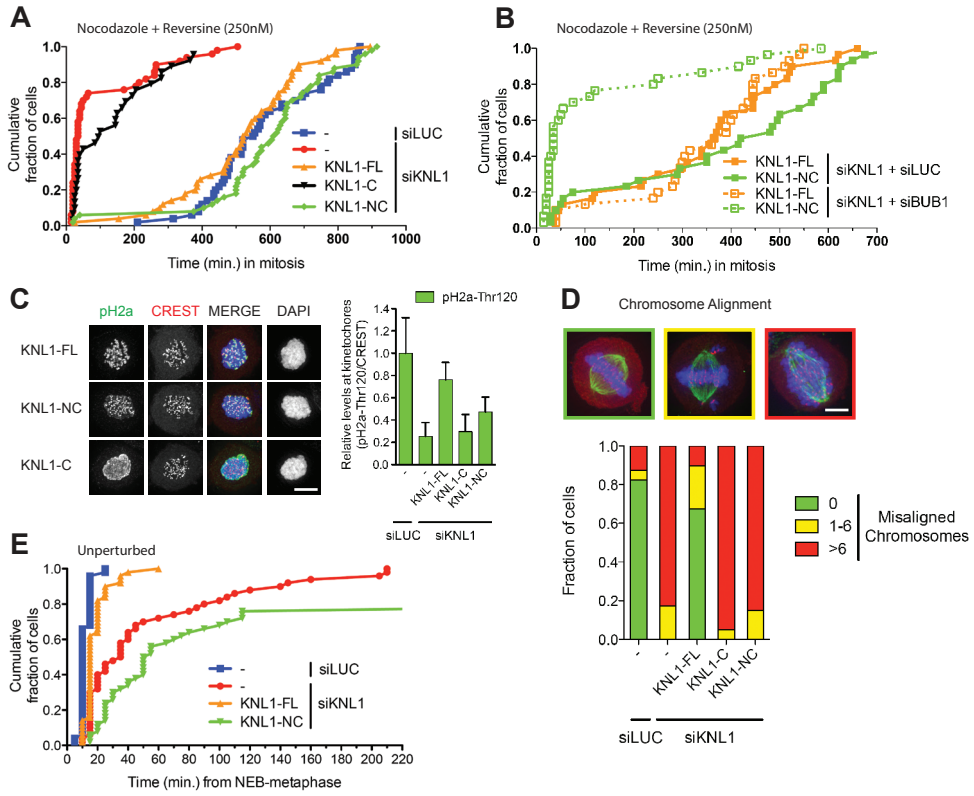


Figure 2 The N-terminal MDLT-KI module in Knl1 is sufficient to support SAC activity but not chromosome biorientation. (A) Time-lapse analysis of Flp-in HeLa cells expressing LAP-Knl1 variants, transfected with siLUC or siKnl1, and treated with nocodazole and 250 nM reversine. Data (n = 40 representative of 3 independent experiments) indicate cumulative fraction of cells that exit from mitosis (as scored by cell morphology using DIC) at the indicated time after NEB. **(B)** As in A, but with transfection of the indicated siRNAs. **(C)** Immunostaining and quantification of centromeric H2A-Thr120 phosphorylation in Flp-in HeLa cells expressing LAP-Knl1 variants and transfected with siLUC or siKnl1. pH2A-T120 is shown in green, centromeres (CREST) in red, and DNA (DAPI) in blue. Bars, 5 µm. pH2A-T120 is quantified over CREST (n = 10 representative of 3 independent experiments). **(D)** Immunostaining and quantification of chromosome alignment in Flp-in HeLa cells expressing LAP-Knl1 variants, transfected with siLUC or siKnl1, and treated with MG132 for 45 min. Tubulin is shown in green, centromeres (CREST) in red, and DNA (DAPI) in blue. Bars, 5 µm. The data shown are from a single representative experiment out of three repeats. For the experiment shown, n = 50. **(E)** Time-lapse analysis of Flp-in HeLa cells expressing LAP-Knl1 variants and transfected with siLUC or siKnl1. Data (n = 40 representative of 3 independent experiments) indicate cumulative fraction of cells that exit from mitosis at the indicated time after NEB (as scored by GFP-H2B).

S2h-j [252,259]). We thus conclude that, although able to bind BUBs (Figure 1a) and support SAC activity (Figure 2a), the N-terminal KI-containing module is dispensable for Knl1 function, at least in our assays. Most likely, therefore, Knl1 can recruit BUBs by alternative means.

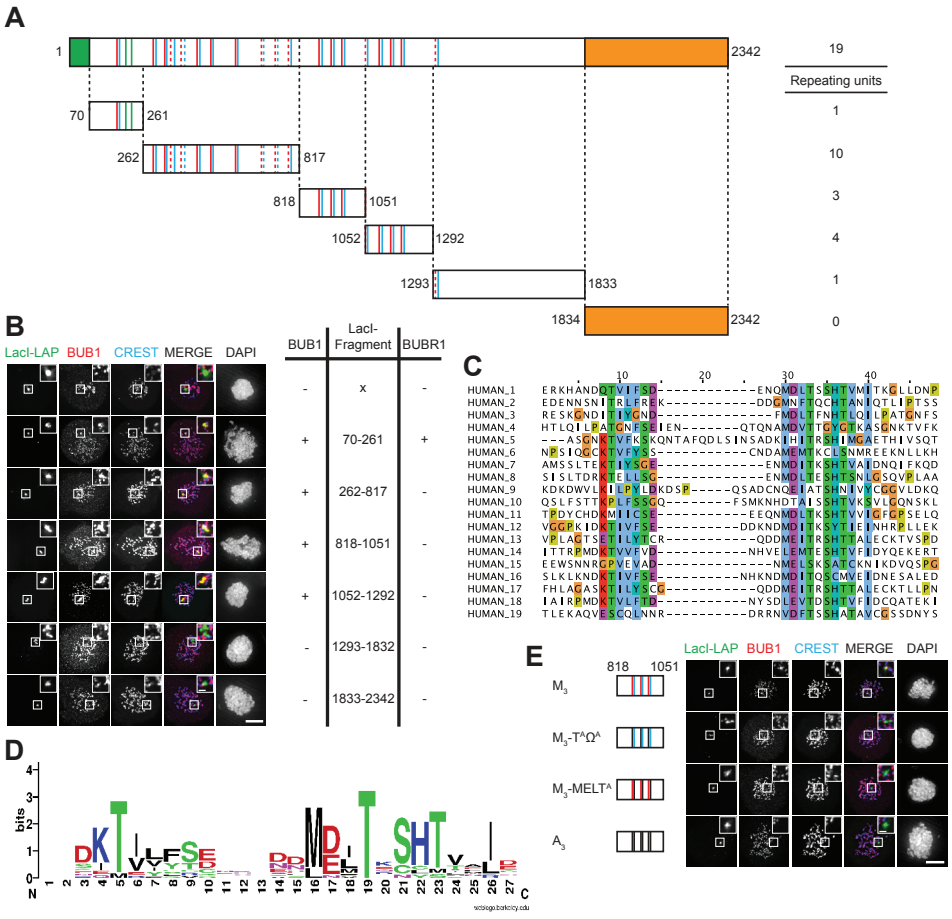


Figure 3 Knl1 contains multiple independent BUB recruitment modules.
(A) Schematic representation of Knl1 showing the microtubule- and PP1-binding domain in green and the kinetochore recruitment domain in orange. KI1 and KI2 motifs are shown as green bars, MELT-like sequences in red, and TΩ-like sequences in blue. Dashed lines indicate the generated LacI-LAP-Knl1 fragments used in B. **(B)** Immunolocalization of Bub1 (red) in nocodazole-treated U2OS-LacO cells transfected with LacI-LAP-Knl1 fragments. LacI-LAP-Knl1 fragments are shown in green, centromeres (CREST) in blue, and DNA (DAPI) in white. Insets show magnifications of the boxed regions. Bars, 5 µm (insets, 0.5 µm). Table indicates the ability (– or +) to recruit Bub1 and BubR1 by the indicated Knl1 fragments (see also Figure S3a,b). **(C)** Alignment of identified TΩ-MELT modules showing conserved (green/purple/red/blue) and atypical (orange/yellow) amino acids. **(D)** Sequence logo of the 19 TΩ-MELT units. **(E)** As in B, but with LacI-LAP-Knl1$^{818–1051}$ (M_3) or mutant variants thereof. These variants are: M_3-T^{A}Ω^A (TxxΩ to AxxA), M_3-MELT^A (MELT to MELA), and A_3 (TxxΩ-MELT to AxxA-AELA), as shown in Figure S3c.

To test whether additional regions in Knl1 could also function as independent BUB recruitment modules, we analyzed the ability of various Knl1 fragments to recruit BUBs to LacO arrays. The LacI-LAP-Knl1$^{70-261}$, the LacI-LAP fusions of Knl1$^{262-817}$, Knl1$^{818-1051}$, and Knl1$^{1052-1292}$ were sufficient to recruit Bub1 to the

LacO array, whereas LacI-LAP, LacI-LAP-Knl1$^{1293–1833}$, and LacI-LAP-Knl1-C were not (Figure 3a,b, Figure S3b). Interestingly, LacI-LAP-Knl1$^{70-261}$ was the only fragment that could recruit detectable amounts of BubR1 to the LacO array (Figure S3a,b). Using the repeat-finding algorithm MEME [218] we noticed that Knl1 consists of 19 repeating modules that include but are not limited to MELT-like sequences (Figure 3a,c). MELT-like sequences, when phosphorylated by Mps1, are thought to participate in BUB recruitment to unattached kinetochores [261–263]. The repeating modules uncovered by MEME consist of a MELT-like sequence flanked on the C-terminal side by SHT and on the N-terminal side by the sequence TΦΦΩ[ST][DE] (where Φ denotes a hydrophobic residue and Ω denotes F or Y), which we will refer to as "TΩ" motifs (Figure 3d). Although overall quite different in sequence, the TΩ motifs have resemblance to KI1, in which the threonine and phenylalanine in the TxxF sequence (KIDTTSFLA) directly interact with Bub1 and are indispensable for Knl1–Bub1 interaction [260]. For convenience, we will refer to these repeating modules as "TΩ-MELT". 10 modules adhere closely to the TΩ-MELT sequence (1, 4, 6, 8, 12–14, 16–18), whereas the remaining nine deviate to some degree in either the TΩ, the MELT, or both motifs (Figure 2c).

To study functionality of the TΩ-MELT motifs, we analyzed their contribution to the ability of the Knl1$^{818–1051}$ fragment to recruit Bub1. This fragment contains three TΩ-MELT repeat modules, and will be referred to as the M_3 fragment. A mutated version of this fragment, in which four amino acids in each TΩ-MELT module were substituted for alanine (TΩ-MELT to AA-AELA), will be referred to as A3. Bub1 recruitment to LacI-Knl1$^{818–1051}$ depended on TΩ-MELT motifs, as the A3 fragment was unable to localize Bub1 to the LacO arrays (Figure 3e, Figure S3c,d). Furthermore, the TΩ and the MELT motifs were each indispensable for the ability of LacI-Knl1$^{818–1051}$ to recruit Bub1 because mutating either TΩ or MELT abolished Bub1 localization (Figure 3e, Figure S3c,d). The TΩ motif was also critical for the N-terminal module that uniquely contains KI1 and KI2 (Figure S3e,f).

Engineered Knl1 proteins reveal differential requirements for TΩ-MELT modules in the SAC and chromosome biorientation

Our observations that the N-terminal module is sufficient but not required for the SAC, that this module is insufficient for proper chromosome biorientation, and that other modules in Knl1 highly resemble this N-terminal module raised the question of whether there is functional redundancy between modules or whether some modules have specialized. To examine this, we generated a Knl1 protein devoid of all TΩ-MELT–like modules but containing the N-terminal-most 86 amino acids (responsible for microtubule and PP1 binding) fused to Knl1-C [60,61,252]. Into this protein, named Knl1$^{\Delta}$, we inserted one or two

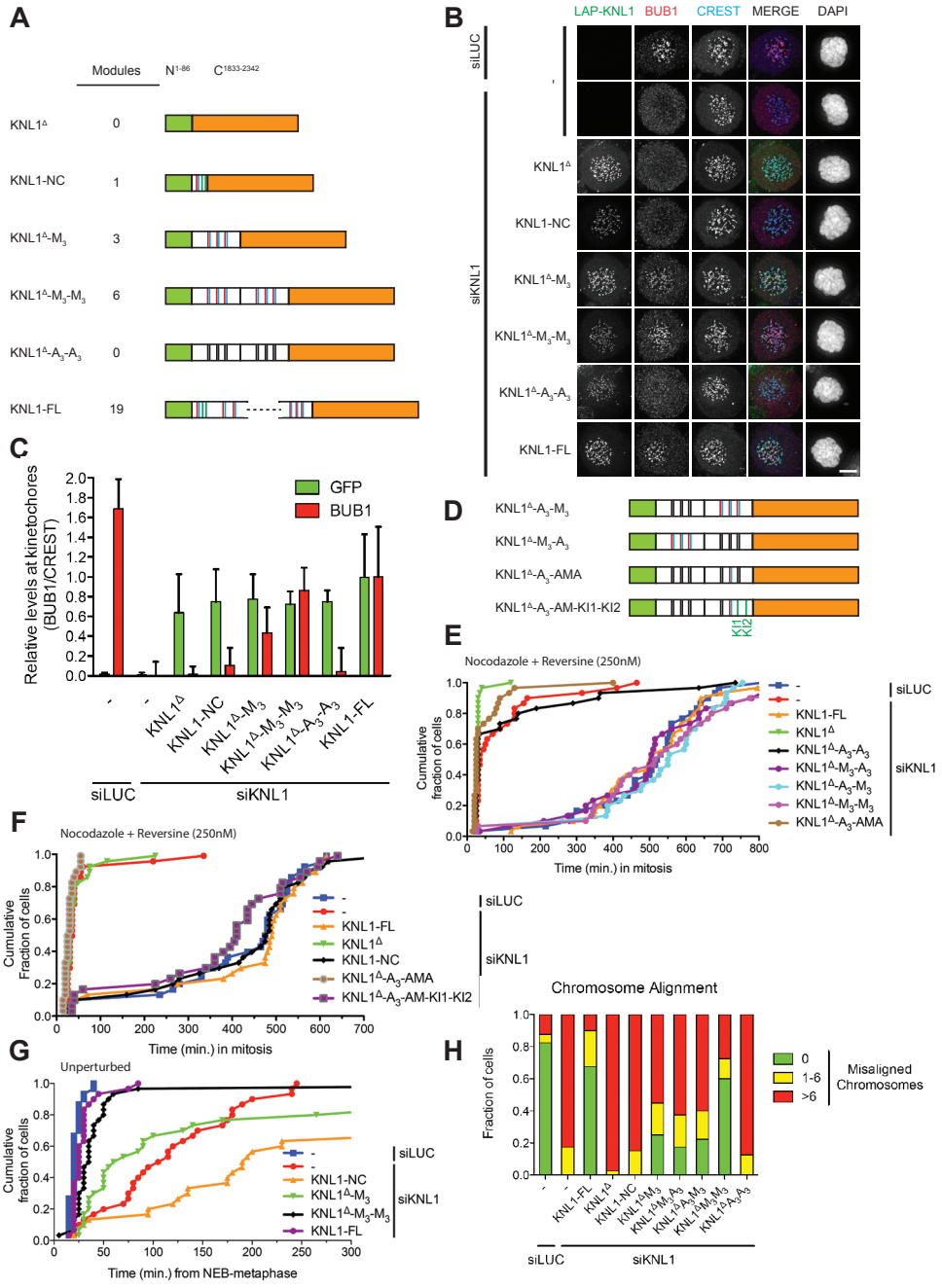


Figure 4 Engineered Knl1 proteins reveal differential requirements for TΩ-MELT modules in the SAC and chromosome biorientation. **(A)** Schematic representation of synthetic LAP-Knl1 constructs, showing the microtubule- and PP1-binding domain in green and the kinetochore recruitment domain in orange. KI1 and KI2 motifs are shown as green bars, MELT-like sequences in blue, and TxxΩ-like sequences in red. See main text for details about constructs. **(B,C)** Repre-

sentative images (B) and quantification (C) of LAP-Knl1-expressing Flp-in HeLa cells transfected with siRNAs to luciferase (siLUC) or to Knl1 (siKnl1) and treated with nocodazole. LAP-Knl1 is shown in green, Bub1 in red, centromeres (CREST) in blue, and DNA (DAPI) in white. Bars, 5 µm. Quantification in C shows total kinetochore signal intensity (+SD) of LAP-Knl1 and BUB proteins over CREST. Data are from >15 cells and representative of 3 experiments. Levels of kinetochore BUBs in control cells and of kinetochore LAP-Knl1 in Knl1-FL-expressing cells are set to 1. **(D)** Schematic as in A. See main text for details about constructs. **(E)** Time-lapse analysis of Flp-in HeLa cells expressing LAP-Knl1 variants, transfected with siLUC or siKnl1, and treated with nocodazole and 250 nM reversine. Data (n = 40 representative of 3 independent experiments) indicate cumulative fraction of cells that exit from mitosis (as scored by cell morphology using DIC) at the indicated time after NEB. **(F)** As in E, with the indicated constructs. **(G)** Time-lapse analysis of Flp-in HeLa cells expressing LAP-Knl1 variants and transfected with siLUC or siKnl1. Data (n = 40 representative of 3 independent experiments) indicate cumulative fraction of cells that exit from mitosis at the indicated time after NEB (as scored by GFP-H2B). **(H)** Quantification of chromosome alignment in Flp-in HeLa cells expressing LAP-Knl1 variants, transfected with siLUC or siKnl1, and treated with MG132 for 45 min. The data shown are from a single representative experiment out of three repeats. For the experiment shown, n = 40.

of the M_3 fragments to create Knl1$^{\Delta}$-M_3 and Knl1$^{\Delta}$-M_3-M_3, respectively (Figure 4a). A_3 fragments were used as control, as well as combinations of M_3 and A_3 fragments, giving rise to Knl1$^{\Delta}$-A_3, Knl1$^{\Delta}$-A_3-A_3, Knl1$^{\Delta}$-A_3-M_3, and Knl1$^{\Delta}$-M_3-A_3 (Figure 3e, Figure 4a,d). Isogenic cell lines with inducible expression of these engineered KNL proteins were generated and analyzed for functionality of various processes upon depletion of endogenous Knl1. Strikingly, the amount of Bub1 detectable at unattached kinetochores followed the amount of repeat modules present in Knl1: a single module (Knl1-NC) recruited low amounts of Bub1, one block of three modules (Knl1$^{\Delta}$-M_3) recruited approximately threefold more Bub1, and two blocks totaling six modules (Knl1$^{\Delta}$-M_3-M_3) doubled that to close to the levels observed in KNL-FL reconstituted cells (Figure 4b,c). Absence of any module (Knl1$^{\Delta}$ or Knl1$^{\Delta}$-A_3-A_3) eliminated Bub1 kinetochore binding. These data are indicative of a direct correlation between the number of functional TΩ-MELT modules and the amount of Bub1 protein at mitotic kinetochores. Interestingly, although BubR1 did not interact with amino acids 818–1051 in Knl1 in the context of the LacO array (Figure S3a,b), Knl1$^{\Delta}$-M_3-M_3 was able to recruit BubR1 to kinetochores (Figure S4a,b). This suggested that that BubR1 recruitment to Knl1 requires the context of kinetochores.

All Knl1 variants that contained at least one M_3 fragment (M_3, M_3-M_3, A_3-M_3, and M_3-A_3) were proficient in recruiting MAD1 and supporting the SAC (Figure S1g,h, Figure 4e). Because SAC activity was also supported by a single module in the context of the N-terminal fragment (Knl1-NC, see Figure 2a), we next examined whether any single module could support the SAC. To this end, one TΩ-MELT module was restored in Knl1$^{\Delta}$-A_3-A_3, creating the Knl1$^{\Delta}$-A_3-AMA protein (Figure 4d). Although able to recruit low levels of kinetochore Bub1 and promote partial H2A-T120 phosphorylation (Figure S4c-e), Knl1$^{\Delta}$-A_3-AMA

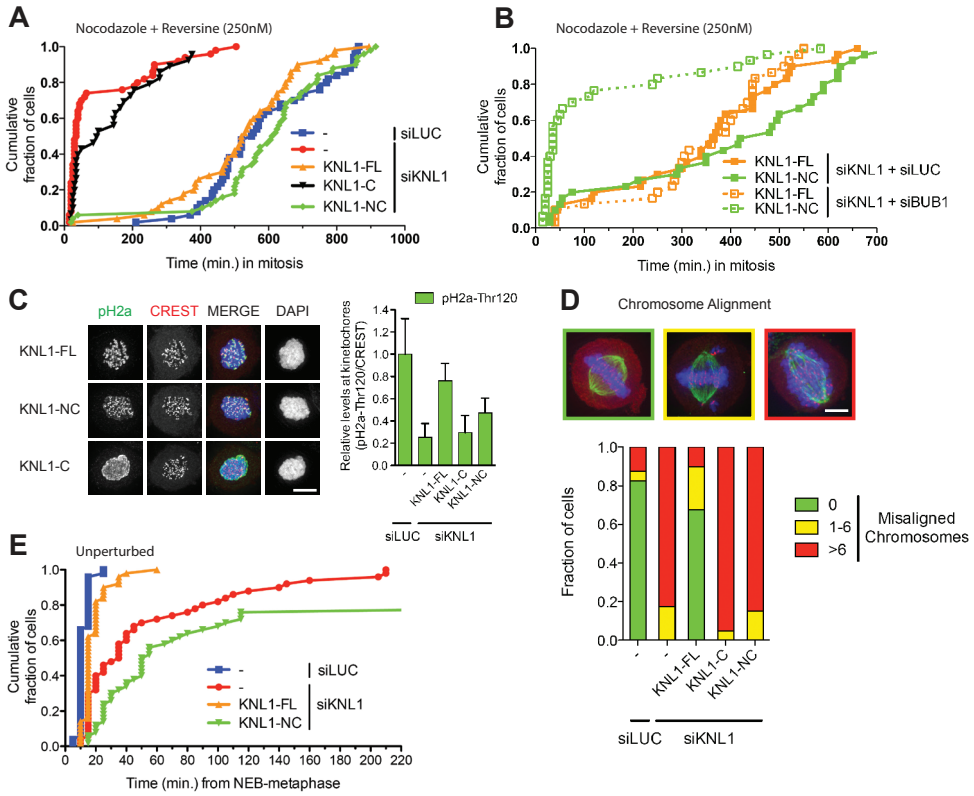


Figure 5 TΩ-MELT modules in Knl1 are redundant and exchangeable. (A) Schematic representation of synthetic LAP-Knl1 constructs. For color codes, see Figure 4a. See main text for details about constructs. **(B,C)** Representative images **(B)** and quantification **(C)** of LAP-Knl1–expressing Flp-in HeLa cells transfected with siRNAs to luciferase (siLUC) or to Knl1 (siKnl1) and treated with nocodazole. LAP-Knl1 is shown in green, Bub1 in red, centromeres (CREST) in blue, and DNA (DAPI) in white. Bars, 5 µm. Quantification in C shows total kinetochore signal intensity (+SD) of LAP-Knl1 and BUB proteins over CREST. Data are from >15 cells and representative of 3 experiments. Levels of kinetochore BUBs in control cells and of kinetochore LAP-Knl1 in Knl1-FL-expressing cells are set to 1. **(D)** Quantification of chromosome alignment in Flp-In HeLa cells expressing LAP-Knl1 variants, transfected with siLUC or siKnl1, and treated with MG132 for 45 min. The data shown are from a single representative experiment out of three repeats. For the experiment shown, n = 40. **(E)** Time-lapse analysis of Flp-in HeLa cells expressing LAP-Knl1 variants, transfected with siLUC or siKnl1, and treated with nocodazole and 250 nM reversine. Data (n = 40 representative of 3 independent experiments) indicate cumulative fraction of cells that exit from mitosis (as scored by cell morphology using DIC) at the indicated time after NEB.

could not recover SAC activity (Figure 4e). We thus conclude that a single module recruits sufficient Bub1 for SAC activation only in the context of the N-terminal fragment, whereas more than one is needed in the context of other Knl1 fragments.

The N-terminal BUB recruitment module is unique in two ways: it is close to the PP1- and microtubule-binding site on Knl1, and it contains the KI motifs that in the context of Knl1-NC significantly contribute to BUB recruitment and SAC activity (Figure 1a, Figure S2c). To examine if either of these is the cause of the difference in ability of Knl1-NC and Knl1$^{\Delta}$-A$_3$-AMA to support SAC activity, we placed KI1 and KI2 downstream of the TΩ-MELT module in Knl1$^{\Delta}$-A$_3$-AMA (resulting in Knl1$^{\Delta}$-A$_3$-AM-KI1-KI2, see Figure 4d). Strikingly, adding KI1 and KI2 to Knl1$^{\Delta}$-A$_3$-AMA endowed the protein with SAC function (Figure 4f) and this correlated with a slight increase in kinetochore Bub1 to close to the levels attained by Knl1-NC (Figure S4f,g). These data therefore indicate that the KI motifs enhance BUB recruitment potential of individual TΩ-MELT modules, and as such allow the N-terminal module to be sufficient for SAC function.

Time-lapse imaging of mitotic progression in the different cell lines showed that increasing amounts of repeat modules gradually decreased the time from nuclear envelope breakdown (NEB) to metaphase (Figure 4g). This corresponded to increased efficiency of chromosome alignment, as assayed in fixed MG132-treated mitotic cells (Figure 4h). Directly in line with Bub1 levels, three modules were more efficient than one, whereas six modules were more efficient than three. In fact, cells expressing Knl1$^{\Delta}$-M$_3$-M$_3$ were almost as efficient in chromosome alignment as control cells or cells expressing Knl1-FL and displayed comparable mitotic timing (Figure 4g,h). High fidelity chromosome segregation in human cells therefore requires between four and six TΩ-MELT modules that combine to recruit sufficient BUBs.

Functional TΩ-MELT modules in Knl1 are redundant and exchangeable

Our observations with the engineered Knl1 proteins suggested that the modules within the M$_3$ fragment may be redundant and that their functionality in SAC activity is independent of exact position in relation to the kinetochore or to the microtubule- and PP1-binding sites. This raised the possibility that redundancy is relatively widespread across the 19 identified repeat modules. To examine this, we designed artificial fragments, based on existing TΩ-MELT modules, and tested their functionality in the context of Knl1$^{\Delta}$. We swapped module 12, 13, and 14 within the M3 fragment for either module 2 or module 17 to create Knl1$^{\Delta}$-2$_3$-2$_3$ or Knl1$^{\Delta}$-17$_3$-17$_3$, respectively (Figure 3c, Figure 5a). We chose module 17 because its sequence adheres closely to the "consensus" TΩ-MELT module sequence (TILYSCGQDDMEITRSHTTAL), and module 2 was chosen because its sequence deviates from that consensus (TRLFREKDDGMNFTQCHTANI) but maintains the TxxΩ and MxxT characteristics (Figure 3c-d, Figure 5a). Knl1$^{\Delta}$-17$_3$-17$_3$ fully restored BUB localization, chromosome alignment, and SAC activity in Knl1-depleted cells (Figure 5b-e, Figure S5a). Interestingly, Knl1$^{\Delta}$-2$_3$-2$_3$ could not support chromosome alignment and SAC activity, which

correlated with low levels of Bub1/BubR1 recruitment to kinetochores and incomplete restoration of centromeric pH2A-T120 (Figure 5b-e, Figure S5b). We therefore conclude that neither chromosome biorientation nor the SAC relies on any specific TΩ-MELT module but that both processes require any combination of modules that can recruit sufficient Bub1. We thus propose that different TΩ-MELT modules have redundant functions and that any array of functional modules that can recruit sufficient Bub1 will promote high fidelity chromosome segregation.

Extensive divergence in sequence and amount of repeat modules in eukaryotic Knl1 homologues

Our findings suggest that human Knl1 has evolved by extensive duplications of the TΩ-MELT modules, possibly followed by degeneration of a number of these sequences. Furthermore, our recent analysis of selected eukaryotic Knl1 homologs [167] showed that the amount of MELT-like sequences varies quite extensively from species to species. To examine if repeating modules exist in these and other Knl1-like sequences, we applied MEME on predicted Knl1 homologues from 15 species across three supergroups of eukaryotic evolution (Figure 6a). Predicted homologs were identified by similarity in the C-terminal coiled-coil region and homology was further strengthened by the presence of an N-terminal PP1-binding RVSF motif. Interestingly, all homologues contained repeating modules, but they diverged extensively in sequence and number. The methionine of the MELT motif is conserved in most species, but the "LT" sequence is often replaced by additional negative charges. A striking example of this are the drosophilids in which the repeating module is based around a MEED-like sequence (Figure 6a) [269]. TΩ-like sequences were apparent in Knl1 homologues of *Branchiostoma floridae* and *Crassostrea gigas*, but MELT-like sequences of most other organisms were complemented with different motifs. In some species (*Nematostella vectensis, Thecamonas trahens*), Knl1 homologues contained two different types of repeating modules. We conclude that Knl1 is a rapidly evolving protein, with extensive variations in the number and sequence of repeating modules across different eukaryotic Knl1 homologs.

Discussion

An extensive array of generic BUB recruitment modules in Knl1

Our data demonstrate that Knl1 is a scaffold that contains multiple independent and redundant repeating modules, which together ensure recruitment of sufficient amounts of BUB proteins to kinetochores (Figure 6b). The ability of Knl1 to recruit BUB proteins and ensure efficient chromosome alignment is independent of protein length, of localization of the recruitment modules within Knl1, and of any particular module, per se: Knl1 function is maintained when

only two copies of modules 12-13-14 or six copies of module 17 are present. Moreover, compared with the SAC, more modules seem required for chromosome alignment, and the efficiency of chromosome alignment directly follows the amount of functional modules present in Knl1, suggesting that the modules act in an additive fashion (Figure 6b).

Mutational analysis of the N-terminal module shows that the TxxΩ motif that we identified as part of the repeating module is critical for BUB recruitment. Previous structural work has shown direct interactions between the TPR domain of Bub1 and the TxxF sequence in the KI1 motif of Knl1 [260]. Considering that there is only one potential TxxF interaction groove within the TPR domain of Bub1, we hypothesize that one functional TΩ-MELT module is capable of recruiting one Bub1 molecule. The contribution of the MELT-like sequences was recently described, and involves the Bub1-interacting protein Bub3 [131].

Unlike Bub1, all LacO-targeted Knl1 fragments except for Knl170–261 failed to recruit BubR1. This may be related to a difference by which the BUBs interact with Knl1. Whereas the TxxF motif in KI1 is critical for interaction with Bub1, a similarly positioned FxxF motif in KI2 is critical for interaction with BubR1, and neither motif can substitute for loss of the other. The repeating modules present in the Knl1 fragments all contain TxxΩ or variants thereof, but never an aromatic residue in the T position. Interestingly, however, a Knl1 fragment that was unable to recruit BubR1 to LacO arrays was able to recruit BubR1 to kinetochores. In fact, Knl1$^{\Delta}$-M$_3$-M$_3$ restored BubR1 kinetochore levels to the same extent as Knl1-FL. BubR1 binding to Knl1 under these conditions is therefore likely indirect and requires one or more kinetochore-localized proteins or activities. Because Bub1 is normally indispensable for BubR1 localization [257,270] and KI2 is not (this study), and because mutations in the TΩ-MELT motifs abolished BubR1 localization, we hypothesize that the predominant mode of BubR1 kinetochore binding is indirect via TΩ-MELT–mediated Knl1–Bub1 interaction, aided by an unidentified kinetochore-localized activity.

In contrast to the TΩ-MELT modules, the role of the N-terminal KI1 and KI2 modules in Knl1 function is unclear. Our recent bioinformatics analysis has indicated that the KI motifs are a recent invention of the vertebrate lineage [167]. Furthermore, the interaction of KI1 with the TPR domain of Bub1 is dispensable for Bub1 recruitment to kinetochores [260], and we show here that in the context of the full-length protein and with our assays, KI1 and KI2 are not required for SAC activity, chromosome alignment, and mitotic progression. These observations raise the question of what the functionality of the KI motifs is. Both within the N-terminal module and the synthetic Knl1$^{\Delta}$-A$_3$-AMA-KI1-KI2 construct, the KI motifs enhance BUB recruitment potential of the TΩ-MELT motifs to levels that support SAC activity. It is therefore likely that the KI motifs

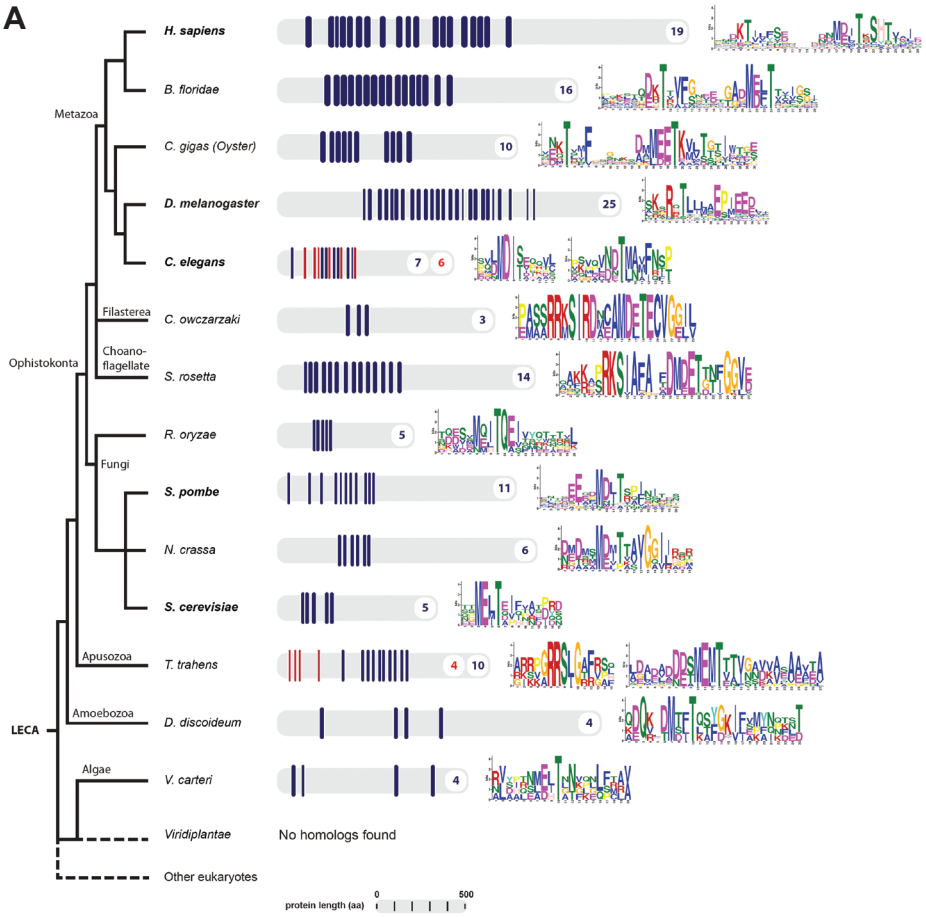

A
H. sapiens
B. floridae
C. gigas (Oyster)
D. melanogaster
C. elegans
C. owczarzaki
S. rosetta
R. oryzae
S. pombe
N. crassa
S. cerevisiae
T. trahens
D. discoideum
V. carteri
Viridiplantae
Other eukaryotes
No homologs found
Metazoa
Filasterea
Choano-flagellate
Ophistokonta
Fungi
Apusozoa
Amoebozoa
Algae
LECA
protein length (aa)
0
500
B
BUBS
KNL1
PP1 & MT-binding site
duplication
KT-recruitment domain
DKTxxΩ.....M[DE][ILVM][ST]xSHT
# of Bub-recruitment modules
p
H2a
MC
KT

Figure 6 TΩ-MELT module evolution and model. (A) Schematic representation of eukaryotic tree of life showing Knl1 homologues from indicated species. Repeating units are shown in blue and red with the number of repeats in corresponding colors. Repeat sequences are shown as sequence logos. **(B)** Model for TΩ-MELT function in human Knl1. Conserved (dark blue) and degenerated (light blue) TΩ-MELT modules (essential amino acids in red) in Knl1 can independently recruit BUB protein complexes (BUBs) to promote H2A-T120 phosphorylation and SAC activity (few modules, low BUB levels) and chromosome biorientation (increasing fidelity with increasing modules and BUB levels).

contribute to some extent to the BUB recruitment ability of full-length Knl1. This may become beneficial under conditions that require maximal BUB levels at kinetochores.

TΩ-MELT module function

Knl1$^{\Delta}$-17$_3$-17$_3$ and Knl1Δ-M3-M3 contain six recruitment modules, yet were able to recruit roughly the same amount of Bub1 to kinetochores as Knl1-FL with its 19 modules. One possible explanation for why Knl1-FL does not recruit more Bub1 is that not all modules are functional in Knl1-FL. Consistent with this, our analysis of Knl1$^{\Delta}$-2$_3$-2$_3$ showed that module 2 is less capable of binding Bub1 than modules 12, 13, 14, and 17. Module 2 contains the motif TF-MN-FT with relatively significant substitutions within the MELT-like motif. Besides module 2, modules 3, 5, 7, 9, 10, 11, 15, and 19, and to a lesser extent module 8, have alterations in either the TΩ and/or the MELT-like motifs, possibly rendering them less or not functional. In addition to sequence, phosphorylation of the motifs also likely contributes to BUB-binding affinity. Some TΩ- (16/18) and MELT-like (12/15/16/17/18) sequences can be phosphorylated by Mps1 in vitro [263], and one was found phosphorylated in mitotic cells (7: MDIp-TKSHpT [bold/underlined letters represent phosphorylated residues] [271]). A Knl1-8A mutant in which all the in vitro phosphorylation sites were mutated to alanine reduced Bub1 kinetochore localization by ~50% [263], showing that TΩ-MELT phosphorylation enhances BUB recruitment. Non-phosphorylatable TΩ- (11/15/19) and MELT-like (9/10) sequences are therefore likely to be less functional than phosphorylatable ones. This additionally raises the question of how many functional modules are phosphorylated at any given moment on one Knl1 molecule on an unattached kinetochore. It is conceivable that expanding the amount of functional modules simply increases the chance that a certain, small number of modules is phosphorylated at steady state, and that the actual amount of Knl1-bound BUBs required for K-MT attachment and the SAC is lower than the amount of modules that we have engineered into Knl1. A systematic biochemical survey of TΩ-MELT functionality and phosphorylation, combined with cell biological analyses will be required to elucidate which TΩ-MELT modules are functional and how they contribute to BUB recruitment.

TΩ-MELT module evolution

Our present and past surveys of eukaryotic homologs of Knl1 have revealed striking differences between species [167]. Most homologs contain an array of repeating modules that is unique to Knl1, but the number and sequence of those modules varies quite extensively. It will be interesting to examine whether BUB–Knl1 interactions in different species require the species-specific repeat module characteristics, or whether these additional motifs contribute to other, unknown module functionality. More in-depth analysis has provided evidence of rapid evolution of the modules in eukaryotes (unpublished data). This, combined with the conserved roles for BUBs in chromosome segregation and our demonstration that the modules in human Knl1 are generic in nature may thus indicate that the extensive species-specific differences in module sequence may not affect BUB binding per se, but may reflect other evolutionary important roles for the modules. Assaying function in human cells of Knl1 containing modules of other species might start to provide some answers to these questions.

Besides sequence, the number of modules per Knl1 homologue also differs strongly. Green algae like *Volvox carteri* have Knl1 homologues with only a few modules, whereas those of species like *Drosophila melanogaster* and *Xenopus tropicalis* have more than 20 (Figure 6a [167]). Possibly, the amount of modules correlates with the amount of BUB signaling required for high fidelity chromosome segregation. Phosphorylation of H2A-T120 is significantly restored with a single module in human cells, and this role of Bub1 in chromosome segregation is conserved also in more primitive species [254]. Perhaps, therefore, H2A phosphorylation and SAC activity require only one or a few modules and this allows more primitive species to survive with few modules in Knl1. More challenging requirements in mitosis for the more complex organisms (for instance, expanding complexity of kinetochores and increasing numbers of microtubules bound per kinetochore) may thus have spurred multiplication of modules to enable recruitment of more BUBs to kinetochores. An exciting possibility therefore is that altering BUB signaling by module expansion and degeneration during evolution is a relatively facile mechanism for adapting the chromosome segregation machinery to changing requirements during mitosis.

Materials and methods

Plasmids

pcDNA5-LAP-KnlFL encodes full-length, N-terminally LAP-tagged, and siRNA-1-resistant wild-type Knl1 (modified codons 258 and 259) and was created by digestion of pEYFP-LAP-Knl1^{FL} (a gift from I. Cheeseman, Whitehead Institute, Cambridge, MA) with XhoI and HpaI to isolate the full-length Knl1 cassette,

which was ligated into the XhoI and PmeI sites of pcDNA5/FRT/TO (Invitrogen). An N-terminal LAP-tag was introduced by subcloning the LAP-tag cassette from pCDNA3-LAP-Mps1$^{\Delta}$200 (Nijenhuis et al., 2013) into the KpnI and XhoI sites of the resulting plasmid. Knl1-NC was generated by PCR and subcloning of Knl1-C (aa 1833–2342, using Xho1–BamH1) and Knl1-N (aa 1–261, using Xho1–Xho1) and subsequent ligation into pcDNA5/FRT/TO-LAP. LacI-Knl1 fragments were generated by PCR and cloned into pLacI-LAP. MELT-block aa 818–1051 and aa 1052–1228 and corresponding variants were synthesized by GenScript and cloned into the Xho1 site of Knl1$^{\Delta}$ (GenScript) using Sal1 and Xho1. Additional blocks were inserted in the Xho1 site of Knl1-NM/A3C.

Cell culture and transfection

U2OS LacO cells (a gift from S. Janicki, The Wistar Institute, Philadelphia, PA) were grown in DMEM supplemented with 8% FBS (Takara Bio Inc.), 200 µg/ml hygromycin, 50 µg/ml penicillin/streptomycin, and 2 mM l-glutamine. HeLa Flp-in cells were grown in 8% Tet-approved FBS (Takara Bio Inc.) supplemented with hygromycin (200 µg ml $^{-1}$) and blasticidin (4 µg ml $^{-1}$). Plasmids were transfected using FuGENE HD (Roche) according to the manufacturer's instructions. To generate stably integrated HeLa Flp-In cells, pcDNA5 constructs were cotransfected with Ogg44 recombinase in a 10:1 ratio [257]. Constructs were expressed by addition of doxycycline (1 µg ml $^{-1}$) for 24 hours. siKnl1 (CASC5#5, J-015673-05, Thermo Fisher Scientific; 5′-GCAUGUAUCUCUUAAGGAA-3′) and siBub1 (5′-GAAUGUAAGCGUUCACGAA-3′) were transfected using HiPerFect (QIAGEN) at 20 nM for 2 days according to the manufacturer's instructions.

Live-cell imaging

For live-cell imaging experiments, cells were transfected with 20 nM siRNA for 24 h, after which cells were arrested in early S-phase for 24 h by addition of 2 mM thymidine, and expression was induced by addition of 1 µg/ml doxycycline. Subsequently, cells were released from thymidine for 8–10 h and arrested in prometaphase by the addition of 830 nM nocodazole with or without 250 nM reversine. Unperturbed mitotic progression was assayed after a 24-h infection with BacMam-H2B-GFP virus (BioTek) followed by a release from thymidine into normal media. Cells were imaged in a heated chamber (37°C and 5% CO_2) using a 20×/0.5 NA UPLFLN objective on a microscope (model IX-81, Olympus) controlled by Cell-M software (Olympus). Images were acquired using a CCD camera (ORCA-ER, Hamamatsu Photonics) and processed using Cell-M software.

Immunofluorescence and antibodies

Asynchronously growing cells were arrested in prometaphase by the addition

of 830 nM nocodazole for 2–3 h. Cells plated on 12-mm coverslips were fixed (with 3.7% paraformaldehyde, 0.1% Triton X-100, 100 mM Pipes, pH 6.8, 1 mM $MgCl_2$, and 5 mM EGTA) for 5–10 min. Coverslips were washed with PBS and blocked with 3% BSA in PBS for 1 h, incubated with primary antibodies for 16 h at 4°C, washed with PBS containing 0.1% Triton X-100, and incubated with secondary antibodies for an additional hour at room temperature. Coverslips were then washed, incubated with DAPI for 2 min, and mounted using ProLong Antifade (Molecular Probes). All images were acquired on a deconvolution system (DeltaVision RT, Applied Precision) with a 100×/1.40 NA U Plan S Apochromat objective (Olympus) using softWoRx software (Applied Precision). Images are maximum intensity projections of deconvolved stacks. For quantification of immunostainings, all images of similarly stained experiments were acquired with identical illumination settings; cells expressing comparable levels of exogenous protein were selected for analysis and analyzed using ImageJ (National Institutes of Health). An ImageJ macro was used to threshold and select all centromeres and all chromosome areas (excluding centromeres) using the DAPI and anticentromere antibody channels as described previously [268]. This was used to calculate the relative mean kinetochore intensity of various proteins ([centromeres–chromosome arm intensity (test protein)] / [centromeres–chromosome arm intensity (CREST)]).

Cells were stained using GFP-booster (ChromoTek), Bub1 (Bethyl Laboratories, Inc.), BubR1 (Bethyl Laboratories, Inc.), H2A-pT120 (ActiveMotif), Mps1 (EMD Millipore), Mad1 (a gift from A. Musacchio, MPI, Dortmund, Germany), CREST/anti-centromere antibodies (Cortex Biochem, Inc.), and/or tubulin (Sigma-Aldrich). Secondary antibodies were goat anti–human Alexa Fluor 647 and goat anti–rabbit and anti–mouse Alexa Fluor 568 (Molecular Probes) for immunofluorescence experiments.

SILAC mass spectrometry

For SILAC mass spectrometry, LAP-Knl1-FL and -NC cells were adapted to light (C12N14 lysine/arginine) and heavy (C13N15 lysine/arginine) medium, respectively. Cells were synchronized in mitosis by a 24-h thymidine block, followed by overnight treatment with nocodazole. Knl1 expression was induced for 24 h using doxycycline and cells were harvested followed by immunoprecipitation and mass spectrometry. Cells were lysed at 4°C in hypertonic lysis buffer (500 mM NaCl, 50 mM Tris-HCl, pH 7.6, 0.1% sodium deoxycholate, and 1 mM DTT) including phosphatase inhibitors (1 mM sodium orthovanadate, 5 mM sodium fluoride, and 1 mM β-glycerophosphate), sonicated, and LAP-Knl1 proteins were coupled to GFP-trap (ChromoTek) for 1 h at 4°C. Purifications were washed three times with high-salt (2 M NaCl, 50 mM Tris-HCl, pH 7.6, 0.1% sodium deoxycholate, and 1 mM DTT) and low-salt wash buffers (50

mM NaCl, 50 mM Tris-HCl, pH 7.6, and 1 mM DTT) and subsequently eluted in 2 M urea, 50 mM Tris-HCL, pH 7.6, and 5 mM IAA. Samples were loaded on a C18 column and run on a nano-LC system coupled to a mass spectrometer (LTQ-Orbitrap Velos; Thermo Fisher Scientific) via a nanoscale LC interface (Thermo Fisher Scientific), as described in a previous study from our lab [147].

Repeat identification

For all Knl1 orthologs used in this study, separate MEME [218] analyses (option "any number of repeats") were performed to detect repeating motifs for which HMMR3 [235] profiles were created. Significant motifs were added to the profiles and searches were repeated until no novel repeats were found. Searches were manually inspected for consistency and significance; clear false-positives were discarded. The resulting repeats were aligned by hand and the alignment was used to construct sequence logos using WebLogo [247], which is embedded in the MEME package. Due to the degenerate nature for some the repeats, we introduced multiple gaps in the alignments. Therefore the sequence logos do not fully reflect the true spacing for the conserved amino acid positions within the repeats.

Contributions

GJPLK conceived the study and managed the project. MV performed the experimental research with help and/or suggestions of ET, MO, VG and WN. ET performed the evolutionary analyses. MV and GJPLK analyzed the data and wrote the manuscript with input from ET and BS.

Acknowledgments

We thank I Cheeseman, S Taylor, and S Janicki for providing reagents, A Musacchio and J Nilsson for sharing unpublished observations, A Schram for critical reading of the manuscript and the Kops and Lens laboratories for discussions. This work was supported by grants from the Netherlands Organization for Scientific Research (NWO-Vici 016.130.661), from the European Research Council (ERC-StG KINSIGN), and from KWF Kankerbestrijding (UU-2012-5427) to GJPL Kops.

Electronic Supplementary Material

Supplementary materials for this chapter are made available online at the following link http://bioinformatics.bio.uu.nl/eelco/thesis/

Figure S1 shows a quantification of the LacO-targeted Knl1 fragments, a sche-

matic of Knl1-NC, comparative mass spectrometry data of Knl1-FL vs. Knl1-NC, and Mps1 and Mad1 kinetochore localization (and quantification thereof) in cells expressing the different Knl1 constructs.

Figure S2 shows representative still images of checkpoint assays, quantification of checkpoint assays without sensitization and in cells expressing KNL-NC mutants, unperturbed metaphase–anaphase timing, representative still images of unperturbed mitosis for the different Knl1 constructs, and immunofluorescence of Bub1 localization (sensitized), mitotic progression, alignment, and checkpoint assays of KI mutant Knl1 constructs.

Figure S3 shows BubR1 recruitment to indicated LacO-targeted Knl1 fragments and quantification of Bub1 and BubR1 recruitment to these loci, an overview of the described M_3 mutations and a quantification of Bub1 recruitment to LacO-targeted M_3 blocks, and immunofluorescence and quantification of LacI-Knl1-N-containing TΩ mutations.

Figure S4 shows immunofluorescence and quantification of BubR1 recruitment to Knl1$^{\Delta}$-M_3-M_3 and immunofluorescence and quantification of pH2A-T120 and Bub1 kinetochore localization in Knl1$^{\Delta}$-M_3-AMA and Bub1 in Knl1$^{\Delta}$-M_3-AM-KI1-KI2 cells.

Figure S5 shows immunofluorescence of BubR1 and pH2A-T120 kinetochore localization in Knl1$^{\Delta}$-2_3-2_3 and Knl1$^{\Delta}$-17_3-17_3 cells.

CHAPTER 5

Widespread recurrent patterns of rapid repeat evolution in the kinetochore scaffold KNL1

Eelco Tromer, Berend Snel & Geert JPL Kops

Genome biology and evolution, 2015

Abstract

The outer kinetochore protein scaffold Knl1 is essential for error-free chromosome segregation during mitosis and meiosis. A critical feature of Knl1 is an array of repeats containing MELT-like motifs. When phosphorylated, these motifs form docking sites for the Bub1–Bub3 dimer that regulates chromosome biorientation and the spindle assembly checkpoint. Knl1 homologs are strikingly different in both the amount and sequence of repeats they harbor. We used sensitive repeat discovery and evolutionary reconstruction to show that the Knl1 repeat arrays have undergone extensive, often species-specific array reorganization through iterative cycles of higher order multiplication in conjunction with rapid sequence diversification. The number of repeats per array ranges from none in flowering plants up to approximately 35–40 in drosophilids. Remarkably, closely related drosophilid species have independently expanded specific repeats, indicating near complete array replacement after only approximately 25–40 Myr of evolution. We further show that repeat sequences were altered by the parallel emergence/loss of various short linear motifs, including phosphosites, which supplement the MELT-like motif, signifying modular repeat evolution. These observations point to widespread recurrent episodes of concerted Knl1 repeat evolution in all eukaryotic supergroups. We discuss our findings in the light of the conserved function of Knl1 repeats in localizing the Bub1–Bub3 dimer and its role in chromosome segregation.

Introduction

Mitotic chromosome segregation in eukaryotes involves the capture and stable attachment of the plus ends of spindle microtubules by all chromosomes in a manner that connects sister chromatids to opposing spindle poles. Large multiprotein assemblies on centromeric DNA, known as kinetochores, facilitate such chromosome–spindle interactions [163]. In addition to providing a link between DNA and the spindle, kinetochores are the signaling hubs for the spindle assembly checkpoint (SAC) and the target of attachment-error correction mechanisms [20,163,272]. The interplay between microtubule attachment, error-correction, and SAC signaling is centered on the KMN network (KNL1-C, MIS12-C, and NDC80-C), an outer-kinetochore multiprotein complex that forms the microtubule-binding interface of kinetochores [20,250]. The focal point of this interplay is KNL1/CASC5/AF15q14/Blinkin (hereafter referred to as Knl1), a largely disordered protein that recruits various mitotic regulators to the kinetochore and is able to directly interact with microtubules [59,61] (Figure 1).

Critical for Knl1's role in ensuring high fidelity chromosome segregation is the recruitment of the paralogs BubR1 and Bub1 (BUBs) to the outer kinetochore. Both BubR1 and Bub1 are bifunctional proteins, being involved in the SAC as well as in regulating stability of kinetochore–microtubule interactions [273]. Their roles in these processes are, however, distinct. BubR1 is a pseudokinase [147] that is a component of a diffusible anaphase inhibitor [139,251,274,275] and regulates stability of kinetochore–microtubule attachments by localizing the phosphatase PP2A-B56 to kinetochores [96–98]. Bub1 regulates error-correction by localizing Aurora B kinase to the inner centromere through the phosphorylation of T120 on the Histone 2A tail [254,255] and likely by localizing BubR1/PP2A to kinetochores [134,257,270], yet its role in the SAC is less well identified [273]. These two BUBs directly interact through their respective TPR (tetratricopeptide repeat) domains with two different KI motifs in the N-terminus of Knl1 [252,258,260]. These motifs are, however, not conserved beyond vertebrates and are not essential for BUB kinetochore binding in human cells [276,277]. Rather, the main BUB-recruitment site on Knl1 is an array of multiple so-called MELT repeats (Met-Glu-Leu-Thr). When phosphorylated by the mitotic kinase Mps1, they form phospho-docking sites for Bub3/Bub1 dimers, hence directly ensuring localization of Bub1 and indirectly of BubR1 to kinetochores [131,261,262].

We and others recently reported that the MELT repeats of human Knl1 are part of larger repeated units that contain (besides a central MELT-like motif) at least two other motifs required for function: A TΩ motif (TxxΩ; where Ω denotes aromatic residues), and a second phospho motif (SHT) C-terminal to the

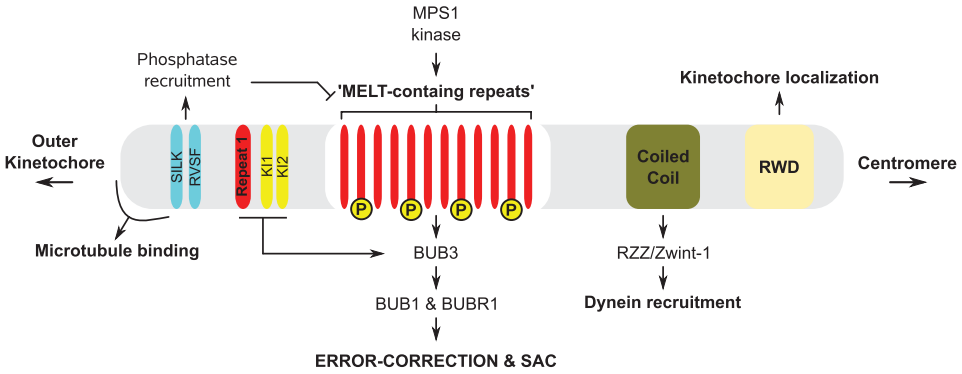


Figure 1 Knl1 is a hub for signaling at the kinetochore–microtubule interface Schematic representation of the domain/motif architecture of human Knl1. Phospho motifs (MELTs) in the disordered middle region of Knl1 function as binding sites for various factors involved in SAC activation and error-correction (Bub3–Bub1/BubR1). KI1 and KI2 increase the affinity of the BUB proteins for repeat 1. In addition this region harbors a basic patch involved in microtubule binding, as well as SILK/RVSF motifs for recruitment of PP1 phosphatase. PP1 can dephosphorylate the phospho-MELT motifs. The C-terminal region contains a tandem RWD (RING-WD40-DEAD) domain that localizes Knl1 to kinetochores and a coiled-coil that interacts with Zwint-1, a factor involved in recruiting the dynein adaptor RZZ–Spindly complex to kinetochores.

MELT motif [141,276,277]. Human Knl1 has approximately 20 of these larger repeats. We showed that only six repeats are capable of recruiting detectable BUB proteins to the kinetochore, which raises the question of the significance of the other 14 repeats. In addition, although pivotal for proper error-correction and SAC function, preliminary analyses hinted at a high degree of variation in Knl1 repeat evolution [167,276]. Although MELT-like motifs were at the core of repeat units of most Knl1 orthologs we analyzed, the remainder of the repeat sequences diverged greatly, and instances were observed where even a MELT-like motif was indiscernible.

We performed phylogenetic analyses to reconstruct Knl1 repeat evolution with the aim to understand its highly divergent patterns and the possible implications for BUB kinetochore recruitment and chromosome segregation in eukaryotes.

Results

Knl1 Orthologs are found in all eukaryotic supergroups

Despite extensive sequence variation we could define Knl1 orthologs (see Materials and Methods) in all eukaryotic supergroups. These include orthologs in the rhizarium (*Bigelowiella natans*), the excavate (*Naegleria gruberi*), archeaplastids (*Galdiera sulphurea*, *Physcomitrella patens* and other land plants) and the cryptophyte (*Guillardia theta*), species in which no Knl1 orthologs were

previously detected [167,276]. A Knl1 ancestor was therefore likely part of the genome of the Last Eukaryotic Common Ancestor (LECA). In all, a total of 110 Knl1 orthologs, displaying a great variety of sequence properties, were used in this study (Sequence File S1). *Of note: after this study we identified novel Knl1 homologs in stramenopile lineages. These sequences can be found in the Sequence File S1 associated to Chapter 2.*

Repeat arrays in Knl1 orthologs display rapid consensus sequence evolution and extensive number changes

To capture the evolutionary behavior of the repeated units in a systematic fashion, we built a framework for short sensitive repeat discovery (see Materials and Methods). The pipeline initiates with a probabilistic search for gapless repeats and in an iterative process refines a statistical sequence consensus profile (hidden Markov model) of the smallest possible single repeat unit. To facilitate the comparison between different taxa, we calculated both inter and intra species repeat unit variation in addition to the number of repeats per array. Our analyses of repeat units in the set of Knl1 orthologs revealed a number of striking observations, summarized in figure 2 and elaborated on thereafter. A brief summary: First, the number of MELT motif-containing repeats differs extensively between eukaryotic species, ranging from 0 in most land plants, up to approximately 35 in flies (Figure 2). Interestingly, we observed recurrent instances of repeat array expansion and/or regression between various taxa of the same clade throughout the eukaryotic tree of life. These include: vertebrates (clawed frog = 31 and zebra fish = 16), chordates (lancelet = 16 and the tunicates = 6-10), insects (silk worm = 8 and mosquito = 33) and fungi (*Spizellomyces punctatis* = 1 and *Yarrowia lipolytica* = 21) (Figure 2). Second, our classification method uncovered a high degree of variation in the repeat consensus sequence both within and between species. For example, expansion of a single repeat is apparent in the ascomycete fungus *Blumeria graminis*, while in zebra fish repeats have decayed and only the MELT motif has been conserved (Figure 2, Figure S1d). Similarly, repeats are highly divergent between Knl1 orthologs, displaying alterations to the canonical MELT motif as well as the presence of additionally conserved motifs, for example, TΩ, SHT and other potential phosphosites ([EDN]x[ST] or Rx[ST]) (e.g., insects in Figure 2). In addition, we observed that motifs that are part of one repeat evolve separately in other species (e.g., MELT and TΩ), which suggests different functions for these motifs and hinting at the modular nature of Knl1 repeat evolution (see "2nd" in Figure 2).

Recurrent Episodes of Extensive Repeat Array Reorganization and Repeat Diversification in Vertebrates and Drosophilids

The widespread diversity in repeat arrays did not permit the reconstruction of a bona fide LECA MELT-repeat array, but instead hinted at lineage-specific

Figure 2 Repeat analyses of Knl1 reveal recurrent patterns during 2 Gyr of eukaryotic evolution Cartoon of the eukaryotic tree of life with selected species from all eukaryotic supergroups containing Knl1 orthologs. The proteins and repeats are represented on scale in the middle. The color of the repeats indicates the degree of similarity to the repeat consensus (see legend). The repeat sequence consensus is depicted as a sequence logo on the right (colors reflect distinct amino acid properties and height of the letters indicates conservation of amino acids). The number of repeats per species is indicated in the light red (MELT-containing repeats) and blue (second repeats). The location of the MELT motif within the repeat is underlined for each species.

drivers and/or functions to explain this pattern of evolution. To determine the evolutionary relationship between the repeats, we resorted to a pairwise similarity matrix approach [278], as the short and divergent nature of the repeats did not allow for the use of common model-guided phylogenetic methods (e.g. WAG using RaxML; see Materials and Methods). Subsequent clustering of the similarity matrices allowed for the visualization and (partial) reconstruction of evolutionary events that gave rise to arrays of both individual and closely related species. We focused on vertebrates and drosophilids because of the optimal sampling of closely related species and well-annotated genomes within these taxa, which allowed for tracing diverse patterns of repeat array reorganization up to single repeat resolution. We observe the following: **(1)** *Short multiplex (2–6) block duplications:* block duplication is the main mechanism through which arrays are reorganized. For human Knl1, we found a triplet block duplication of the repeats 12–14 and 16–18 (Figure 3a, Figure S1a) [276]. With the exception of the Chinese tree shrew (which had an additional duplication, Figure S2), all placental mammals share the human array topology (see alignment S4), which was therefore likely part of their common ancestor (~65 Mya) [279]. Comparison with orthologs of the non-placental mammals opossum, Tasmanian devil (marsupials) and platypus (monotreme), revealed multiple block duplications of different size (2–6) in approximately the same region as the placental mammal duplication (Figure 3, Figure S1b and see dynamic region in Figure S2). **(2)** *Homogenization:* we observed additional instances of very recent single-copy repeat expansion that resulted in an almost complete overwriting of the array (hereafter referred to as homogenization). Most notably in lamprey (*Petromyzon marinus*, Figure S1c) and the ascomycete *Blumeria graminis* (Figure 2), the repeat arrays are highly similar within one species. The low number of substitutions in the DNA hints to a recent and rapid repeat regeneration event (Figure S3). **(3)** *Array size maintenance and repeat loss:* we noticed incomplete repeat units and discontinuous patterns of overlapping block duplication indicating that the repeats in the dynamic region of mammalian Knl1 were partially overwritten (see "+" signs in Figure S1a–c and the gaps in Figure S4). In addition, we observed that repeats in the middle of the dynamic region in platypus were more similar to each other compared with repeats at the outside of the array, indicating unequal crossover as a potential mechanism for array maintenance (Figure 3, Figure S1b). Some of the repeat units in mammals exhibit divergence from the repeat consensus ('* signs' in Figure S1a–c), acquiring multiple mutations in important residues, leading to decay and ultimately loss of these repeats. Strikingly, similarity between repeat 1,7 and 11 and those within the duplicated triplet block in human Knl1 correlates with their capacity to recruit BUB proteins, suggesting that diverged repeats loose their function [141,276]. In zebra fish, no order in which duplications were generated could be inferred and decay has occurred at multiple repeats, as both the TΩ and the SHT motif are lost (Figure 1, Figure S1d).

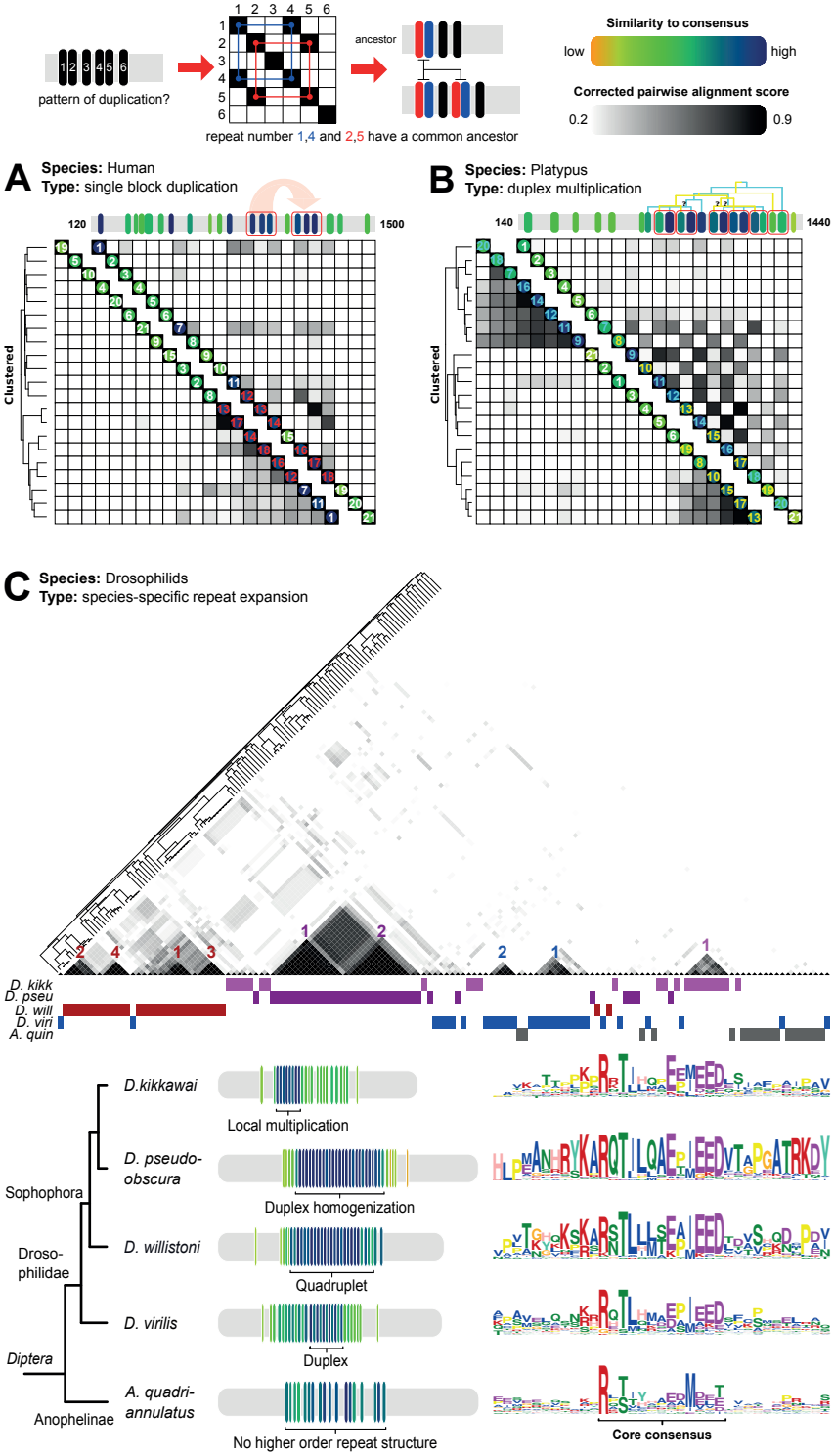

pattern of duplication?
ancestor
repeat number 1,4 and 2,5 have a common ancestor
Similarity to consensus
low
high
Corrected pairwise alignment score
0.2
0.9
A
Species: Human
Type: single block duplication
120
1500
Clustered
B
Species: Platypus
Type: duplex multiplication
140
1440
Clustered
C
Species: Drosophilids
Type: species-specific repeat expansion
D. kikk
D. pseu
D. will
D. viri
A. quin
D.kikkawai
D. pseudo-obscura
D. willistoni
D. virilis
A. quadri-annulatus
Sophophora
Droso-philidae
Diptera
Anophelinae
Local multiplication
Duplex homogenization
Quadruplet
Duplex
No higher order repeat structure
Core consensus

Figure 3 Patterns of repeat array reorganization in mammals and drosophilids Individual repeats are scored based on similarity to the repeat consensus (similar to Figure 2). The example matrix at the top depicts the duplication of a twin repeat block (1,2–4,5). Similarity matrices (clustered [bottom-left] and unclustered [upper-right]) show patterns of repeat duplication; above the matrices scaled linear representations of the repeat array. Repeat numbers are colored according to their shared ancestry. **(A)** A single block duplication of repeat triplet 12–13–14 or 16–17–18 shaped human Knl1**. (B)** Overlapping twin block multiplications point to a complex history of platypus Knl1 evolution. **(C)** Pseudohomogenization and near full array replacement in four *Drosophila* species. Colors below the matrix indicate which repeat in the matrix belongs to which species. Colored numbers correspond to position in amplicon of the respective species. Alignment of sequence logos indicates species-specific changes in consensus sequence. *Anopheles quadriannulatus* is a species of mosquito and is used to show *Drosophila*-specific increase in duplication rate.

All types of repeat evolution described also occurred within the drosophilid genus (25–40 Ma). (Figure S5). Four species (*Drosophila pseudoobscura*, *Drosophila virilis*, *Drosophila kikkawai*, and *Drosophila willistoni*) diverged their arrays to such extent, that we could only infer 2 one-to-one orthologous repeats (*Drosophila pseudoobscure* 2–3 and *Drosophila kikkawai* 11–12). Strikingly, each of these four species independently expanded specific repeats through subsequent rounds of extensive multiplication resulting in (partial) homogenization. This significantly altered the length of the array as well as the species-specific consensus sequence (Figure 3c, alignment S2).

Modular Evolution of Short Conserved Motifs in the Repeats

Recurrent episodes of array reorganization (expansion and contraction) may well be rooted in the selection for changes of the repeat consensus. To understand how the contents of the repeats such as those of the drosophilids have diverged, we tracked the behavior of the repeat consensus over approximately 550 Ma of arthropod evolution [280]. To that end, the repeat consensus sequence logos of 50 arthropods were manually aligned and centered at the MELT-like motif and other recognizable motifs such as the N-terminal TΩ (Figure 4a). We found that the MELT-like motif is altered at position 0, −1, −2 (relative to the Thr), intermediately changing from ME[LF]T in most species to DMSLT in moths, butterflies and the beetle *Dendroctonus ponderosae*, MEET in mosquitos (*Anophelini*), and finally EP[MI]EEE in drosophilids. The phosphoconsensus of TΩ switched between predominantly basic residues [KR] and acidic residue [DE] (see *Hymenoptera*) at position −2 relative to Thr. This creates a potential phosphorylation site for Aurora B-like basophilic or Plk1/Mps1-like acidophilic kinases, respectively. Knl1 is a known substrate for such kinases in opisthokont model organisms [167]. We also noticed a conserved proline at +4 (relative to Thr), which was also present in the repeats of the fungus *Yarrowia lipolytica* and red algae *Galderia sulphurea* (Figure 2), indicating parallel gain and a potential shared functionality. The differential loss and emergence of conserved short motifs, (for example TΩ and other phosphosites)

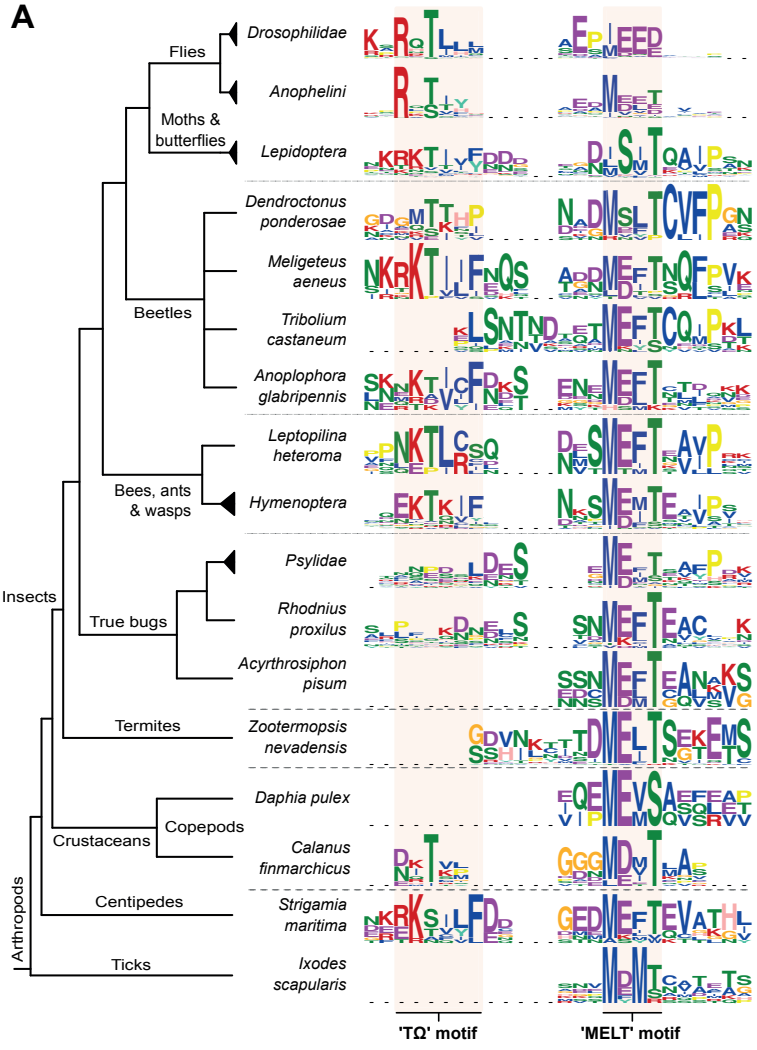

A
Flies
Moths & butterflies
Beetles
Bees, ants & wasps
Insects
True bugs
Termites
Crustaceans
Copepods
Centipedes
Arthropods
Ticks
Drosophilidae
Anophelini
Lepidoptera
Dendroctonus ponderosae
Meligeteus aeneus
Tribolium castaneum
Anoplophora glabripennis
Leptopilina heteroma
Hymenoptera
Psylidae
Rhodnius proxilus
Acyrthrosiphon pisum
Zootermopsis nevadensis
Daphia pulex
Calanus finmarchicus
Strigamia maritima
Ixodes scapularis
'TΩ' motif
'MELT' motif
B
Legend
Types of MELT motifs
Empty slot
TΩ motif
Phosphorylation site consensus
Specific conserved amino acids
Repeat slots
N-term
middle
MELT
C-term
1
2
3
4
gain/loss/switch
Observed motifs

Figure 4 Repeat sequence consensus evolution of arthropods (A) Alignment of repeat consensus sequences (weblogo) of arthropods based on the TΩ and MELT motif (red shaded). **(B)** Abstraction of conserved features indicates that repeats in arthropods consist of blocks that can be lost and gained. The repeat is subdivided into four "slots" (N-term, middle, MELT, and C-term) that contain all the observed motifs in arthropod evolution. Letters in blocks indicate the conservation of that amino acid or motif (P, proline; C, cysteine; GG, (double) glycine; "–", aspartate or glutamate; Φ, bulky hydrophobic residues; Ω, aromatic residues; phenylalanine or tyrosine).

signifies the modular character of the Knl1 repeat evolution. To reconstruct the repeat consensus evolution of all eukaryotes, we abstracted the repeats into a presence/absence pattern of frequently conserved short motifs, divided over four regions within repeats (Figure S6). We traced the origin of the TΩ motif to the base of the opisthokonts, with parallel loss in most fungi and early-branching animals (*Trichoplax adhaerens*, sea anemone, and sponges). Furthermore, we observe additional parallel events similar to those in arthropods (Figure 4b), such as TΩ phosphorylation consensus switching, MELT to MSLT/MEET and frequent changes of downstream conserved sites (glycines, proline, cysteine, and hydrophobic stretches) (see * signs for parallel events in Figure S6).

No Clear Indication for Positive Selection on Primate Knl1 Repeat arrays

As the evolutionary reconstruction reveals episodes of repeat array rearrangement and diversification, we wondered whether repeats in closely related species would be under positive selection (higher non-synonymous vs. synonymous substitution rate). We therefore fitted a concatenated alignment of the Knl1 repeats of 13 primates to various models of sequence evolution to estimate the dN/dS ratio using PAML [281] (see Materials and Methods, Figure S7). Although there seem to be different selective pressures impinging on the Knl1 repeat arrays in different species (Figure S7a), we could not detect significant positive selection on different sites (Figure S7b). Considering all sites, primate Knl1 repeats appear to be under weak purifying selection (dN/dS = 0.55).

Discussion

Our analyses and reconstructions reveal great diversity in the evolution of Knl1 repeat sequences. This diversity is the result of a myriad of mutations (repeat point mutation, loss, and duplication) further acted upon by selective forces. Together the interplay of these processes has driven a multitude of compound outcomes such as repeat homogenization and changes in repeat array length and consensus between closely related species (Figure 5). Similar patterns of rapid repeat evolution have been observed for proteins involved in adaptive evolution, for example in VERL, a protein involved in egg-sperm interaction in abalones [282], in Prdm9, a protein involved in homologous recombination during meiosis [283], and in the arms race between zinc-finger proteins and

retrotransposons [284]. Repeats in some core cellular proteins such as structural BRC repeats in the DNA-damage-related protein Brca2 [285,286] and a phosphomotif in the C-terminal domain of RNA polymerase [287] have likewise undergone striking repeat evolution in specific clades. To our knowledge however, our study is the first to trace such extensive dynamic repeat evolution for a disordered signaling protein across all eukaryotic supergroups.

Patterns and Mechanisms of Extensive Array Reorganization

Single-repeat or block-repeat multiplication is the result of duplications iterating in relatively quick succession. We find duplications undergoing no further dynamics, for example, approximately 65 Ma of evolution (placental mammals). In contrast, we also find cases where a block or single repeat underwent very recent iterating duplications (lamprey and drosophilids), indicating the episodic nature of Knl1 repeat evolution. Scars of overlapping block multiplications and a higher similarity of repeats in the middle of arrays (Figure 3b, Figure S1b) point to unequal crossover to maintain stable repeat arrays (Figure 5), similar to what was described for centromeric DNA repeat evolution [14]. Interestingly, high numbers of repeated units increase local sequence homology and thereby the chance of replication slippage and unequal crossover [288]. It is however unclear why the arrays never appear to be longer than approximately 35 units. This may have to do with the potential negative impact on chromosome segregation by a large number of Bub1–Bub3 recruitment modules, or of problematic protein folding/aggregation in case of extended unstructured regions. In any case, the array size limitation is indicative of purifying selection against excessive multiplications.

Patterns of Repeat Unit Consensus Evolution

The Knl1 repeat consensus sequence evolved in a modular fashion. It consists of several short conserved motifs, which are recurrently gained (indicative of convergent motif evolution) and lost at both up- and downstream positions relative to the MELT motif. The Knl1 repeat thus serves as a unit that contains multiple motif slots. This unit is dynamic in the motif content of its slots as well as dynamic in duplication and losses. Although the motifs slots seem to evolve dynamically on large time scales, on shorter time scales species-specific alignments of repeats units reveal conservation of each motif consensus by purifying selection, allowing us in fact to detect them as such (see sequence logos). Simultaneously, episodes of extensive array reorganization could lead to the expansion of specific repeat isoforms (signified by homogenization events), indicating how species have rapidly evolved their repeat consensus sequence.

Drivers of Repeat Evolution: A Role for Bub1–Bub3?

The wide array of evolutionary processes impinging on the Knl1 repeat ar-

ray raises the question what function of the repeats is driving these processes? We envision two distinct but non-mutually exclusive possibilities: 1) the altering number of repeats signifies different requirements for the number of Bub1–Bub3 molecules needed on a kinetochore or the length of the protein. As the number of functional repeats in human Knl1 dictates the efficiency of attachment error-correction [276], selective pressures may have called for rapid adaptability of the number of BUB molecules that can bind kinetochores. In such a scenario, the appearance of additional motifs could reflect differences in the Bub3 structure and/or regulatory pathways that impinge on Bub3 kinetochore recruitment. Recent work from our lab on human Knl1 showed that a vertebrate-specific SHT motif, C-terminal to the MELT motif, is an additional phosphomotif that interacts with a basic patch on the surface of Bub3 [141]. This patch is present in numerous Bub3 homologs of nonvertebrates, indicating co-option of pre-existing Bub3 features for interaction with the SHT motif in the ancestor of vertebrates. It is therefore possible that the various motifs in diverse eukaryotes bind to various conserved core features of the Bub3 structure. Of interest is also the loop region within Bub1 that stabilizes the interaction, which diversifies rapidly throughout eukaryotic evolution. Finally, some of the motifs may have evolved to accommodate different cell division kinases/phosphatases, possibly explaining changes in phospho-motif sequences. Further detailed molecular and functional analyses of the repeat motifs and their mode of interaction with the Bub3–Bub1 dimer, kinases, and –or phosphatases will be required to understand the repeat evolution. 2) A minimal requirement for Bub3 binding is maintained through purifying selection on the core MELT-like motif and the changes in number and sequence of additionally conserved motifs (e.g., the additional phosphosites) signify other, yet unknown functions of Knl1 repeat divergence. The observed repeat (pseudo) homogenization events in *Blumeria graminis*, lamprey, and several drosophilids are reminiscent of genetic conflicts, such as the compensatory evolution of centromere sequences and centromere-binding proteins to prevent genetic conflict during asymmetric meiosis, known as centromere drive [29]. The centromere-drive hypothesis describes an arms race between centromere sequence variants with higher probabilities of being retained in the oocyte (rather than the evolutionary invisible polar bodies) and centromere-binding proteins that negate this bias [15,289]. Interestingly, in nematodes Knl1 is involved in biorientation of acentrosomal meiosis [290] and Knl1 protein expression is highest at the sperm acrosome in humans [291]. Nevertheless, there is currently no evidence that Knl1 binds centromere sequences directly, and rapid evolution of its repeats occurs also in species with symmetric meiosis. Other forms of genetic conflict that may explain Knl1 repeat evolution include defense against supernumerary/selfish (B-) chromosomes that utilize kinetochore proteins and the mitotic spindle to segregate [292], or in the evasion of hijacking of the mitotic machinery by intracellular pathogens.

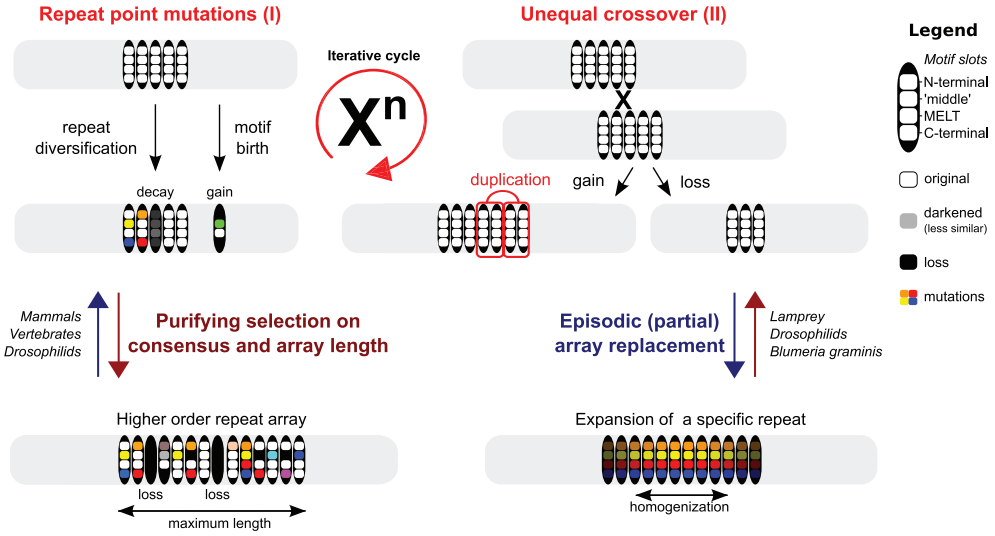


Figure 5 Model of repeat evolution in Knl1. Knl1 repeat units (black bars) are depicted as having four "motif slots". The color white indicates the ancestral state of the repeat; black the loss of the respective slot; and further coloring signifies subsequent mutations. Arrays are subjected to continuous repeat turnover (gain/loss) through iterative cycles of unequal crossover (**II**) in combination with repeat point mutation (**I**) leading to repeat diversification, potential decay (loss), and de novo motif emergence. Repeat arrays are stabilized by purifying selection to maintain a sufficient number of functional repeats (dark red). Intermittent episodes of extensive single copy expansion allow for rapid evolution of the consensus and/or array length, which is reminiscent of adaptive evolution (dark blue). Species names indicate which type of behavior is seen for that species.

Materials and Methods

Sequences

Classical homology searches using BLAST (Basic Local Alignment Search Tool) failed to detect sufficient homology for Knl1 genes. We therefore performed iterated sensitive homology searches with HMMer [235,293], using a permissive E-value and bit-score cut-off to include diverged homologs. Given that we detected a single homolog per genome we considered them orthologs. We included orthologs based on the presence of a N-terminal PP1-recruitment motifs (SILK/RVSF), MELT-like repeats, conserved regions in the C-terminus including a recently discovered RWD domain [54], and a C-terminal coiled-coil region. Incompletely predicted genes were searched against whole-genome shotgun contigs (http://www.ncbi.nlm.nih.gov/genbank/wgs) using tBLASTn. Significant hits were manually predicted using AUGUSTUS [245] and GENESCAN [244]. For the sequences that we used in this study, see Sequence File S1.

Repeat Discovery Pipeline

The MEME [218] algorithm (option: anr) was used to search for gapless amino acid repeat sequences, which were aligned using MAFFT [212] (option: einsi). Sensitive profile HMM searches (permissive E-value of 10) of the aligned repeats were iterated until convergence [235]. Due to the sensitivity of the profile HMM searches, the results were manually scrutinized for obvious errors.

Sequence Logos and Similarity Matrices

The repeat consensus sequence was depicted as a sequence logo using Weblogo2 (MEME color scheme). To prevent over interpretation of gaps and infrequent amino acids, columns in the repeat alignment with less than 20% occupancy were removed. The deviation from the consensus of individual repeats was calculated by normalizing pairwise alignment scores (Smith–Waterman) for the highest average score of all repeats and corrected for their respective length. We visualized repeat evolution history by projection of the normalized and corrected Smith–Waterman scores onto a similarity matrix (as described by Björklund et al. [278]). Subsequent clustering enabled the classification of repeats with shared ancestry. Due to incomplete and dispersed clustering, further manual assignment of clusters and thus repeat phylogeny was necessary. The short length and limited amount of conserved sites between repeat units did not allow us to fit the Knl1 repeat data to a model of sequence evolution in order to reconstruct its evolution, due to lack of power and likely over –or under fitting of model parameters (at least need ~50 amino acids per repeat unit for good results).

Contributions

BS and GJPLK managed and conceived the project. ET performed all the research. ET, BS and GJPLK analyzed the data and wrote the manuscript.

Acknowledgments

The authors thank the members of the Kops and Snel lab for fruitful discussion and critical reading of the manuscript. The authors also thank members of the Malik, Henikoff and Biggins labs at FHCRC for discussions. This work is supported by a grant from the Netherlands Organization for Scientific Research (NWO-Vici 865.12.400).

Electronic Supplementary Material

Supplementary materials for this chapter are made available online at the following link http://bioinformatics.bio.uu.nl/eelco/thesis/

Sequence File S1 is a fasta file of the homologous sequences of all eukaryotic Knl1 proteins used in this study.

Alignment PAML analysis contains a multiple sequence alignment of concatenated repeats (xxxTxxFxxxMELTxSHTxxx) of 13 selected primate Knl1 sequences without gaps (PAML format, FASTA format for codons and amino acid sequences).

Alignment S1 show a multiple sequence alignment of mammalian + turtle repeats on which the similarity matrix of Figure S2 is based.

Alignment S2 shows a multiple sequence alignment of drosophilid + mosquito repeats on which the similarity matrix of Figure 3c is based.

Alignment S3 shows a multiple sequence alignment of drosophilid repeats on which the similarity matrix of Figure S5 is based.

Alignment S4 contains a multiple sequence alignment of full-length Knl1 of various placental mammals on which the 21 repeats of human are projected.

Figure S1 shows repeat alignments related to Figure 3a,b, showcases repeat homogenization in *Petromyzon marinus* (Lamprey) and extensive degeneration of repeat in *Danio rerio* (Zebrafish).

Figure S2 shows clustered similarity matrices for four mammal species + turtle as outgroup and a manual reconciliation of the repeat duplication history.

Figure S3 DNA alignment of the repeats of *Blumeria graminis* and *Petromyzon marinus*.

Figure S4 alignment of larger blocks containing multiple repeats for a number of marsupial species.

Figure S5 contains an expanded analysis similar to Figure 3c for 19 drosophilid species.

Figure S6 shows the modular evolution of the repeat consensus sequence in eukaryotes similar to Figure 4b. Loss, gain and other mutational events of repeat slots are projected onto the eukaryotic tree of life.

Figure S7 shows the PAML analyses for quantifying selective pressures on the Knl1 repeat regions in various primates.

5

CHAPTER 6

Phylogenomics-guided discovery of a conserved cassette of short linear motifs in BubR1 essential for the spindle checkpoint

Eelco Tromer, Debora Bade, Berend Snel & Geert JPL Kops

Open Biology, 2016

Abstract

The spindle assembly checkpoint (SAC) maintains genomic integrity by preventing progression of mitotic cell division until all chromosomes are stably attached to spindle microtubules. The SAC critically relies on the paralogs Bub1 and BubR1/Mad3, which integrate kinetochore–spindle attachment status with generation of the anaphase inhibitory complex MCC. We previously reported on the widespread occurrences of independent gene duplications of an ancestral 'MadBub' gene in eukaryotic evolution and the striking parallel subfunctionalization that lead to loss of kinase function in BubR1/Mad3-like paralogs. Here, we present an elaborate subfunctionalization analysis of the Bub1/BubR1 gene family and perform de novo sequence discovery in a comparative phylogenomics framework to trace the distribution of ancestral sequence features to extant paralogs throughout the eukaryotic tree of life. We show that known ancestral sequence features are consistently retained in the same functional paralog: GLEBS/CMI/CDII/kinase in the Bub1-like and KEN1/KEN2/D-Box in the BubR1/Mad3-like. The recently described ABBA motif can be found in either or both paralogs. We however discovered two additional ABBA motifs that flank KEN2. This cassette of ABBA1-KEN2-ABBA2 forms a strictly conserved module in all ancestral and BubR1/Mad3-like proteins, suggestive of a specific and crucial SAC function. Indeed, deletion of the ABBA motifs in human BubR1 abrogates the SAC and affects APC/C–Cdc20 interactions. Our detailed comparative genomics analyses thus enabled discovery of a conserved cassette of motifs essential for the SAC and shows how this approach can be used to uncover hitherto unrecognized functional protein features.

Introduction

Chromosome segregation during cell divisions in animals and fungi is monitored by a cell cycle checkpoint known as the spindle assembly checkpoint (SAC) [101,167,272]. The SAC couples absence of stable attachments between kinetochores and spindle microtubules to inhibition of anaphase by assembling a four-subunit inhibitor of the anaphase-promoting complex (APC/C), known as the MCC [20,145,251]. The molecular pathway that senses lack of attachment and produces the MCC relies on two related proteins known as Bub1 and BubR1/Mad3 [272]. Bub1 is a serine/threonine kinase that localizes to kinetochores and promotes recruitment of MCC subunits and of factors that stimulate its assembly [140,141,257]. These events are largely independent of Bub1 kinase activity, however, which instead is essential for the correction process of attachment errors [254,257,294]. BubR1/Mad3 is one of the MCC subunits, responsible for directly preventing APC/C activity and anaphase onset [251,274,295]. It does so by contacting multiple molecules of the APC/C co-activator Cdc20, preventing APC/C substrate access and binding of the E2 enzyme UbcH10 [108,121,145,251]. The BubR1/Mad3–Cdc20 contacts occur via various short linear motifs (SLiMs) known as ABBA, KEN and D-box [108,111,120,121,251,296,297]. Like Bub1, BubR1 also impacts on the attachment error-correction process via a KARD motif that recruits the PP2A-B56 phosphatase [96–98,112]. This may not however be a universal feature of BubR1/Mad3-like proteins, because many lack a KARD-like motif.

Bub1 and BubR1/Mad3 are paralogs. We previously showed they originated by similar but independent gene duplications from an ancestral MadBub gene in many lineages, and that the two resulting gene copies then subfunctionalized in remarkably comparable ways [147]. An ancestral N-terminal KEN motif (KEN1: essential for the SAC) and an ancestral C-terminal kinase domain (essential for attachment error-correction) were retained in only one of the paralogous genes in a mutually exclusive manner in virtually all lineages (i.e. one gene retained KEN but lost kinase, while the other retained kinase but lost KEN). One exception to this 'rule' is vertebrate clade, where both paralogs have a kinase-like domain. The kinase domain of human BubR1 however lacks enzymatic activity (i.e. is a pseudokinase) but instead confers stability onto the BubR1 protein [147].

The similar subfunctionalization of Bub1 and BubR1/Mad3-like paralogs was inferred from analysis of two domains (TPR and kinase) and one motif (KEN1). We set out to analyze whether any additional features specifically segregated to Bub1- or BubR1/Mad3-like proteins after duplications by designing an unbiased feature discovery pipeline and tracing feature evolution. The pipeline extracted all known and various previously unrecognized conserved motifs from

Bub1/BubR1 family gene members. Two of these are novel ABBA motifs that flank KEN2 specifically in BubR1/Mad3-like proteins; we show that this highly conserved ABBA-KEN2-ABBA cassette is crucial for the SAC in human cells.

Results and discussion

Refined phylogenomic analysis of the MadBub gene family pinpoints 16 independent gene duplication events across the eukaryotic tree of life

To enable detailed reconstruction of subfunctionalization events of all known functional features after duplication of ancestral MadBub genes, we expanded our previously published set of homologs [147] through broader sampling of sequenced eukaryotic genomes, focusing on sequences closely associated with duplication events (see for sequences Sequence File S1). Phylogenetic analyses of a multiple sequence alignment of the TPR domain (the only domain shared by all MadBub family members) of 149 MadBub homologs (Supplementary Procedures and Discussion, Figure S1) corroborated the 10 independent duplications previously described [147] and allowed for a more precise determination of the age of the duplications. Strikingly, we found evidence for a number of additional independent duplications: three duplications in stramenopile species of the SAR super group (Albuginaceae (#10 in Figure 1b), *Ectocarpus siliculosis* (#11) and *Aureococcus anophagefferens* (#12)) and one at the base of basidiomycete fungi (puccinioimycetes (#4)). The BubR1 paralog in teleost fish underwent a duplication and fission event, of which the C-terminus product was retained only in the lineage leading to zebra fish (*Danio rerio* (#7)). Lastly, through addition of recently sequenced genomes we could specify a duplication around the time plants started to colonize land (bryophytes (#13)) and an independent duplication in the ancestor of higher plants (tracheophytes (#14)), followed by a duplication in the ancestor of the flowering plants (magnoliaphytes (#15)). These gave rise to three MadBub homologs, signifying additional subfunctionalization of the paralogs in the plant model organism *Arabidopsis thaliana*. It thus seems to be the case that such striking parallel subfunctionalization as we originally identified is indeed predictive for more of its occurrence in lineages whose genome sequences have since been elucidated.

De novo discovery, phylogenetic distribution and fate after duplication of functional motifs in the MadBub gene family

Previous analyses revealed a recurrent pattern of mutually exclusive retention of an N-terminal KEN-box and a C-terminal kinase domain after duplication of an ancestral MadBub gene [147,298]. These patterns suggested the hypothesis of paralog subfunctionalization towards either inhibition of the APC/C in the cytosol (retaining the KEN-box) or attachment-error correction at the kineto-

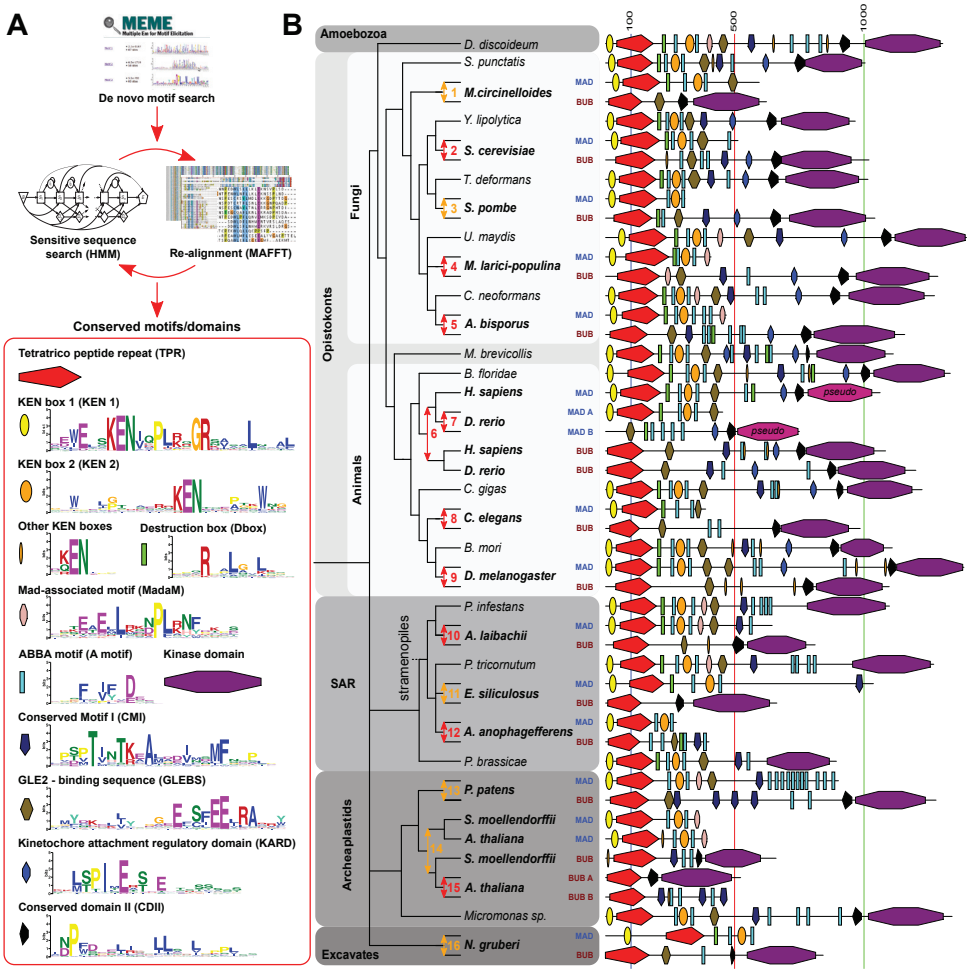


Figure 1 Fate of conserved functional sequence features after 16 independent duplications of the MadBub gene family throughout eukaryotic evolution (A) Overview of the de novo sequence discovery pipeline ConFeaX including the ancestral conserved features of a search against the eukaryotic MadBub gene family. The consensus sequences of the detected conserved motifs are depicted as a sequence logo (colors reflect distinct amino acid properties and height of the letters indicates conservation of amino acids). Each feature is assigned a differently colored shape. **(B)** Cartoon of the evolutionary scenario of 16 independent duplications of the MadBub gene family throughout eukaryotic evolution, including a projection of conserved features onto the linear protein representation (on scale). Gene duplications are indicated by an arrow (red: high confidence, orange: likely). The subfunctionalized paralogs Mad and Bub are colored brown and blue, respectively. Numbers indicate the clades in which the duplications occurred: 1, *Mucorales*; 2, *Saccharomycetaceae*; 3, schizosaccharomycetes; 4, pucciniomycetes: 5, agaricomycetes (excluding early branching species); 6, vertebrates; 7, teleost fish; 8, nematodes; 9, diptera (flies); 10, *Albuginaceae* (oomycete); 11, Ectocarpales (brown algae); 12, *Aureococcus* (harmful algae bloom); 13, bryophytes (mosses); 14, tracheophytes (vascular plants); 15, magnoliaphytes (flowering plants); 16, *Naegleria*.

chore (retaining the kinase domain). Given the extensive sequence divergence of MadBub homologs and a scala of different known functional elements, we reasoned that a comprehensive analysis of MadBub gene duplicates would provide opportunities for the discovery of novel and co-evolving ancestral features. For clarity, we refer to the Bub1-like paralog (C-terminal kinase domain) as Bub and the BubR1/Mad3-like paralog (N-terminal KEN-box) as Mad throughout the rest of this paper.

To capture conserved ancestral features of diverse eukaryotic MadBub homologs, we constructed a sensitive de novo motif and domain discovery pipeline (ConFeaX: conserved feature extraction) similar to our previous approach used to characterize KNL1 evolution [133]. In short, the MEME algorithm [218] was used to search for significantly similar gapless amino acid motifs, and extended motifs were aligned by MAFFT [212]. Alignments were modeled using HMMER [235] and sensitive profile HMM searches were iterated and specifically optimized using permissive E-values/bit-scores until convergence (Material and methods and Figure 1a). Owing to the degenerate nature of the detected SLiMs, we manually scrutinized the results for incorrectly identified features and supplemented known motif instances, when applicable. We preferred ConFeaX to other de novo motif discovery methods [231,232], as it does not rely on high quality full-length alignment of protein sequences and allows detection of repeated or dynamic non-syntenic conserved features (which is a common feature for SLiMs). It is therefore better tuned to finding conserved features over long evolutionary distances in general and specifically in this case where recurrent duplication and subfunctionalization hamper conventional multiple sequence alignment based analysis.

ConFeaX identified known functional motifs and domains and in some cases extended their definition: KEN1 [146], KEN2 [120], GLEBS [299], KARD [96–98], CMI (also known as CDI [257]), D-box [120], CDII (a co-activator domain of Bub1 [257,300]) and the recently discovered ABBA motif (here termed ABBA3) [111,112,141,296] (Figure 1a, table S2, Sequence File 2). The TPR and the kinase domain were annotated using profile searches of previously established models [147] and excluded from de novo sequence searches. KEN1 and KEN2 could be discriminated by differentially conserved residues surrounding the core KEN-box (Figure 1a). Those surrounding KEN1 are involved in the formation of the helix-turn-helix motif that positions BubR1/Mad3 towards Cdc20 [251], while two pseudo-symmetrically conserved tryptophan residues with unknown function specifically defined KEN2. Furthermore, we found that the third position of the canonical ABBA motif is often occupied by a proline residue and the first position in ascomycetes (fungi) is often substituted for a polar amino acid (KRN) (Figure 1a), signifying potential lineage-specific changes in Cdc20–ABBA interactions. Last, we also discovered a novel motif predomi-

nantly associated with the Mad paralog in basidiomycetes, plants, amoeba and stramenopiles but not metazoa, which we termed Mad-associated motif (MadaM) (Figure 1a).

Projection of the conserved ancestral features onto the MadBub gene phylogeny provided a highly detailed overview of MadBub motif evolution (Figure 1b, Figure S1b). We found that the core functional motifs and domains (TPR, KEN1, KEN2, ABBA, D-box, GLEBS, MadaM, CMI, CDII and kinase) are present throughout the eukaryotic tree of life, representing the core features that were probably part of the SAC signalling network in the last eukaryotic common ancestor (LECA). Of note are lineages (nematodes, flatworms (*Schistosoma mansoni)*, dinoflagellates (*Symbiodinium minutum*) and early branching fungi (microsporidia and *Conidiobolus coronatus*)) for which multiple features were either lost or considerably divergent (Figure S1b). Especially interesting is *Caenorhabditis elegans* in which both KEN boxes and the GLEBS domain appear to have been degenerated (ceMad = san-1) and the CMI motif is lost (ceBub = bub-1), indicating extensive rewiring or a less essential role of the SAC in nematode species, as has been suggested recently [110,301].

Our motif discovery analyses revealed the Cdc20/Cdh1-interacting ABBA motif to be much more abundant than the single instances that were previously reported for BubR1 and Bub1 in humans [111,141,296]. We observed three different contexts for the ABBA motifs (Figure 1b, Figure S1b): (i) in repeat arrays (e.g. Mad of *Physcomitrella patens*, basidiomycetes and stramenopiles), (ii) in the vicinity of CMI (many instances) and/or D-box/KEN (e.g. human) and (iii) as two highly conserved ABBA motifs flanking KEN2 (virtually all species). Because of the positional conservation of the latter, we have termed these ABBA1 and ABBA2. Any additional ABBA motifs were pooled in the category 'ABBA-other'.

In order to track the fate of the features discovered using ConFeaX, we quantified their co-presences and -absences, as a proxy for co-evolution, by calculating the Pearson correlation coefficient (*r*) for the profiles of each domain/motif pair of 16 duplicated MadBub homologs (Figure 1b) [219]. Subsequent average clustering of the Pearson distance ($d = 1 - r$) revealed two sets of co-segregating and anti-correlated conserved features (Figure 2a,b) consistent with our hypothesis that MadBub gene duplication caused parallel subfunctionalization of features towards the kinetochore (mainly Bub) and the cytosol (Mad) [147]. GLEBS, CMI, ABBA-other, KARD, CDII and the kinase domain formed a coherent cluster of features with bona fide function at the kinetochore. For a detailed discussion on several intriguing observations regarding presence/absence of these motifs in several eukaryotic lineages, and what this may mean for Bub/Mad and SAC function in these lineages, see the Supplementary Pro-

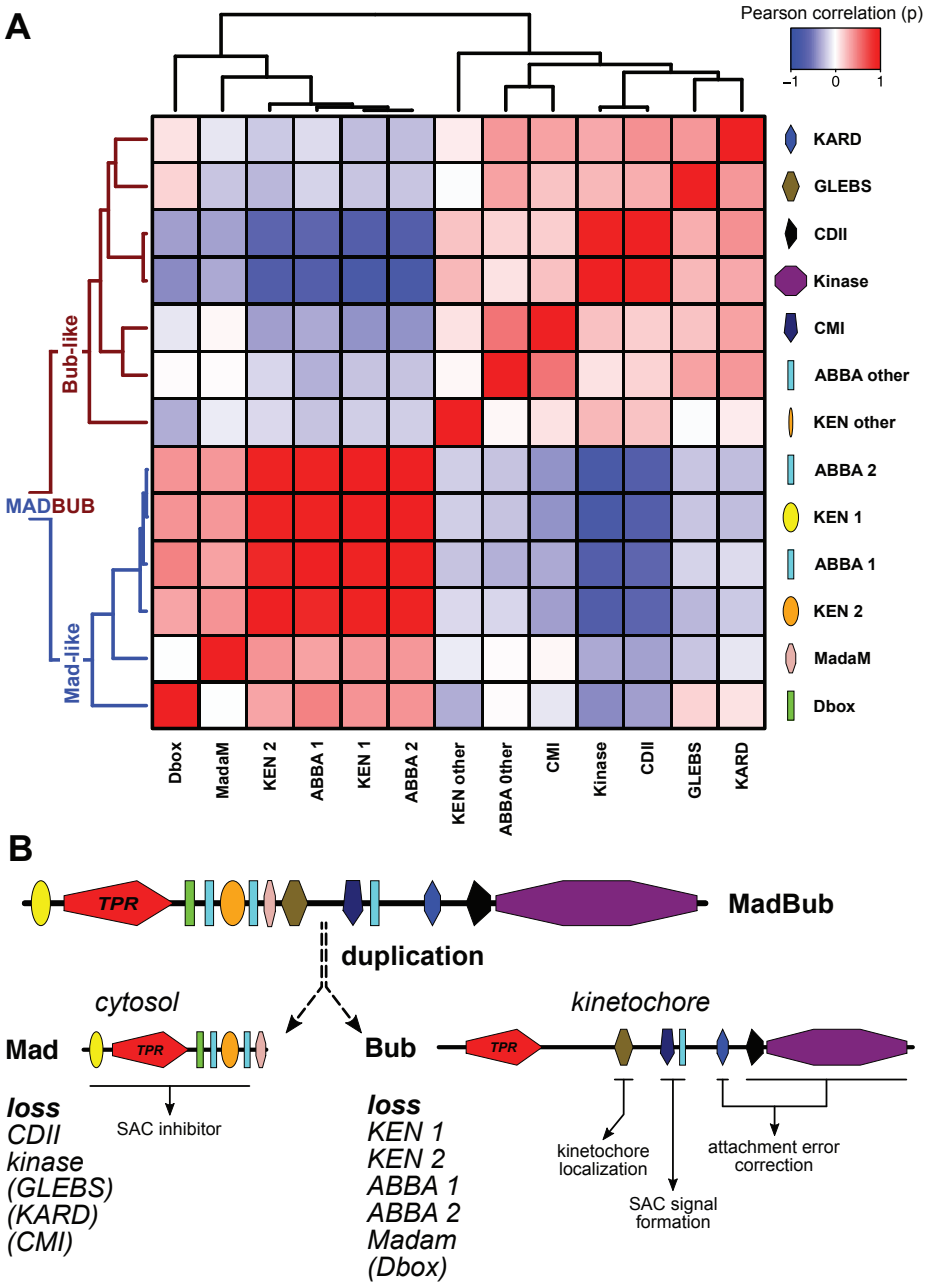


Figure 2 Co-evolution of conserved features signify subfunctionalization of Mad and Bub after MadBub duplication. (A) Average clustering based on Pearson distances of conserved ancestral feature correlation matrix (d = 1 – *r)* of all MadBub paralogs from figure 1. Red and blue indicate co-presence or -absence of features in the same paralog, respectively. **(B)** Evolutionary scenario of MadBub subfunctionalization: Mad (cytosol) as a SAC effector and Bub (kinetochore) involved in SAC signal formation and kinetochore microtubule attachment.

cedures and Discussion (Electronic Supplementary Material). A second cluster contained known motifs that bind and interact with (multiple) Cdc20 molecules, including KEN1, KEN2 and (to a lesser extent) the D-box. Our newly discovered ABBA motifs that flank KEN2 were tightly associated with KEN2 and KEN1 (Figure 2). As such, the ABBA1-KEN2-ABBA2 cassette (Figure 3a) co-segregated with Mad function during subfunctionalization of MadBub gene duplicates. Although the D-box often co-occurs with the KEN–ABBA cluster, this motif was occasionally lost (e.g. archeaplastids, *Schizosaccharomyces pombe* and *Aureococcus anophagefferens*). Finally, MadaM co-segregated with the Cdc20-interacting motifs (Figure 2a), suggesting a Mad-specific role for this newly discovered motif (possibly in MCC function and/or Cdc20-binding) in species harboring it such as plants, basidiomycetes and stramenopiles.

The conserved ABBA1-KEN2-ABBA2 cassette is essential for SAC signaling in human cells

The strong correlation of the ABBA1-KEN2-ABBA2 cassette with KEN1 and the D-box urged us to examine the role of these motifs in BubR1-dependent SAC signaling in human cells. We therefore generated stable isogenic HeLa-FlpIn cell lines expressing doxycyclin-inducible versions of LAP-tagged BubR1 [302]. These included: ΔABBA1, ΔABBA2, ΔABBA1 + 2, alanine-substitutions of the two KEN2-flanking tryptophans (W1-A, W2-A and W1/2-A), KEN1-AAA, KEN2-AAA, ΔABBA3 and ΔD-box (Figure 3a–c). The SAC was severely compromised in cells depleted of endogenous BubR1 by RNAi, as measured by inability to maintain mitotic arrest upon treatment with S-trityl-l-cysteine (STLC) [303] (median (m) = 50 min from nuclear envelope breakdown to mitotic exit, compared with control (m > 500 min)) (Figure 3d,e). SAC proficiency was restored by expression of siRNA-resistant LAP-BubR1 (m > 500 min). As shown previously [120,304,305], mutants of KEN1, KEN2 and the D-box strongly affected the SAC. Importantly, BubR1 lacking ABBA1 or ABBA2 or both, or either of the two tryptophans, could not rescue the SAC (Figure 3e). We observed a consistently stronger phenotype for the mutated motifs on the N-terminal side of KEN2 (ΔABBA1 (m = 65 min) and W1-A (m = 165 min)) compared with those on the C-terminal side (ΔABBA2 (m = 200 min) and W2-A (m = 260 min)). The double ABBA (1/2) and tryptophan (1/2) mutants were however further compromised (m = 50 and 110 min, respectively), suggesting non-redundant functions. As expected from the interaction of ABBA motifs with the WD40 domain of Cdc20 [111,121], BubR1 lacking ABBA1 and/or ABBA2 was less efficient in binding APC/C-Cdc20 in mitotic human cells, to a similar extent as mutations in KEN1 (figure 3f). In our hands, the ABBA1 and ABBA2 mutants were strongly deficient in SAC signaling and APC/C-Cdc20 binding while the previously described ABBA motif (ABBA3) was not (Figure 3d,e). Previous studies suggested that ABBA3 might play a role in SAC silencing [16,42], which

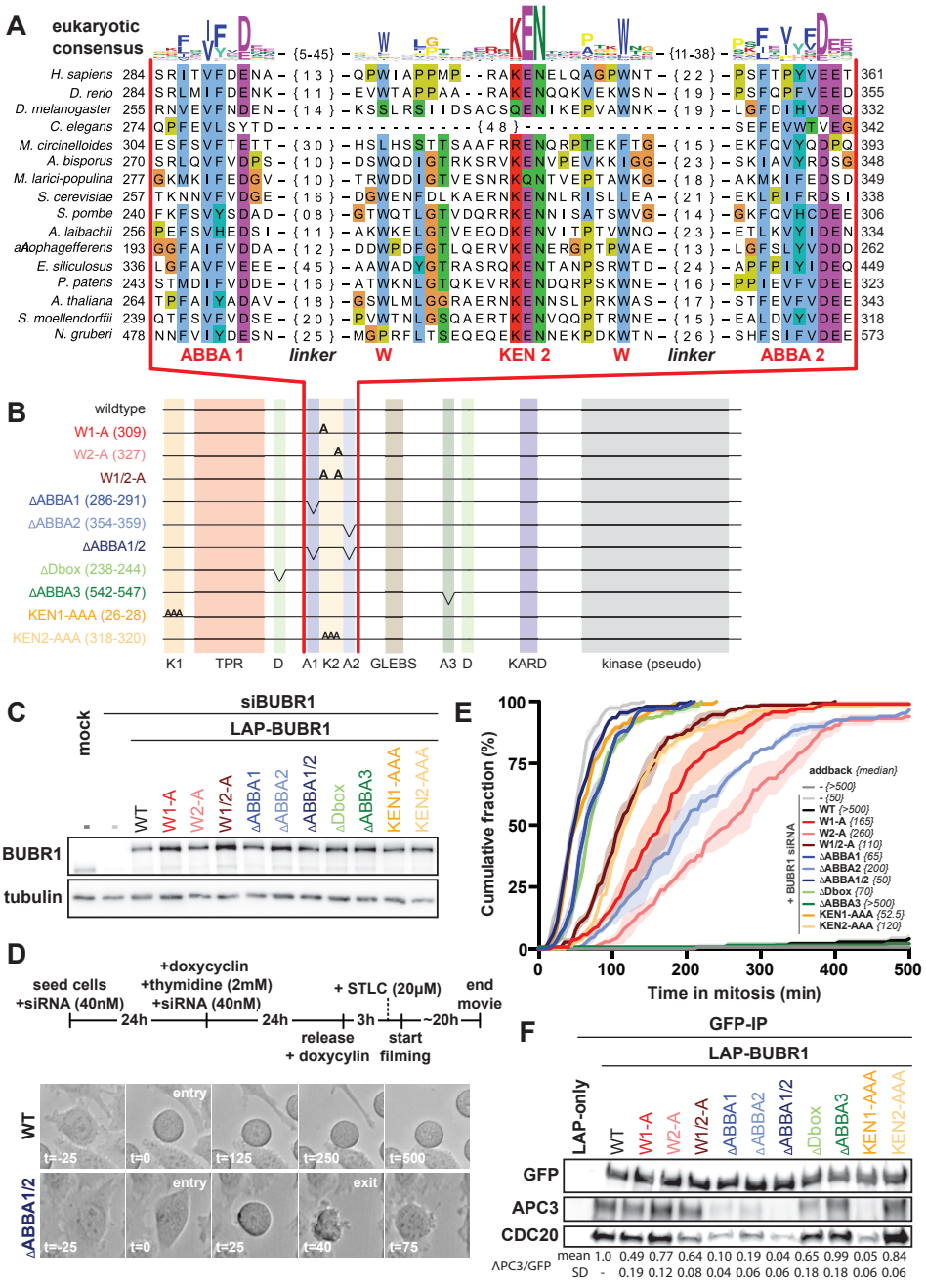


Figure 3 The evolutionary conserved cassette ABBA1-KEN2-ABBA2 in BubR1 is essential for SAC signaling. (A) Alignment of ABBA1-KEN2-ABBA2 cassette (red). Linkers (black) between ABBA motifs and KEN2 are indicated by {n}. The sequence logo on top is representative for all eukaryotic sequences (colors reflect distinct amino acid properties and height of the letters indicates conservation of amino acids). **(B)** Schematic representation of LAP-hBubR1 mutants. Color-coding is consistent for each mutant in this figure. **(C)** Immunoblots of BubR1 and tubulin

of mitotic lysates of HeLa-FlpIn cell lines stably expressing LAP-tagged BubR1 proteins. Cells were treated with siRNA (40 nM) for 48h and cells were released and arrested into Taxol after double thymidine block. **(E)** Time-lapse analysis of HeLa-FlpIn cells expressing hBubR1 mutants, treated with 20 μM STLC. Data (N = 3 with n = 50 per experiment) indicate the mean of cumulative fraction of cells that exit mitosis after nuclear envelope breakdown. Transparent regions represent the standard error of the mean. Values between braces {} indicate the median value. Cells were scored by cell morphology using DIC imaging; see **(D)** for examples of SAC deficient (ΔABBA1/2) and proficient cells (wild-type). Only YFP-positive cells were considered for analyses. **(F)** Immunoblots of GFP, APC3 and Cdc20 in LAP-BubR1 precipitations (LAP-pulldown) in whole cell lysates of mitotic HeLa-FlpIn cells expressing LAP-BubR1 mutant constructs. The mean and standard deviation values of three independent APC3/GFP co-immunoprecipitation experiments for all mutant LAP-BubR1 cell lines are normalized to wild-type LAP-BubR1 and depicted below the immunoblots.

raises the possibility that ABBA3 may somehow counteract binding of ABBA1 and/or ABBA2 to Cdc20. In conclusion, the ABBA1-KEN2-ABBA2 cassette in BubR1 is essential for APC/C inhibition by the SAC.

We here discovered a symmetric cassette of SLiMs containing two Cdc20-binding ABBA motifs and KEN2. This cassette strongly co-occurs with KEN1 in Mad-like and MadBub proteins throughout eukaryotic evolution and has important contributions to the SAC in human cells. Our co-precipitation experiments along with the known roles for ABBA-like motifs and KEN2 and their recent modeling into the MCC-APC/C structure [108,121] strongly suggest that the ABBA1-W1-KEN2-W2-ABBA2 cassette interacts with one or multiple Cdc20 molecules. Together with KEN1, these interactions probably regulate affinity of MCC for APC/C or its positioning once bound to APC/C. The constellation of interactions between two Cdc20 molecules (Cdc20MCC and Cdc20APC/C) and the various Cdc20-binding motifs in one molecule of BubR1 (3× ABBA, 2× KEN and a D-box) is not immediately clear, and will have to await detailed atomic insights. One suggestion that arises from our study is that the ABBA3 motif that is modeled into the APC/C-MCC structure by *Alfieri et al.* [121] might well be the ABBA2 motif. The symmetric arrangement of the cassette may be significant in this regard, as is the observation that (despite a highly conserved WD40 structure of Cdc20) the length of spacing between the ABBA motifs and KEN2 is highly variable between species. A more detailed understanding of SAC function may be aided by ConFeaX-driven discovery of lineage-specific conserved features in the MadBub family when more genome sequences become available, as well as of features in other SAC proteins families.

Material and methods

Phylogenomic analysis

We performed iterated sensitive homology searches with jackhmmer [43] (based on the TPR, kinase, CMI, GLEBS and KEN boxes) using a permissive E-value and bitscore cut-off to include diverged homologs on UniProt (2016_08) and Ensemble Genomes 32 (http://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer). Incompletely predicted genes were searched against whole genome shotgun contigs (wgs, http://www.ncbi.nlm.nih.gov/genbank/wgs) using tblastn. Significant hits were manually predicted using AUGUSTUS [245] and GENESCAN [244]. In total, we used 152 MadBub homologs (Sequence File S1). The TPR domains of 148 sequences were aligned using MAFFT-LINSI [212]; only columns with 80% occupancy were considered for further analysis. Phylogenetic analysis of the resulting multiple sequence alignment was performed using RAxML [214] (Figure S1a). Model selection was performed using Prot Test [306] (Akaike information criterion): LG + G was chosen as the evolutionary model.

Conserved feature extraction and subfunctionalization analysis

ConFeaX starts with a probabilistic search for short conserved regions (max. 50) using the MEME algorithm (option: any number of repeats) [218]. Significant motif hits are extended on both sides by five residues to compensate for the strict treatment of alignment information by the MEME algorithm. Next, MAFFT-LINSI [212]introduces gaps and the alignments are modeled using the HMMER package [235] and used to search for hits that are missed by the MEME algorithm. Subsequent alignment and HMM searches were iterated until convergence. For SLiMs with few conserved positions, specific optimization of the alignments and HMM models using permissive E-values/bit-scores was needed (e.g. ABBA motif and D-box). Sequence logos were obtained using weblogo2 [247]. Subsequently, from each of the conserved features, a phylogenetic profile was derived (present is '1' and absent is '0') for all duplicated MadBub sequences as presented in figure 1. For all possible pairs, we determined the correlation using Pearson correlation coefficient [219]. Average clustering based on Pearson distances ($d = 1 - r$) was used to indicate subfunctionalization.

Cell culture, transfection and plasmids

HeLa-FlpIn TRex cells were grown in DMEM high glucose supplemented with 10% Tet-free FBS (Clontech), penicillin/streptomycin (50 mg ml^{-1}) and alanyl-glutamine (Sigma; 2 mM). pcDNA5-constructs were co-transfected with pOgg44 recombinase in a 10:1 ratio [257] using FuGEHE HD (Roche) as a transfection reagent. After transfection, the medium was supplemented with puromycin (1 µg ml^{-1}) and blasticidin (8 µg ml^{-1}) until cells were fully confluent in a 10 cm culture dish. siBubR1 (5'-AGAUCCUGGCUAACUGUUCUU-3'

custom Dharmacon) was transfected using Hiperfect (Qiagen) at 40 nM for 48 hours, according to the manufacturer's guidelines. RNAi-resistant LAP (YF-P)-BubR1 was sub-cloned from plC58 [302] into pcDNA5.1-puro using AflII and BamHI restriction sites. To acquire mutants, site-directed mutagenesis was performed using the quickchange strategy (for primer sequences see table S3).

Live cell imaging

For live cell imaging experiments, the stable HeLa-FlpIn-TRex cells were transfected with 40 nM siRNA (start and at 24 hours). After 24 hours, the medium was supplemented with thymidine (2.5 mM) and doxycyclin (2 μg ml^{-1}) for 24 hours to arrest cells in early S-phase and to induce expression of the stably integrated construct, respectively. After 48 hours, cells were released for 3 hours and arrested in prometaphase of the mitotic cell cycle (after approximately 8–10 hours) by the addition of the Eg5 inhibitor S-trityl-l-cysteine (STLC, 20 μM). HeLa cells were imaged (DIC) in a heated chamber (37°C, 5% CO2) using a CFI S Plan Fluor ELWD 20x/NA 0.45 dry objective on a Nikon Ti-Eclipse wide field microscope controlled by NIS software (Nikon). Images were acquired using an Andor Zyla 4.2 sCMOS camera and processed using NIS software (Nikon) and ImageJ.

Immunoprecipitation and western blot

HeLa-FlpIn-TRex cells were induced with doxycyclin (2 μg ml^{-1}) 48 hours before harvesting. Synchronization by thymidine (2 mM) for 24 hours and release for 10 hours into Taxol (2 μM) arrested cells in prometaphase. Cells were collected by mitotic shake-off. Lysis was done in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5% NP40, 1 mM EDTA, 1 mM DTT, protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktails 2 and 3 (Sigma). Complexes were purified using GFP-Trap beads (ChromoTek) for 15 min at 4°C. Precipitated proteins were washed with lysis buffer and eluted in 5× SDS sample buffer. Primary antibodies were used at the following dilutions for western blotting: BubR1 (A300-386A, Bethyl) 1:2000, alpha-tubulin (T9026, Sigma) 1:5000, GFP (Custom) 1:10 000, APC1 (A301-653A, Bethyl) 1:2500, APC3 (gift from Phil Hieter) 1:2000, Mad2 (Custom) 1:2000, Cdc20 (A301-180A, Bethyl) 1:1000. Western blot signals were detected by chemiluminescence using an ImageQuant LAS 4000 (GE Healthcare) imager.

Contributions

ET performed the motif search, phylogenetic analysis and SAC assays. DB performed the immunoprecipitation and western blot analyses. GJPLK and BS conceived and managed the project. ET, BS and GJPLK wrote the manuscript.

Acknowledgements

The authors thank the Snel and Kops labs for discussion and feedback. We thank Bas de Wolf and Laura Demmers for making cell lines. We specifically thank Jolien van Hooff for extensive discussions.

Electronic Supplementary Material

Supplementary materials for this chapter are made available online at the following link http://bioinformatics.bio.uu.nl/eelco/thesis/

Sequence file S1 harbors the full-length sequences of all MadBub orthologs used in this study.

Sequence file S2 contains the subsequences of conserved features of the MadBub gene family detected by ConFeaX + TPR and kinase domain.

Procedures and Discussion this section contains a detailed description and discussion of the phylogenetic analyses, motif searches and evolutionary reconstructions associated to Figure 1, Figure S1 and Figure 2.

Figure S1 contains an overview of phylogenetic analysis of the MadBub family as presented in Figure 1.

Figure S2 shows a multiple sequence alignment of the ABBA1-KEN2-ABBA2 cassette in all species used in this study.

Table S1 is a conversion table for species IDs used in this study and their corresponding full species names.

Table S2 is a matrix of all conserved features in all MadBub orthologs.

Table S3 contains an overview of the primers used for molecular cloning.

CHAPTER 7

Discussion

Discussion

This thesis

In this thesis we set out to perform comparative genomics studies in combination with in-depth comparative sequence analyses to illuminate the molecular evolution and function of the kinetochore network in eukaryotes. Our work in chapter 2 represents the first large-scale systematic study of kinetochore-associated proteins in a wide variety of eukaryotic genomes and provides a framework for the (future) interpretation of the evolutionary dynamics of the kinetochore. Our evolutionary reconstructions imply that LECA likely possessed a complex kinetochore network and that the kinetochores of different lineages strongly diverged through extensive gene loss in combination with recurring duplications and occasional inventions and even displacement. Striking examples include the mutual exclusive presence-absence patterns of the microtubule plus-end tracking Dam1 and Ska complex (see also [87]), the recurrent loss of the largest part of a 16-subunit inner kinetochore complex (CCAN) that is essential in vertebrates and fungi and the degenerate composition of kinetochores in various alveolate and excavate lineages. Degeneration occurred occasionally to such extent that Trypanosoma brucei is devoid of any conventional kinetochore subunit and utilizes an analogous system (see also [307]). These patterns hold even when we perform an in depth check on gene prediction problems, despite this check revealing many individual absences to be false negative (chapter 3).

To harness the wealth of sequence data of our manually determined ortholog sets and to capture patterns of highly divergent sequence evolution that are characteristic of many kinetochore subunits, we developed a *de novo* sequence discovery workflow (ConFeaX) to trace the eukaryote-wide (co-)evolution of short linear motifs, domains and proteins of the kinetochore network (chapter 3). This versatile approach proved useful as the advanced detection of conserved elements and their (co-)evolutionary reconstructions provided us with testable hypotheses on various aspects of eukaryotic kinetochore biology. For example, ConFeaX guided us to uncover functional short linear motifs for Knl1, BubR1 and Spindly in human cells (see chapter 3-6).

More genomes, more proteins and technological challenges

Although our comparative analyses revealed many aspects of eukaryotic kinetochore function and evolution, several improvements and additional analyses are warranted. Our genome set is inherently biased towards opisthokont lineages, known model organisms and (parasitic) animal pathogens, because of the focus on benefit to humans in the decision which genomes to sequence. Our set therefore lacks genomes that represent key positions in the eukaryotic

tree. Recent technological advances in the field of single cell sequencing and large-scale projects such as The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) however, now provide transcriptomics data on an ever-growing number of (unculturable) eukaryotes [308]. The massive influx of such new sequencing data currently results in the ongoing revision of eukaryotic phylogeny and makes it possible to determine which lineages diverged close to LECA. For example, recent evidence suggests that Excavates are not monophyletic and indicate that *Malawimonas* and *Collodictyon* are unikonts and cluster together with opisthokonts and amoebozoa [159]. The addition of new genomes and transcriptomes may therefore not only provide a better view on eukaryotic diversity but also allow for an improved reconstruction of the LECA kinetochore. The collection of lineage-specific genome sets will allow for the reconstruction of key evolutionary events that pertain to a specific clade. For example, it would be interesting to collect an excavate-specific genome set to uncover the evolutionary history of the kinetochore in trypanosomes and shed light on the origins of its ~20 unconventional subunits [165,166].

The compilation of large eukaryotic sequence databases opens up the possibility for the use of methods like GREMLIN [309] and EVFOLD [310], which utilize sequence co-variation to predict residue-residue contacts of proteins. These algorithms have been used for the *de novo* prediction of various structures and co-evolving protein subunits in prokaryotes [311,312] and may well aid in the characterization of structural features of currently understudied kinetochore proteins. Although our methods are heavily dependent on manual curation, and therefore not geared towards the handling of large datasets, the establishment of specific HMM profiles for each orthologous kinetochore protein (chapter 2) should allow for the rapid detection and characterization of kinetochore composition in a large variety of eukaryotes. In addition, the increase in the number of genomes and sequences will likely result in the establishment of more informative HMM profiles that will aid in the detection of highly diverged homologs. A potential danger of a large database size is the inability to sufficiently track errors such as incomplete gene prediction, which may result in the incorrect assignment of absences and an overestimation of gene/motifs loss. Our preliminary analyses in chapter 3 indicated such issues in ~20% of the genes upon carefully curation of a smaller training set. Future studies are therefore warranted to assess to what extent gene prediction problems may influence evolutionary inferences.

To inclusively model the (co-)evolution of conserved features of the kinetochore in eukaryotes, our ConFeaX workflow should be further applied to the orthologs of the ~80 kinetochore proteins that we determined in this thesis. While our pipeline is specifically tailored towards the characterization of rapidly evolving protein families, kinetochore subunits that consist of one structur-

al domain, such as Mad2 (HORMA) and Bub3 (WD40), are not suitable to our method and can be analyzed with existing alignment-based conservation tools such as phyloHMM [232] and SLiMprints [231]. Furthermore, our analyses should be supplemented with a number of kinetochore protein complexes that we did not include in our studies e.g. CenpA loading factors such as, Mis18bp, Mis18 □/□ and HJURP, and modulators of microtubule dynamics at the kinetochores, like Stu2, MCAK and ch-TOG. Over time, similar analysis of molecular systems such as condensin/cohesin, Cyclin-Cdks and the large variety of spindle-associated factors should provide the basis for an integrated model of chromosome segregation and cell division in eukaryotes.

In the remainder of this chapter, I would like to outline two potentially interesting venues for future research, which could be made possible by the developments described in this thesis.

Origins of the kinetochore

Duplications of genes and whole genomes have had a profound impact on the evolution of function in eukaryotes and are generally associated with an increase in molecular network complexity [3]. Various eukaryotic protein complexes and pathways are characterized by a high number of subsequent duplications that underlie their present-day appearances e.g. vesicle-tethers and coatamers that constitute organellar membrane trafficking [11], small GTPases [313] and parts of the spliceosomal machinery [314]. As such, a large part of the comparative genomics endeavor is specifically aimed at the reconstruction of gene duplication events to infer the origins and ongoing evolutionary trajectories of molecular systems. While work in this thesis was mostly focused on the evolution of the kinetochore after the divergence of LECA, our reconstruction of its composition reveals a remarkable amount of subunits with similar structural domains, suggesting that the kinetochore is likely the result of a number of duplications (ancient paralogs) that occurred before LECA (Figure 1) and that the contours of a simpler more primitive kinetochore might be delineated.

Duplications within the kinetochore include common eukaryotic domains: the microtubule-binding Calponin Homology (CH) domain (Ndc80 and Nuf2), the DNA-interacting histone fold (CenpT, CenpW, CenpS, CenpX and CenpA), the RWD domain, which is either present in a single (Mad1, Csm1, Spc24, Spc25) –or a double configuration (Knl1, Zwint-1, CenpO and CenpP), the HORMA domain (Mad2 and p31^{comet}), the WD40 domain (Bub3 and Cdc20), the kinase domain (Plk1, Aurora, MadBub and Mps1) and various internal complex duplications e.g. the Mis12- [47,46], Dam1- and Ska complex [87]. Using a method like scroll saw [313], eukaryote-wide phylogenies of these domains may illuminate the evolutionary trajectories of the aforementioned kinetochore

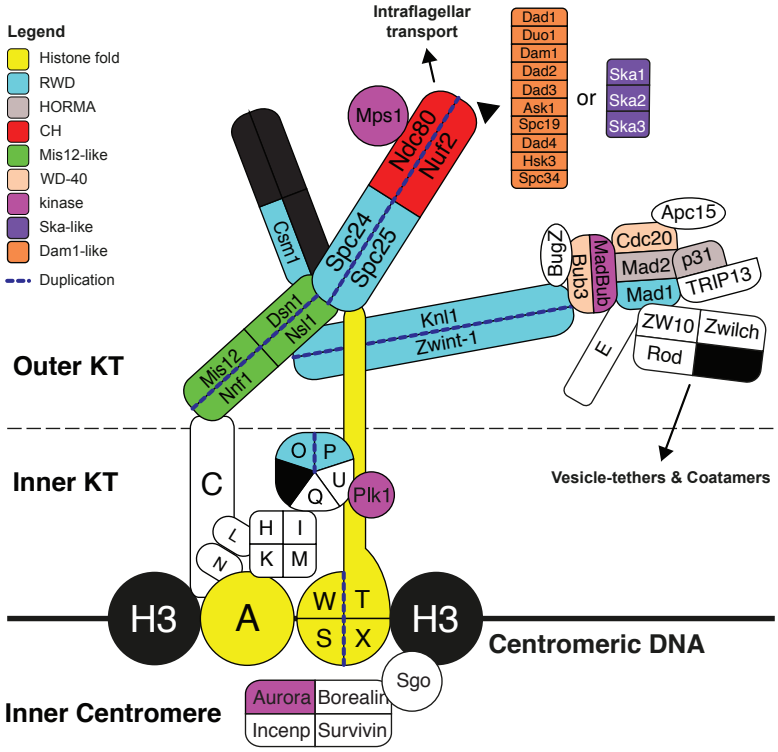


Figure 1 Pre-LECA duplications illuminate the origins of the kinetochore. Cartoon of the kinetochore composition that was inferred to be present in LECA (see Chapter 2,3). Colors of the proteins indicate the shared presence of common structural domains in kinetochore proteins. Black depicts non-LECA subunits that are essential for the function of the associated subcomplexes in lineages in which they are studied (Csm1 – Monopolin complex and Rod-Zwilch-ZW10 – Spindly). The presence of homologous domains points to ancient duplications for the origins of kinetochore complexity. The dashed line delineates a tentative internal duplication of kinetochore subunits that are part of the main centromere-microtubule axis and form obligate dimers. The CH domains of Ndc80:Nuf2 and the RZZ complex are homologous to proteins that are involved in intraflagellar transport and vesicle-tethering systems, respectively.

subunits. Evolutionary reconstructions will likely indicate either one or both models for the origins of the kinetochore: (1) kinetochore complexity arose through internal duplications and (2) the components of the kinetochore originate from duplications of cellular systems of distinct origin. Strikingly, for kinetochore subunits that are part of main centromere-microtubule axis, their domains (RWD, histone, CH and Mis12-like) form obligate heterodimers. Without any prior analyses it is therefore tempting to speculate that at least one internal duplication gave rise to kinetochore complexity in LECA (see dashed line in Figure 1). Furthermore, a quick scan of the literature on the cellular functions of eukaryote-wide homologous genes provides a number of potential hypotheses on the mixed origin of the kinetochore. The RWD domain is

present in a large variety of E3 ubiquitin ligases and the kinase Gcn2, which is involved in starvation signalling [315–317]. The E3 ubiquitin ligase FancL contains an RWD domain and is a member of the fanconi anemia (FA) pathway, which is involved in DNA damage repair [318]. Strikingly, two bona fide kinetochore proteins CenpS and CenpX are also part of the FA pathway [188], possibly indicating a shared ancestry between this molecular system and the kinetochore. Interestingly, the CH domain of Ndc80 was found to be similar to that of several microtubule-binding proteins involved in intraflagellar transport [319]. Last, the RZZ complex is homologous to vesicle tethering systems of which multiple subsequent duplications facilitated the extensive divergence of internal membrane systems in eukaryotes e.g the COPII complex and the nuclear pore [11,320].

Altogether these homologies suggest that the kinetochore is of mixed origin, implicating three major innovations in eukaryotes -the flagellum, the chromatin and internal membrane sorting systems- as its potential ancestors. Subsequent intra process duplications and a few gene inventions would then have expanded the kinetochore to the composition that we infer to have been present in LECA. This would also suggest that during eukaryogenesis mitosis might only have emerged after these other ancestors had already been established, at least in primitive form. Detailed reconstruction of the pre-LECA duplications will be needed to inform with more certainty on the origins of the kinetochore and may add to a working model for the evolution of mitotic and meiotic cell division in the developing ancestor of eukaryotes.

Comparative molecular cell biology of kinetochores in eukaryotes

A striking observation from our studies is that kinetochore compositions are highly diverse across eukaryotic species. Absences of complexes or protein features perceived as indispensable in fungal and metazoan lineages defy common intuitions on the strong evolutionary conservation of essential cellular components, indicating the limited capacity of studies of classical model organisms to illuminate the extent of eukaryotic molecular diversity. Similarly, although microtubule-based chromosome segregation is conserved, there is a remarkable large variation in many aspects of its mechanisms in eukaryotic cells [321]. Differences exist in features that pose numerous potential functional challenges to kinetochore systems. These include the fate and integrity of the nuclear envelope (breakdown vs. maintenance) [161,164], the variable geometry (single orthogonal axis vs. bilateral axis) [322] and position (intra- vs. extranuclear) [322] of the spindle apparatus throughout mitosis, the nature of its microtubule-organizing centers (centrioles-based vs. others) [323], and the chromosomal distribution of centromeres (point-, holo- and regional centromeres). Interestingly, the molecular underpinnings for these differences are

largely unknown. In light of such extensive differences in kinetochore compositions and mitotic mechanisms, numerous fundamental questions arise. What drives kinetochore diversity? How is kinetochore function wired in diverse species? How do the different architectures enable kinetochore function? What are the functional differences between kinetochores and what are the consequences of this for chromosome segregation and cell division?

To answer these questions, a first step is to set up a comparative proteomics study to molecularly interrogate the biochemical composition of kinetochores in a number of unexplored and diverse eukaryotic species. The outcomes of these studies will illuminate whether divergent kinetochore compositions merely represent 'simplified' architectures that are characterized by various absences of 'conventional' (opisthokont-like) complexes or that these kinetochores are in part or completely made up of unconventional analogous proteins (e.g. in trypanosomes). Extensive functional characterization and evolutionary genomics analysis of newly discovered eukaryotic kinetochore components in combination with the categorization of diverse mitotic and cellular features will present new opportunities for more complete comparative analyses and provide a good starting point to assess the true extent of the molecular and evolutionary flexibility of the kinetochore in eukaryotes.

Crucial for the success of these comparative analyses will be the selection of species. While our comparative genomics analysis provide a solid basis for such a selection, species should generally fulfill the following criteria: (1) sufficient absence of conventional kinetochore complexes, (2) part of an understudied clade with respect to kinetochore research, (3) divergent mechanisms of chromosome segregation and (4) susceptibility to genetic manipulation. While most studies have focused on opisthokont lineages, the selection of species from the vast majority of other eukaryotic supergroups seems most logical, these include plants, algae, amoeba, ciliates and other single-celled protozoans. Although their kinetochore compositions show limited variation, various uncommon model species from these clades are experimentally tractable, including the green algae *Chlamydomonas reinhardtii*, the amoebazoan *Dictyostelium discoideum*, and the moss *Physcomitrella patens*.

The most interesting candidate species for future studies are various alveolate species, as both their kinetochores and mitotic mechanisms are strongly divergent (Figure 2). The ciliate *Tetrahymena thermophila* (Figure 2a), has been used as a model organism for many decades and harbors two distinct nuclei [324]. During asexual cell division, only the epigenetically silenced micronucleus undergoes mitosis and was found to contain CenpA-containing centromeres [325]. The macronucleus consists of multiple copies of the micro nuclear genome and divides through random fission. The intracellular apicomplexan

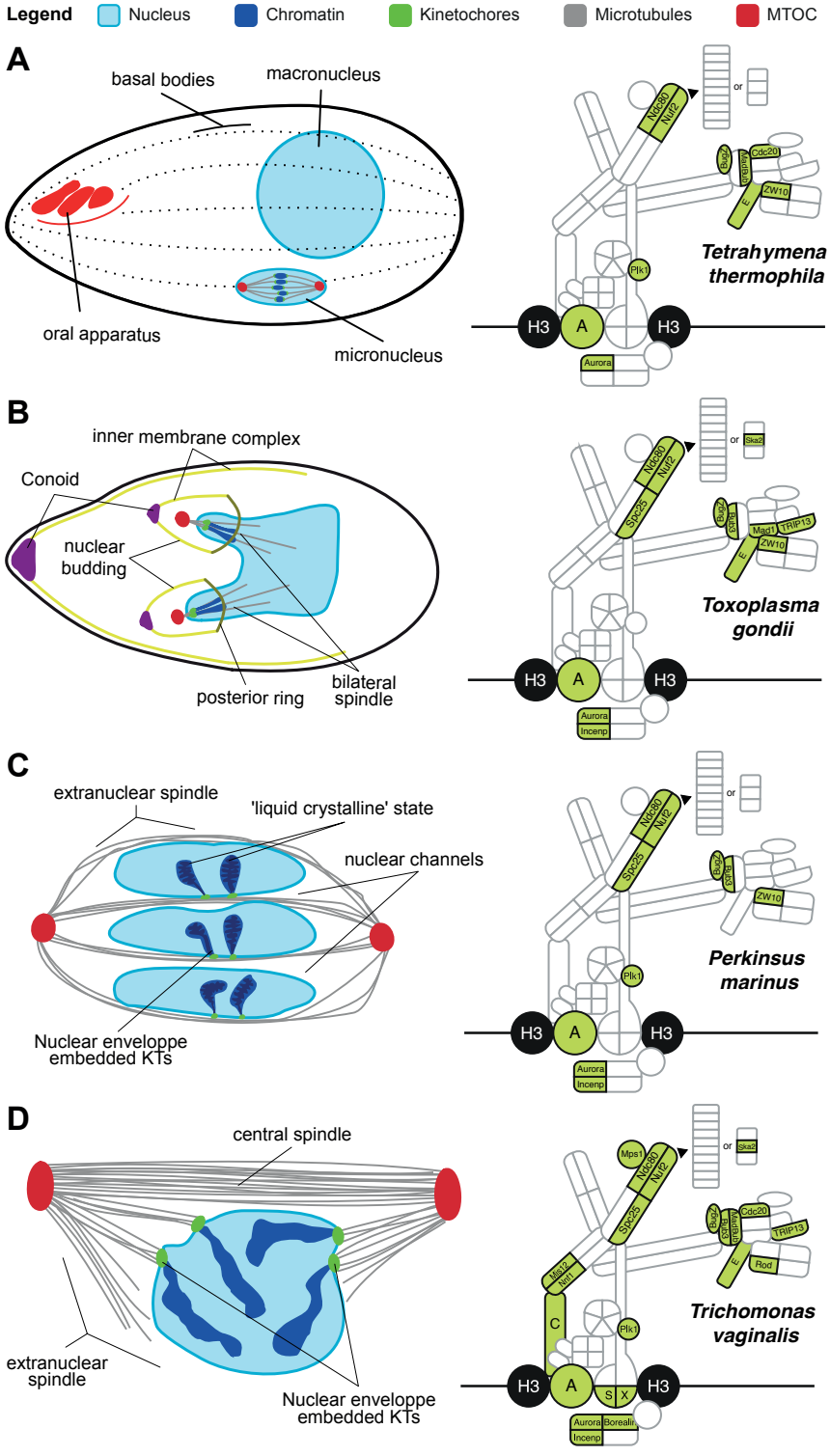

Legend
Nucleus
Chromatin
Kinetochores
Microtubules
MTOC
A
basal bodies
macronucleus
oral apparatus
micronucleus
Ndc80
Nuf2
or
Cdc20
Mad1
ZW10
E
Plk1
H3
A
H3
Aurora
Tetrahymena thermophila
B
inner membrane complex
Conoid
nuclear budding
posterior ring
bilateral spindle
Ska2
Spc25
Bub2
Bub1
TRIP13
Toxoplasma gondii
Incenp
C
extranuclear spindle
'liquid crystalline' state
nuclear channels
Nuclear enveloppe embedded KTs
Perkinsus marinus
D
central spindle
extranuclear spindle
Nuclear enveloppe embedded KTs
Mps1
Hec1
Nuf1
C
S
X
Borealin
Rod
Trichomonas vaginalis

Figure 2 Alveolate -and excavate species contain divergent kinetochore compositions and perform chromosome segregation in distinct ways. The cartoons on the left depict an interpretation of EM studies of metaphase/anaphase in various organisms and highlight features that characterize chromosome segregation mechanisms and cellular physiology in these species (note: proportions may deviate from the original data). Recurring colors indicate the position of structures involved in chromosome segregation (MTOC: microtubule-organizing center). The panels on the right show the kinetochore composition, which is similar to that of Figure 1. Green/white subunits indicate the presence and absence of subunits, respectively. **(A)** *Tetrahymena thermophila* is a ciliate that contains two nuclei (micro, macro). The micronucleus undergoes a closed mitosis and the chromosomes of the macronucleus are randomly distributed upon binary fission during cell division [336]. **(B)** Cell division of Apicomplexans (*Toxoplasma gondii*) is characterized by a mechanism known as 'internal budding'. Two new cells form inside the cytoplasm and divide the nucleus and its chromosomes between them. Kinetochores are attached to the centrocone (MTOC) throughout the cell cycle [329]. The spindle has a bilateral geometry and 'true metaphase' is therefore never observed [322,326]. **(C)** Dinoflagellates are one of the strangest eukaryotic lineages (here depicted: a cartoon of *Crypthecodinium cohnii* [331]). In EM studies, their chromosomes are described as being in a 'liquid crystalline' state. During mitosis, microtubules of an extranuclear spindle run through channels in the nucleus and connect to kinetochores that are embedded with the nuclear envelope. *Perkinsus marinus* (a sister-clade of dinoflagellates) is genetically tractable and has similar extranuclear spindles and nuclear envelope-embedded kinetochores [333] (composition depicted on the right). (**D**) Various excavate lineages, including the human parasite *Trichomonas vaginalis*, have independently evolved an extranuclear spindle and nuclear envelope-embedded kinetochores. A prominent feature of their mitotic figures is a large central spindle [322,334,337].

parasites *Toxoplasma gondii* (Figure 2b) and *Plasmodium falciparum* perform genome partioning through a peculiar nuclear budding mechanism [326]. Initial studies have identified CenpA [327,328] and revealed that kinetochores are connected to the spindle pole structure (centrocone) throughout the cell division [329]. Dinoflagellates are perhaps the most interesting group of organisms to study since their genomes are organized in a fundamentally different manner. Their large genomes are divided over liquid crystalline chromosomes that are in a permanently condensed state. Strikingly, members of this lineage replaced histone proteins with dinoflagellate viral nucleoproteins DNVP that likely act as the main packaging factor of the genome [330]. In addition, EM studies have revealed that their kinetochores are embedded within the nuclear envelope and connect to microtubules of an extranuclear spindle-like structure that run through channels that traverse an intact nucleus [322,331]. Although dinoflagellate species are currently not amenable to genetic manipulation, recent technical advances now allow for the molecular interrogation of dinoflagellate-related lineages such as *Perkinsus marinus* [332] (Figure 2c), which also inserts its kinetochores in the nuclear membrane [333]. Strikingly, the excavate lineage of the Parabasalids, containing species such as *Trichomonas vaginalis* (Figure 2d), have independently acquired nuclear membrane-embedded kinetochores and may provide an alternative model to study the function and composition of such odd kinetochores [334,335].

Concluding remarks

Over the last two decades, the rapid increase in the number of available genomes revolutionized our ability to reconstruct evolutionary events that have driven the molecular complexity of eukaryotic cellular systems. This thesis highlights such developments and showcases the use of comparative genomics and comparative sequence analyses to uncover functional aspects of kinetochore biology in eukaryotes. While kinetochores are at the heart of chromosome segregation in eukaryotes, extensive compositional diversity as well as the rapid sequence evolution of its subunits is a recurring theme in this thesis and presents us with a fundamental paradox: how can kinetochores be essential and divergent at the same time? Many have sought explanations in the form of an adaptive conflict e.g. within genomes (transposable elements), host-pathogen interactions (viruses) and even between sexes (meiotic drive). Perhaps the best explanation for this phenomenon is the centromere drive hypothesis that poses an evolutionary battle between centromeres and kinetochores for the selection of chromosomes that will end up in gametes, eventually causing reproductive boundaries between species. As such, kinetochores drive speciation and its inherently instable nature may well be one of the main components for the origin of eukaryotic diversity. Characterization of kinetochore function in mitosis and meiosis in a large diversity of eukaryotes will therefore not only benefit our understanding of diseases such as cancer, but holds great promise to illuminate our understanding of the evolutionary process itself.

Addendum

References
Samenvatting in het Nederlands
Curriculum Vitae
List of publications
Dankwoord

References

1. Venter JC. 2001 The Sequence of the Human Genome. *Science* **291**, 1304–1351.
2. Lander ES *et al*. 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
3. Van de Peer Y, Mizrachi E, Marchal K. 2017 The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424.
4. Ku C *et al*. 2015 Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* **524**, 427–432.
5. Albalat R, Cañestro C. 2016 Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391.
6. Soucy SM, Huang J, Gogarten JP. 2015 Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482.
7. Adl SM *et al*. 2012 The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–493.
8. Koonin EVE V. *et al*. 2010 The Incredible Expanding Ancestor of Eukaryotes. *Cell* **140**, 606–608.
9. Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Bettencourt-Dias M. 2011 Evolution: Tracing the origins of centrioles, cilia, and flagella. *J. Cell Biol.* **194**, 165–75.
10. Dacks JB, Field MC, Buick R, Eme L, Gribaldo S, Roger AJ, Brochier-Armanet C, Devos DP. 2016 The changing view of eukaryogenesis – fossils, cells, lineages and how they all come together. *J. Cell Sci.* **129**, 3695–3703.
11. Rout MP, Field MC. 2017 The Evolution of Organellar Coat Complexes and Organization of the Eukaryotic Cell. *Annu. Rev. Biochem.* **86**, 637–657.
12. Aravind L, Watanabe H, Lipman DJ, Koonin E V. 2000 Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci.* **97**, 11319–11324.
13. Morgan DO. 2006 *The Cell Cycle: Principles of Control*. London: New Science Press Ltd.
14. Melters DP *et al*. 2013 Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10.
15. Malik HS, Henikoff S. 2009 Major Evolutionary Transitions in Centromere Complexity. *Cell* **138**, 1067–1082.
16. Petry S. 2016 Mechanisms of Mitotic Spindle Assembly. *Annu. Rev. Biochem.* **85**, 659–683.
17. Prosser SL, Pelletier L. 2017 Mitotic spindle assembly in animal cells: a fine balancing act. *Nat. Rev. Mol. Cell Biol.* **18**, 187–201.
18. Musacchio A, Desai A. 2017 A Molecular View of Kinetochore Assembly and Function. *Biology (Basel)*. **6**, E5.
19. Maiato H, Gomes AM, Sousa F, Barisic M. 2017 Mechanisms of Chromosome Congression during Mitosis. *Biology (Basel)*. **6**, 1–13.
20. Sacristan C, Kops GJPL. 2015 Joined at the hip: kinetochores, microtubules, and spindle assembly checkpoint signaling. *Trends Cell Biol.* **25**, 21–28.
21. Steiner FA, Henikoff S. 2014 Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *Elife* **3**, e02025.
22. Melters DP, Paliulis L V, Korf IF, Chan SWL. 2012 Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosom. Res.* **20**, 579–593.
23. Maddox PS, Oegema K, Desai A, Cheeseman IM. 2004 'Holo'er than thou: Chromosome segregation and kinetochore function in C. elegans. *Chromosom. Res.* **12**, 641–653.

24. Drinnenberg IA, DeYoung D, Henikoff S, Malik HS. 2014 Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife* **3**.
25. Heckmann S, Jankowska M, Schubert V, Kumke K, Ma W, Houben A. 2014 Alternative meiotic chromatid segregation in the holocentric plant Luzula elegans. *Nat. Commun.* **5**, 4979.
26. McKinley KL, Cheeseman IM. 2016 The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **17**, 16–29.
27. Rosin L, Mellone BG. 2016 Co-evolving CENP-A and CAL1 Domains Mediate Centromeric CENP-A Deposition across Drosophila Species. *Dev. Cell* **37**, 136–147.
28. Malik HS, Bayes JJ. 2006 Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. *Biochem. Soc. Trans.* **34**, 569–573.
29. Henikoff S, Ahmad K, Malik HS. 2001 The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102.
30. Talbert PB, Bryson TD, Henikoff S. 2004 Adaptive evolution of centromere proteins in plants and animals. *J. Biol.* **3**, 18.
31. Krassovsky K, Henikoff JG, Henikoff S. 2012 Tripartite organization of centromeric chromatin in budding yeast. *Proc. Natl. Acad. Sci.* **109**, 243–248.
32. Masumoto H. 1989 A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* **109**, 1963–1973.
33. Gascoigne KE, Cheeseman IM. 2013 CDK-dependent phosphorylation and nuclear exclusion coordinately control kinetochore assembly state. *J. Cell Biol.* **201**, 23–32.
34. Gascoigne KE, Takeuchi K, Suzuki A, Hori T, Fukagawa T, Cheeseman IM. 2011 Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* **145**, 410–422.
35. Cheeseman IM, Desai A. 2008 Molecular architecture of the kinetochore–microtubule interface. *Nat. Rev. Mol. Cell Biol.* **9**, 33–46.
36. Weir JR *et al.* 2016 Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature* **537**, 249–253.
37. Nishino T, Takeuchi K, Gascoigne KE, Suzuki A, Hori T, Oyama T, Morikawa K, Cheeseman IM, Fukagawa T. 2012 CENP-T-W-S-X forms a unique centromeric chromatin structure with a histone-like fold. *Cell* **148**, 487–501.
38. McKinley KL, Sekulic N, Guo LY, Tsinman T, Black BE, Cheeseman IM. 2015 The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol. Cell* **60**, 886–898.
39. Bancroft J, Auckland P, Samora CP, McAinsh AD. 2015 Chromosome congression is promoted by CENP-Q- and CENP-E-dependent pathways. *J. Cell Sci.* **128**, 171–184.
40. Kang YH, Park CH, Kim TS, Soung NK, Bang JK, Kim BY, Park JE, Lee KS. 2011 Mammalian polo-like kinase 1-dependent regulation of the PBIP1-CENP-Q complex at kinetochores. *J. Biol. Chem.* **286**, 19744–19757.
41. Li Y *et al.* 2015 A versatile reporter system for CRISPR-mediated chromosomal rearrangements. *Genome Biol.* **16**, 111.
42. Lee KS, Oh D-Y, Kang YH, Park J-E. 2008 Self-regulated mechanism of Plk1 localization to kinetochores: lessons from the Plk1-PBIP1 interaction. *Cell Div.* **3**, 4.
43. Cheeseman IM, Anderson S, Jwa M, Green EM, Kang J seog, Yates JR, Chan CSM, Drubin DG, Barnes G. 2002 Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell.* **111**, 163–172.
44. Tirupataiah S, Jamir I, Srividya I, Mishra K. 2014 Yeast Nkp2 is required for accurate chromosome segregation and interacts with several components of the central kinetochore. *Mol. Biol. Rep.* **41**, 1–11.

45. Liu X, McLeod I, Anderson S, Yates JR, He X. 2005 Molecular analysis of kinetochore architecture in fission yeast. *EMBO J.* **24**, 2919–2930.
46. Petrovic A *et al.* 2016 Structure of the MIS12 Complex and Molecular Basis of Its Interaction with CENP-C at Human Kinetochores. *Cell* **167**, 1028–1040.e15.
47. Dimitrova YN, Jenni S, Valverde R, Khin Y, Harrison SC. 2016 Structure of the MIND Complex Defines a Regulatory Focus for Yeast Kinetochore Assembly. *Cell* **167**, 1014–1027.
48. Huis in 't Veld PJ *et al.* 2016 Molecular basis of outer kinetochore assembly on CENP-T. *Elife* **5**, 45–59.
49. Milks KJ, Moree B, Straight AF. 2009 Dissection of CENP-C-directed Centromere and Kinetochore Assembly. *Mol. Biol. Cell* **20**, 4246–4255.
50. Cheeseman IM, Chappie JS, Wilson-Kubalek EM, Desai A. 2006 The Conserved KMN Network Constitutes the Core Microtubule-Binding Site of the Kinetochore. *Cell* **127**, 983–997.
51. Drinnenberg IA, Henikoff S, Malik HS. 2016 Evolutionary Turnover of Kinetochore Proteins: A Ship of Theseus? *Trends Cell Biol.* **26**, 498–510.
52. 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.
53. Ciferri C *et al.* 2008 Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell*. **133**, 427–439.
54. Petrovic A *et al.* 2014 Modular Assembly of RWD Domains on the Mis12 Complex Underlies Outer Kinetochore Organization. *Mol. Cell* **53**, 591–605.
55. Valverde R, Ingram J, Harrison SC. 2016 Conserved Tetramer Junction in the Kinetochore Ndc80 Complex. *Cell Rep.* **17**, 1915–1922.
56. Gonen S, Akiyoshi B, Iadanza MG, Shi D, Duggan N, Biggins S, Gonen T. 2012 The structure of purified kinetochores reveals multiple microtubule-attachment sites. *Nat. Struct. & Mol. Biol.* **19**, 925–929.
57. Corbett KD, Harrison SC. 2012 Molecular Architecture of the Yeast Monopolin Complex. *Cell Rep.* **1**, 583–589.
58. Malvezzi F, Litos G, Schleiffer A, Heuck A, Mechtler K, Clausen T, Westermann S. 2013 A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. *EMBO J.* **32**, 409–423.
59. Caldas G V., DeLuca JG. 2014 KNL1: Bringing order to the kinetochore. *Chromosoma* **123**, 169–181.
60. Liu D, Vleugel M, Backer CB, Hori T, Fukagawa T, Cheeseman IM, Lampson M a. 2010 Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase. *J. Cell Biol.* **188**, 809–820.
61. Welburn JPI *et al.* 2010 Aurora B phosphorylates spatially distinct targets to differentially regulate the kinetochore-microtubule interface. *Curr. Biol.* **38**, 399–405.
62. Espeut J, Cheerambathur DK, Krenning L, Oegema K, Desai A. 2012 Microtubule binding by KNL-1 contributes to spindle checkpoint silencing at the kinetochore. *J. Cell Biol.* **196**, 469–482.
63. Kops GJPL, Kim Y, Weaver B a a, Mao Y, McLeod I, Yates JR, Tagaya M, Cleveland DW. 2005 ZW10 links mitotic checkpoint signaling to the structural kinetochore. *J. Cell Biol.* **169**, 49–60.
64. Zhang G, Lischetti T, Hayward DG, Nilsson J. 2015 Distinct domains in Bub1 localize RZZ and BubR1 to kinetochores to regulate the checkpoint. *Nat. Commun.* **6**, 7162.
65. DeLuca JG, Gall WE, Musacchio A, Salmon ED. 2006 Kinetochore Microtubule Dynamics and Attachment Stability Are Regulated by Hec1. *Cell* **127**, 969–982.

66. DeLuca JG, Moree B, Hickey JM, Kilmartin J V., Salmon ED. 2002 hNuf2 inhibition blocks stable kinetochore-microtubule attachment and induces mitotic cell death in HeLa cells. *J. Cell Biol.* **159**, 549–555.
67. Wigge PA, Kilmartin J V. 2001 The Ndc80p Complex from Saccharomyces cerevisiae Contains Conserved Centromere Components and Has a Function in Chromosome Segregation. *J. Cell Biol.* **152**, 349–360.
68. McCleland ML. 2003 The highly conserved Ndc80 complex is required for kinetochore assembly, chromosome congression, and spindle checkpoint activity. *Genes Dev.* **17**, 101–114.
69. Desai A, Rybina S, Müller-Reichert T, Shevchenko A, Shevchenko A, Hyman A, Oegema K. 2003 KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in C. elegans. *Genes Dev.* **17**, 2421–2435.
70. DeLuca JG, Dong Y, Hergert P, Strauss J, Hickey JM, Salmon ED, McEwen BF. 2005 Hec1 and nuf2 are core components of the kinetochore outer plate essential for organizing microtubule attachment sites. *Mol. Biol. Cell* **16**, 519–531.
71. Wei RR, Al-Bassam J, Harrison SC. 2007 The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nat. Struct. Mol. Biol.* **14**, 54–59.
72. Guimaraes GJ, Dong Y, McEwen BF, DeLuca JG. 2008 Kinetochore-Microtubule Attachment Relies on the Disordered N-Terminal Tail Domain of Hec1. *Curr. Biol.* **18**, 1778–1784.
73. Miller SA, Johnson ML, Stukenberg PT. 2008 Kinetochore Attachments Require an Interaction between Unstructured Tails on Microtubules and Ndc80Hec1. *Curr. Biol.* **18**, 1785–1791.
74. Rattner JB, Bazett-Jones DP. 1989 Kinetochore structure: electron spectroscopic imaging of the kinetochore. *J. Cell Biol.* **108**, 1209–1219.
75. McEwen BF, Hsieh CE, Mattheyses AL, Rieder CL. 1998 A new look at kinetochore structure in vertebrate somatic cells using high-pressure freezing and freeze substitution. In *Chromosoma*, pp. 366–375.
76. Cooke CA, Schaar B, Yen TJ, Earnshaw WC. 1997 Localization of CENP-E in the fibrous corona and outer plate of mammalian kinetochores from prometaphase through anaphase. *Chromosoma* **106**, 446–455.
77. Rattner JB, Rao A, Fritzler MJ, Valencia DW, Yen TJ. 1993 CENP-F is a ca 400 kDa kinetochore protein that exhibits a cell-cycle dependent localization. *Cell Motil. Cytoskeleton* **26**, 214–226.
78. Mosalaganti S *et al.* 2017 Structure of the RZZ complex and molecular basis of its interaction with Spindly. *J. Cell Biol.* **216**, 961–981.
79. Gama JB *et al.* 2017 Molecular mechanism of dynein recruitment to kinetochores by the Rod–Zw10–Zwilch complex and Spindly. *J. Cell Biol.* **216**, 943–960.
80. Friese A, Faesen AC, Huis In 't Veld PJ, Fischböck J, Prumbaum D, Petrovic A, Raunser S, Herzog F, Musacchio A. 2016 Molecular requirements for the inter-subunit interaction and kinetochore recruitment of SKAP and Astrin. *Nat. Commun.* **7**, 11407.
81. Miller MP, Asbury CL, Biggins S. 2016 A TOG protein confers tension sensitivity to kinetochore-microtubule attachments. *Cell* **165**, 1428–1439.
82. Domnitz SB, Wagenbach M, Decarreau J, Wordeman L. 2012 MCAK activity at microtubule tips regulates spindle microtubule length to promote robust kinetochore attachment. *J. Cell Biol.* **197**, 231–237.
83. Westermann S, Wang H-W, Avila-Sakar A, Drubin DG, Nogales E, Barnes G. 2006 The Dam1 kinetochore ring complex moves processively on depolymerizing microtubule ends. *Nature* **440**, 565–569.
84. Lampert F, Hornung P, Westermann S. 2010 The Dam1 complex confers microtubule

plus end–tracking activity to the Ndc80 complex. *J. Cell Biol.* **189**, 641–649.
85. Schmidt JC *et al*. 2012 The Kinetochore-Bound Ska1 Complex Tracks Depolymerizing Microtubules and Binds to Curved Protofilaments. *Dev. Cell* **23**, 968–980.
86. Lampert F, Mieck C, Alushin GM, Nogales E, Westermann S. 2013 Molecular requirements for the formation of a kinetochore-microtubule interface by Dam1 and Ndc80 complexes. *J. Cell Biol.* **200**, 21–30.
87. 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.
88. Magidson V, O'Connell CB, Lon□arek J, Paul R, Mogilner A, Khodjakov A. 2011 The spatial arrangement of chromosomes during prometaphase facilitates spindle assembly. *Cell* **146**, 555–567.
89. Carmena M, Wheelock M, Funabiki H, Earnshaw WC. 2012 The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat. Rev. Mol. Cell Biol.* **13**, 789–803.
90. Kim Y, Holland AJ, Lan W, Cleveland DW. 2010 Aurora kinases and protein phosphatase 1 mediate chromosome congression through regulation of CENP-E. *Cell* **142**, 444–455.
91. Chan YW, Jeyaprakash a. A, Nigg E a., Santamaria A. 2012 Aurora B controls kinetochore-microtubule attachments by inhibiting Ska complex-KMN network interaction. *J. Cell Biol.* **196**, 563–571.
92. Chan YW, Nigg E a, Santamaria A. 2012 Aurora B controls kinetochore-microtubule attachments by modulating the interaction between the Ska complex and the KMN network . *Growth (Lakeland)* **196**, 563–571.
93. Liu D, Vader G, Vromans MJM, Lampson MA, Lens SMA. 2009 Sensing chromosome bi-orientation by spatial separation of aurora B kinase from kinetochore substrates. *Science* **323**, 1350–1353.
94. Campbell CS, Desai A. 2013 Tension sensing by Aurora B kinase is independent of survivin-based centromere localization. *Nature* **497**, 118–121.
95. Hengeveld RCC, Vromans MJM, Vleugel M, Hadders MA, Lens SMA. 2017 Inner centromere localization of the CPC maintains centromere cohesion and allows mitotic checkpoint silencing. *Nat. Commun.* **8**, 15542.
96. Suijkerbuijk SJEE, Vleugel M, Teixeira A, Kops GJPLPL. 2012 Integration of kinase and phosphatase activities by BUBR1 ensures formation of stable kinetochore-microtubule attachments. *Dev. Cell* **23**, 745–55.
97. Kruse T, Zhang G, Larsen MSY, Lischetti T, Streicher W, Kragh Nielsen T, Bjorn SP, Nilsson J. 2013 Direct binding between BubR1 and B56-PP2A phosphatase complexes regulate mitotic progression. *J. Cell Sci.* **126**, 1086–1092.
98. Xu P, Raetz E a, Kitagawa M, Virshup DM, Lee SH. 2013 BUBR1 recruits PP2A via the B56 family of targeting subunits to promote chromosome congression. *Biol. Open* **2**, 479–86.
99. Sivakumar S, Janczyk P, Qu Q, Brautigam CA, Stukenberg PT, Yu H, Gorbsky GJ. 2016 The human SKA complex drives the metaphase-anaphase cell cycle transition by recruiting protein phosphatase 1 to kinetochores. *Elife* **5**.
100. Zhang Q, Sivakumar S, Chen Y, Gao H, Yang L, Yuan Z, Yu H, Liu H. 2017 Ska3 Phosphorylated by Cdk1 Binds Ndc80 and Recruits Ska to Kinetochores to Promote Mitotic Progression. *Curr. Biol.* **27**, 1477–1484.e4.
101. Musacchio A. 2015 The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Curr. Biol.* **25**, R1002–R1018.
102. Markova K, Uzlikova M, Tumova P, Jirakova K, Hagen G, Kulda J, Nohynkova E. 2016

Absence of a conventional spindle mitotic checkpoint in the binucleated single-celled parasite Giardia intestinalis. *Eur. J. Cell Biol.* **95**, 355–367.
103. Vicente J-J, Cande WZ. 2014 Mad2, Bub3, and Mps1 regulate chromosome segregation and mitotic synchrony in Giardia intestinalis, a binucleate protist lacking an anaphase-promoting complex. *Mol. Biol. Cell* **25**, 2774–2787.
104. Barford D. 2011 Structural insights into anaphase-promoting complex function and mechanism. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 3605–3624.
105. Kulkarni K, Zhang Z, Chang L, Yang J, da Fonseca PCA, Barford D. 2013 Building a pseudo-atomic model of the anaphase-promoting complex. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 2236–2243.
106. Chang L, Zhang Z, Yang J, McLaughlin SH, Barford D. 2014 Molecular architecture and mechanism of the anaphase-promoting complex. *Nature* **17**, 13–17.
107. Pines J. 2011 Cubism and the cell cycle: the many faces of the APC/C. *Nat. Rev. Mol. Cell Biol.* **12**, 427–438.
108. Yamaguchi M *et al.* 2016 Cryo-EM of Mitotic Checkpoint Complex-Bound APC/C Reveals Reciprocal and Conformational Regulation of Ubiquitin Ligation. *Mol. Cell* **63**, 593–607.
109. Chang L, Zhang Z, Yang J, McLaughlin SH, Barford D. 2015 Atomic structure of the APC/C and its mechanism of protein ubiquitination. *Nature* **522**, 450–454.
110. Davey NE, Morgan DO. 2016 Building a Regulatory Network with Short Linear Sequence Motifs: Lessons from the Degrons of the Anaphase-Promoting Complex. *Mol. Cell* **64**, 12–23.
111. Di Fiore B *et al.* 2015 The ABBA Motif Binds APC/C Activators and Is Shared by APC/C Substrates and Regulators. *Dev. Cell* **32**, 358–372.
112. Davenport J, Harris LD, Goorha R. 2006 Spindle checkpoint function requires Mad2-dependent Cdc20 binding to the Mad3 homology domain of BubR1. *Exp. Cell Res.* **312**, 1831–1842.
113. Di Fiore B, Wurzenberger C, Davey NE, Pines J. 2016 The Mitotic Checkpoint Complex Requires an Evolutionary Conserved Cassette to Bind and Inhibit Active APC/C. *Mol. Cell* **64**, 1144–1153.
114. Tromer E, Bade D, Snel B, Kops GJPL. 2016 Phylogenomics-guided discovery of a novel conserved cassette of short linear motifs in BubR1 essential for the spindle checkpoint. *Open Biol.* **6**, 160315.
115. van Leuken R, Clijsters L, Wolthuis R. 2008 To cell cycle, swing the APC/C. *Biochim. Biophys. Acta* **1786**, 49–59.
116. Amon A, Irniger S, Nasmyth K. 1994 Closing the cell cycle circle in yeast: G2 cyclin proteolysis initiated at mitosis persists until the activation of G1 cyclins in the next cycle. *Cell* **77**, 1037–1050.
117. Hagting A. 2002 Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. *J. Cell Biol.* **157**, 1125–1137.
118. Zur A, Brandeis M. 2001 Securin degradation is mediated by fzy and fzr, and is required for complete chromatid separation but not for cytokinesis. *EMBO J* **20**, 792–801.
119. Primorac I, Musacchio A. 2013 Panta rhei: The APC/C at steady state. *J. Cell Biol.* **201**, 177–189.
120. Burton JL, Solomon MJ. 2007 Mad3p, a pseudosubstrate inhibitor of APCCdc20 in the spindle assembly checkpoint. *Genes Dev.* **21**, 655–667.
121. Alfieri C, Chang L, Zhang Z, Yang J, Maslen S, Skehel M, Barford D. 2016 Molecular basis of APC/C regulation by the spindle assembly checkpoint. *Nature* **536**, 1–19.

122. Rieder CL, Cole RW, Khodjakov A, Sluder G. 1995 The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J. Cell Biol.* **130**, 941–948.
123. Collin P, Nashchekina O, Walker R, Pines J. 2013 The spindle assembly checkpoint works like a rheostat rather than a toggle switch. *Nat. Cell Biol.* **15**, 1378–85.
124. Dick AE, Gerlich DW. 2013 Kinetic framework of spindle assembly checkpoint signalling. *Nat. Cell Biol.* **15**, 1370–1377.
125. Etemad B, Kuijt TEF, Kops GJPL. 2015 Kinetochore–microtubule attachment is sufficient to satisfy the human spindle assembly checkpoint. *Nat. Commun.* **6**, 8987.
126. Tauchman EC, Boehm FJ, DeLuca JG. 2015 Stable kinetochore–microtubule attachment is sufficient to silence the spindle assembly checkpoint in human cells. *Nat. Commun.* **6**, 10036.
127. Ji Z, Gao H, Jia L, Li B, Yu H. 2017 A sequential multi-target Mps1 phosphorylation cascade promotes spindle checkpoint signaling. *Elife* **6**, e22513.
128. Faesen AC, Thanasoula M, Maffini S, Breit C, Müller F, van Gerwen S, Bange T, Musacchio A. 2017 Basis of catalytic assembly of the mitotic checkpoint complex. *Nature* **23**, 498–502.
129. Hiruma Y, Sacristan C, Pachis ST, Adamopoulos A, Kuijt T, Ubbink M, von Castelmur E, Perrakis A, Kops GJPL. 2015 Competition between MPS1 and microtubules at kinetochores regulates spindle checkpoint signaling. *Science* **348**, 1264–1267.
130. Ji Z, Gao H, Yu H. 2015 Kinetochore attachment sensed by competitive Mps1 and microtubule binding to Ndc80C. *Science* **348**, 1260–1264.
131. Primorac I, Weir JR, Chiroli E, Gross F, Hoffmann I, van Gerwen S, Ciliberto A, Musacchio A. 2013 Bub3 reads phosphorylated MELT repeats to promote spindle assembly checkpoint signaling. *Elife* **2**.
132. Vleugel M, Omerzu M, Groenewold V, Hadders MA, Lens SMA, Kops GJPL. 2015 Sequential multisite phospho-regulation of KNL1-BUB3 interfaces at mitotic kinetochores. *Mol. Cell* **57**, 824–835.
133. Tromer E, Snel B, Kops GJPL. 2015 Widespread Recurrent Patterns of Rapid Repeat Evolution in the Kinetochore Scaffold KNL1. *Genome Biol. Evol.* **7**, 2383–2393.
134. Overlack K *et al.* 2015 A molecular basis for the differential roles of Bub1 and BubR1 in the spindle assembly checkpoint. *Elife* **4**, e05269.
135. De Antoni A *et al.* 2005 The Mad1/Mad2 Complex as a Template for Mad2 Activation in the Spindle Assembly Checkpoint. *Curr. Biol.* **15**, 214–225.
136. Mapelli M, Massimiliano L, Santaguida S, Musacchio A. 2007 The Mad2 Conformational Dimer: Structure and Implications for the Spindle Assembly Checkpoint. *Cell* **131**, 730–743.
137. Hardwick KG, Johnston RC, Smith DL, Murray AW. 2000 MAD3 encodes a novel component of the spindle checkpoint which interacts with Bub3p, Cdc20p, and Mad2p. *J. Cell Biol.* **148**, 871–882.
138. Fraschini R, Beretta A, Sironi L, Musacchio A, Lucchini G, Piatti S. 2001 Bub3 interaction with Mad2, Mad3 and Cdc20 is mediated by WD40 repeats and does not require intact kinetochores. *EMBO J.* **20**, 6648–6659.
139. Sudakin V, Chan GKT, Yen TJ. 2001 Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *J. Cell Biol.* **154**, 925–936.
140. London N, Biggins S. 2014 Mad1 kinetochore recruitment by Mps1-mediated phosphorylation of Bub1 signals the spindle checkpoint. *Genes Dev.* **28**, 140–152.
141. Vleugel M, Hoek TA, Tromer E, Sliedrecht T, Groenewold V, Omerzu M, Kops GJPL. 2015 Dissecting the roles of human BUB1 in the spindle assembly checkpoint. *J. Cell*

Sci. **128**, 2975–2982.
142. Karess R. 2005 Rod-Zw10-Zwilch: a key player in the spindle checkpoint. *Trends Cell Biol.* **15**, 386–392.
143. Défachelles L, Raich N, Terracol R, Baudin X, Williams B, Goldberg M, Karess RE. 2015 RZZ and Mad1 dynamics in Drosophila mitosis. *Chromosom. Res.* **23**, 333–342.
144. Diaz-Martinez LA, Tian W, Li B, Warrington R, Jia L, Brautigam CA, Luo X, Yu H. 2015 The Cdc20-binding Phe box of the spindle checkpoint protein BubR1 maintains the mitotic checkpoint complex during mitosis. *J. Biol. Chem.* **290**, 2431–2443.
145. Izawa D, Pines J. 2014 The mitotic checkpoint complex binds a second CDC20 to inhibit active APC/C. *Nature* **517**, 631–634.
146. Murray AW, Marks D. 2001 Can sequencing shed light on cell cycling? *Nature* **409**, 844–846.
147. Suijkerbuijk SJE *et al.* 2012 The Vertebrate Mitotic Checkpoint Protein BUBR1 Is an Unusual Pseudokinase. *Dev. Cell* **22**, 1321–1329.
148. Gassmann R, Holland AJ, Varma D, Wan X, Civril F, Cleveland DW, Oegema K, Salmon ED, Desai A. 2010 Removal of Spindly from microtubule-attached kinetochores controls spindle checkpoint silencing in human cells. *Genes Dev.* **24**, 957–971.
149. Barisic M, Sohm B, Mikolcevic P, Wandke C, Rauch V, Ringer T, Hess M, Bonn G, Geley S. 2010 Spindly/CCDC99 is required for efficient chromosome congression and mitotic checkpoint regulation. *Mol. Biol. Cell* **21**, 1968–1981.
150. Uzunova K *et al.* 2012 APC15 mediates CDC20 autoubiquitylation by APC/CMCC and disassembly of the mitotic checkpoint complex. *Nat. Struct. Mol. Biol.* **19**, 1116–1123.
151. Foster SA, Morgan DO. 2012 The APC/C Subunit Mnd2/Apc15 Promotes Cdc20 Autoubiquitination and Spindle Assembly Checkpoint Inactivation. *Mol. Cell* **47**, 921–932.
152. Habu T, Matsumoto T. 2013 p31(comet) inactivates the chemically induced Mad2-dependent spindle assembly checkpoint and leads to resistance to anti-mitotic drugs. *Springerplus* **2**, 562.
153. Hagan RS, Manak MS, Buch HK, Meier MG, Meraldi P, Shah J V, Sorger PK, Doxsey SJ. 2011 p31comet acts to ensure timely spindle checkpoint silencing subsequent to kinetochore attachment. *Mol. Biol. Cell* **22**, 4236–4246.
154. Date DA, Burrows AC, Summers MK. 2014 Phosphorylation regulates the p31comet-mitotic arrest-deficient 2 (Mad2) interaction to promote spindle assembly checkpoint (SAC) activity. *J. Biol. Chem.* **289**, 11367–11373.
155. Eytan E, Wang K, Miniowitz-Shemtov S, Sitry-Shevah D, Kaisari S, Yen TJ, Liu S-T, Hershko A. 2014 Disassembly of mitotic checkpoint complexes by the joint action of the AAA-ATPase TRIP13 and p31comet. *Proc. Natl. Acad. Sci.* **111**, 12019–12024.
156. Ye Q, Rosenberg SC, Moeller A, Speir JA, Su TY, Corbett KD. 2015 TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. *Elife* **4**, e07367.
157. Nelson CR, Hwang T, Chen PH, Bhalla N. 2015 TRIP13PCH-2 promotes Mad2 localization to unattached kinetochores in the spindle checkpoint response. *J. Cell Biol.* **211**, 503–516.
158. Nijenhuis W, Vallardi G, Teixeira A, Kops GJPL, Saurin AT. 2014 Negative feedback at kinetochores underlies a responsive spindle checkpoint signal. *Nat. Cell Biol.* **16**, 1257–1264.
159. Burki F, Kaplan M, Tikhonenkov D V., Zlatogursky V, Minh BQ, Radaykina L V., Smirnov A, Mylnikov AP, Keeling PJ. 2016 Untangling the early diversification of eukaryotes:

a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings. Biol. Sci.* **283**, 20152802.
160. Katz LA 2012 Origin and Diversification of Eukaryotes. *Annu. Rev. Microbiol.* **66**, 411–427.
161. Boettcher B, Barral Y. 2013 The cell biology of open and closed mitosis. *Nucleus* **4**, 160–165.
162. Cheeseman IM. 2014 The Kinetochore. *Cold Spring Harb. Perspect. Biol.* **6**, a015826–a015826.
163. Santaguida S, Musacchio A. 2009 The life and miracles of kinetochores. *EMBO J.* **28**, 2511–31.
164. De Souza CPC, Osmani SA. 2007 Mitosis, not just open or closed. *Eukaryot. Cell* **6**, 1521–1527.
165. Akiyoshi B, Gull K. 2014 Discovery of Unconventional Kinetochores in Kinetoplastids. *Cell* **156**, 1247–1258.
166. D'Archivio S, Wickstead B. 2017 Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J. Cell Biol.* **216**, 379–391.
167. Vleugel M, Hoogendoorn E, Snel B, Kops GJPL. 2012 Evolution and Function of the Mitotic Checkpoint. *Dev. Cell* **23**, 239–250.
168. Meraldi P, McAinsh AD, Rheinbay E, Sorger PK, Patrick Meraldi ADMERPKS. 2006 Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol.* **7**, R23.
169. Eme L, Trilles A, Moreira D, Brochier-Armanet C. 2011 The phylogenomic analysis of the Anaphase Promoting Complex and its targets points to complex and modern-like control of the cell cycle in the last common ancestor of eukaryotes. *BMC Evol. Biol.* **11**, 265.
170. Gutiérrez-Caballero C, Cebollero LR, Pendás AM. 2012 Shugoshins: from protectors of cohesion to versatile adaptors at the centromere. *Trends Genet.* **28**, 351–360.
171. Zamariola L, De Storme N, Tiang CL, Armstrong SJ, Franklin FCH, Geelen D. 2013 SGO1 but not SGO2 is required for maintenance of centromere cohesion in Arabidopsis thaliana meiosis. *Plant Reprod.* **26**, 197–208.
172. Wang X, Dai W. 2005 Shugoshin, a guardian for sister chromatid segregation. *Exp. Cell Res.* **310**, 1–9.
173. Kensche PR, van Noort V, Dutilh BE, Huynen MA. 2008 Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. *J R Soc Interface* **5**, 151–170.
174. Maaten L van der, Hinton G. 2008 Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605.
175. Morett E, Korbel JO, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P. 2003 Systematic discovery of analogous enzymes in thiamin biosynthesis. *Nat Biotechnol* **21**, 790–795.
176. Fokkens L, Snel B, Like Fokkens BS. 2009 Cohesive versus flexible evolution of functional modules in eukaryotes. *PLoS Comput. Biol.* **5**.
177. Schneider A, Seidl MF, Snel B. 2013 Shared Protein Complex Subunits Contribute to Explaining Disrupted Co-occurrence. *PLoS Comput Biol* **9**, e1003124.
178. Jiang H *et al.* 2014 A Microtubule-Associated Zinc Finger Protein, BuGZ, Regulates Mitotic Chromosome Alignment by Ensuring Bub3 Stability and Kinetochore Targeting. *Dev. Cell* **28**, 268–281.
179. Toledo CM *et al.* 2014 BuGZ Is Required for Bub3 Stability, Bub1 Kinetochore Function, and Chromosome Alignment. *Dev. Cell*

180. Wan Y, Zheng X, Chen H, Guo Y, Jiang H, He X, Zhu X, Zheng Y. 2015 Splicing function of mitotic regulators links R-loop–mediated DNA damage to tumor cell killing. *J Cell Biol* **209**, 235–246.
181. Jiang H, Wang S, Huang Y, He X, Cui H, Zhu X, Zheng Y. 2015 Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly. *Cell* **163**, 108–122.
182. Hirose H *et al*. 2004 Implication of ZW10 in membrane trafficking between the endoplasmic reticulum and Golgi. *EMBO J.* **23**, 1267–1278.
183. Chan YW, Fava LL, Uldschmid A, Schmitz MHA, Gerlich DW, Nigg EA, Santamaria A. 2009 Mitotic control of kinetochore-associated dynein and spindle orientation by human Spindly. *J Cell Biol* **185**, 859–874.
184. Griffis ER, Stuurman N, Vale RD. 2007 Spindly, a novel protein essential for silencing the spindle assembly checkpoint, recruits dynein to the kinetochore. *J Cell Biol* **177**, 1005–1015.
185. Yamamoto TG, Watanabe S, Essex A, Kitagawa R. 2008 SPDL-1 functions as a kinetochore receptor for MDF-1 in Caenorhabditis elegans. *J Cell Biol* **183**, 187–194.
186. Nagpal H, Fukagawa T. 2016 Kinetochore assembly and function through the cell cycle. *Chromosoma* **125**, 645–659.
187. Przewloka MR, Venkei Z, Bolanos-Garcia VM, Debski J, Dadlez M, Glover DM. 2011 CENP-C Is a Structural Platform for Kinetochore Assembly. *Curr. Biol.* **21**, 399–405.
188. Singh TR *et al*. 2010 MHF1-MHF2, a Histone-Fold-Containing Protein Complex, Participates in the Fanconi Anemia Pathway via FANCM. *Mol. Cell* **37**, 879–886.
189. Yan Z *et al*. 2010 A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Mol Cell* **37**, 865–878.
190. Westhorpe FG, Straight AF. 2013 Functions of the centromere and kinetochore in chromosome segregation. *Curr. Opin. Cell Biol.* **25**, 334–340.
191. Wang H *et al*. 2004 Human Zwint-1 specifies localization of Zeste White 10 to kinetochores and is essential for mitotic checkpoint signaling. *J Biol Chem* **279**, 54590–54598.
192. Pagliuca C, Draviam VM, Marco E, Sorger PK, De Wulf P. 2009 Roles for the conserved Spc105p/Kre28p complex in kinetochore-microtubule binding and the spindle assembly checkpoint. *PLoS One* **4**, e7640.
193. Jakopec V, Topolski B, Fleig U. 2012 Sos7, an Essential Component of the Conserved Schizosaccharomyces pombe Ndc80-MIND-Spc7 Complex, Identifies a New Family of Fungal Kinetochore Proteins. *Mol. Cell. Biol.* **32**, 3308–3320.
194. Famulski JK, Vos L, Sun X, Chan G. 2008 Stable hZW10 kinetochore residency, mediated by hZwint-1 interaction, is essential for the mitotic checkpoint. *J. Cell Biol.* **180**, 507–20.
195. Wang K, Sturt-Gillespie B, Hittle JC, Macdonald D, Chan GK, Yen TJ, Liu S-T. 2014 Thyroid Hormone Receptor Interacting Protein 13 (TRIP13) AAA-ATPase Is a Novel Mitotic Checkpoint-silencing Protein. *J. Biol. Chem.* **289**, 23928–23937.
196. Hara K *et al*. 2010 Crystal Structure of Human REV7 in Complex with a Human REV3 Fragment and Structural Implication of the Interaction between DNA Polymerase and REV1. *J. Biol. Chem.* **285**, 12299–12307.
197. Suzuki H, Kaizuka T, Mizushima N, Noda NN. 2015 Structure of the Atg101–Atg13 complex reveals essential roles of Atg101 in autophagy initiation. *Nat. Struct. & Mol. Biol.*
198. Jao CC, Ragusa MJ, Stanley RE, Hurley JH. 2013 A HORMA domain in Atg13 mediates PI 3-kinase recruitment in autophagy. *Proc. Natl. Acad. Sci.* **110**, 5486-5491
199. Loiodice I, Loïodice I, Alves A, Rabut G, van Overbeek M, Ellenberg J, Sibarita J-B,

Doye V, Loiodice I. 2004 The Entire Nup107-160 Complex, Including Three New Members, Is Targeted as One Entity to Kinetochores in Mitosis. *Mol. Biol. Cell* **15**, 3333–3344.
200. Orjalo A V, Arnaoutov A, Shen Z, Boyarchuk Y, Zeitlin SG, Fontoura B, Briggs S, Dasso M, Forbes DJ. 2006 The Nup107-160 Nucleoporin Complex Is Required for Correct Bipolar Spindle Assembly. *Mol. Biol. Cell* **17**, 3806–3818.
201. Koster MJE, Snel B, Timmers HTM. 2015 Genesis of Chromatin and Transcription Dynamics in the Origin of Species. *Cell* **161**, 724–736.
202. Luo X, Tang Z, Rizo J, Yu H. 2002 The Mad2 spindle checkpoint protein undergoes similar major conformational changes upon binding to either Mad1 or Cdc20. *Mol. Cell* **9**, 59–71.
203. Sironi L, Mapelli M, Knapp S, De Antoni A, Jeang K-TT, Musacchio A. 2002 Crystal structure of the tetrameric Mad1-Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint. *EMBO J* **21**, 2496–2506.
204. Ding D, Muthuswamy S, Meier I. 2012 Functional interaction between the Arabidopsis orthologs of spindle assembly checkpoint proteins MAD1 and MAD2 and the nucleoporin NUA. *Plant Mol. Biol.* **79**, 203–216.
205. Schittenhelm RB, Heeger S, Althoff F, Walter A, Heidmann S, Mechtler K, Lehner CF. 2007 Spatial organization of a ubiquitous eukaryotic kinetochore protein network in Drosophila chromosomes. *Chromosoma* **116**, 385–402.
206. Przewloka MR, Zhang W, Costa P, Archambault V, D'Avino PP, Lilley KS, Laue ED, McAinsh AD, Glover DM. 2007 Molecular analysis of core kinetochore composition and assembly in Drosophila melanogaster. *PLoS One* **2**, 10484–10489.
207. Williams B *et al*. 2007 Mitch – a rapidly evolving component of the Ndc80 kinetochore complex required for correct chromosome segregation in Drosophila. *J. Cell Sci.* **120**, 3522–3533.
208. Cheeseman IM, Niessen S, Anderson S, Hyndman F, Yates JR, Oegema K, Desai A. 2004 A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes Dev.* **18**, 2255–2268.
209. Nakajima Y, Tyers RG, Wong CCL, Yates JR, Drubin DG, Barnes G. 2009 Nbl1p: A Borealin/Dasra/CSC-1-like Protein Essential for Aurora/Ipl1 Complex Function and Integrity in Saccharomyces cerevisiae. *Mol. Biol. Cell* **20**, 1772–1784.
210. Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009 Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**, 289–297.
211. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008 NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9.
212. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
213. 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.
214. Stamatakis A. 2014 RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
215. Malik HS, Henikoff S. 2003 Phylogenomics of the nucleosome. *Nat Struct Mol Biol* **10**, 882–891.
216. Letunic I, Bork P. 2016 Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245.
217. The UniProt Consortium. 2017 UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169.
218. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble

WS. 2009 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208.
219. Wu J, Kasif S, DeLisi C. 2003 Identification of functional links between genes using phylogenetic profiles. *Bioinformatics* **19**, 1524–1530.
220. Mi H, Muruganujan A, Thomas PD. 2013 PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41**, D377–D386.
221. Cunningham F *et al.* 2015 Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669.
222. Vader G. 2015 Pch2TRIP13: controlling cell division through regulation of HORMA domains. *Chromosoma* **124**, 333–339.
223. Richmond D, Rizkallah R, Liang F, Hurt MM, Wang Y. 2013 Slk19 clusters kinetochores and facilitates chromosome bipolar attachment. *Mol. Biol. Cell* **24**, 566–77.
224. Sarangapani KK *et al.* 2014 Sister kinetochores are mechanically fused during meiosis I in yeast. *Science* **346**, 248–251.
225. Gregan J *et al.* 2007 The Kinetochore Proteins Pcs1 and Mde4 and Heterochromatin Are Required to Prevent Merotelic Orientation. *Curr. Biol.* **17**, 1190–1200.
226. Rabitsch KP, Petronczki M, Javerzat JP, Genier S, Chwalla B, Schleiffer A, Tanaka TU, Nasmyth K. 2003 Kinetochore recruitment of two nucleolar proteins is required for homolog segregation in meiosis I. *Dev. Cell*. **4**, 535–548.
227. Corbett KD, Yip CK, Ee LS, Walz T, Amon A, Harrison SC. 2010 The monopolin complex crosslinks kinetochore components to regulate chromosome-microtubule attachments. *Cell* **142**, 556–567.
228. Tzafrir I, McElver JA, Liu CM, Yang LJ, Wu JQ, Martinez A, Patton DA, Meinke DW. 2002 Diversity of TITAN functions in Arabidopsis seed development. *Plant Physiol.* **128**, 38–51.
229. Tzafrir I *et al.* 2004 Identification of genes required for embryo development in Arabidopsis. *Plant Physiol.* **135**, 1206–20.
230. Schulz I, Erle A, Gräf R, Krüger A, Lohmeier H, Putzler S, Samereier M, Weidenthaler S. 2009 Identification and cell cycle-dependent localization of nine novel, genuine centrosomal components in Dictyostelium discoideum. In *Cell Motility and the Cytoskeleton*, pp. 915–928.
231. Davey NE, Cowan JL, Shields DC, Gibson TJ, Coldwell MJ, Edwards RJ. 2012 SLiMPrints: Conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res.* **40**, 10628–10641.
232. Nguyen Ba AN, Yeh BJ, van Dyk D, Davidson AR, Andrews BJ, Weiss EL, Moses AM. 2012 Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci. Signal.* **5**, rs1.
233. 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**, 3433–4.
234. Delorenzi M, Speed T. 2002 An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625.
235. Eddy SSR *et al.* 2011 Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195.
236. Sarkar S, Shenoy RT, Dalgaard JZ, Newnham L, Hoffmann E, Millar JBA, Arumugam P. 2013 Monopolin Subunit Csm1 Associates with MIND Complex to Establish Monopolar Attachment of Sister Kinetochores at Meiosis I. *PLoS Genet.* **9**, e1003610.
237. Holland AJ, Reis RM, Niessen S, Pereira C, Andres DA, Spielmann HP, Cleveland DW, Desai A, Gassmann R. 2015 Preventing farnesylation of the dynein adaptor Spindly contributes to the mitotic defects caused by farnesyltransferase inhibitors.

Mol. Biol. Cell **26**, 1845–1856.
238. Moudgil DK, Westcott N, Famulski JK, Patel K, Macdonald D, Hang H, Chan GKT. 2015 A novel role of farnesylation in targeting a mitotic checkpoint protein, human Spindly, to kinetochores. *J. Cell Biol.* **208**, 881–896.
239. Wang M, Casey PJ. 2016 Protein prenylation: unique fats make their mark on biology. *Nat. Rev. Mol. Cell Biol.* **17**, 110–122.
240. Hoogenraad CC, Akhmanova A. 2016 Bicaudal D Family of Motor Adaptors: Linking Dynein Motility to Cargo Binding. *Trends Cell Biol.* **26**, 327–340.
241. Hoogenraad CC, Akhmanova A, Howell SA, Dortland BR, De Zeeuw CI, Willemsen R, Visser P, Grosveld F, Galjart N. 2001 Mammalian golgi-associated Bicaudal-D2 functions in the dynein-dynactin pathway by interacting with these complexes. *EMBO J.* **20**, 4041–4054.
242. Schlager MA *et al.* 2010 Pericentrosomal targeting of Rab6 secretory vesicles by Bicaudal-D-related protein 1 regulates neuritogenesis. *EMBO J.* **29**, 1637–51.
243. Splinter D *et al.* 2012 BICD2, dynactin, and LIS1 cooperate in regulating dynein recruitment to cellular structures. *Mol. Biol. Cell* **23**, 4226–41.
244. Burge C, Karlin S. 1997 Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94.
245. Stanke M, Tzvetkova A, Morgenstern B. 2006 AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7 Suppl 1**, S11.1-8.
246. Dosztányi Z, Csizmók V, Tompa P, Simon I. 2005 The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347**, 827–39.
247. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004 WebLogo: A sequence logo generator. *GENOME Res.* **14**, 1188–1190.
248. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015 JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394.
249. Cheeseman IM, Desai A. 2008 Molecular architecture of the kinetochore-microtubule interface. *Nat. Rev. Mol. Cell Biol.* **9**, 33–46.
250. Foley EA, Kapoor TM. 2012 Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore. *Nat. Rev. Mol. Cell Biol.* **14**, 25–37.
251. Chao WCH, Kulkarni K, Zhang Z, Kong EH, Barford D. 2012 Structure of the mitotic checkpoint complex. *Nature* **484**, 208–213.
252. Kiyomitsu T, Obuse C, Yanagida M. 2007 Human Blinkin/AF15q14 is required for chromosome alignment and the mitotic checkpoint through direct interaction with Bub1 and BubR1. *Dev. Cell* **13**, 663–76.
253. Foley EAEA, Maldonado MM, Kapoor TMTM. 2011 Formation of stable attachments between kinetochores and microtubules depends on the B56-PP2A phosphatase. *Nat. Cell Biol.* **13**, 1265–1271.
254. Kawashima S a, Yamagishi Y, Honda T, Ishiguro K, Watanabe Y. 2010 Phosphorylation of H2A by Bub1 prevents chromosomal instability through localizing shugoshin. *Science* **327**, 172–177.
255. Yamagishi Y, Honda T, Tanno Y, Watanabe Y. 2010 Two Histone Marks Establish the Inner Centromere and Chromosome Bi-Orientation. *Science* **330**, 239–243.
256. Tang Z, Shu H, Oncel D, Chen S, Yu H. 2004 Phosphorylation of Cdc20 by Bub1 Provides a Catalytic Mechanism for APC/C Inhibition by the Spindle Checkpoint. *Mol. Cell* **16**, 387–397.
257. Klebig C, Korinth D, Meraldi P. 2009 Bub1 regulates chromosome segregation in a kinetochore-independent manner. *J. Cell Biol.* **185**, 841–858.

258. Bolanos-Garcia VM, Nilsson J, Blundell TL. 2012 The architecture of the BubR1 tetratricopeptide tandem repeat defines a protein motif underlying mitotic checkpoint-kinetochore communication. *Bioarchitecture* **2**, 23–27.
259. Kiyomitsu T, Murakami H, Yanagida M. 2011 Protein interaction domain mapping of human kinetochore protein Blinkin reveals a consensus motif for binding of spindle assembly checkpoint proteins Bub1 and BubR1. *Mol. Cell. Biol.* **31**, 998–1011.
260. Krenn V V, Wehenkel AA, Li XX, Santaguida SS, Musacchio AA. 2012 Structural analysis reveals features of the spindle checkpoint kinase Bub1-kinetochore subunit Knl1 interaction. *J. Cell Biol.* **196**, 451–467.
261. Shepperd L a., Meadows JC, Sochaj AM, Lancaster TC, Zou J, Buttrick GJ, Rappsilber J, Hardwick KG, Millar JB a. 2012 Phosphodependent recruitment of Bub1 and Bub3 to Spc7/KNL1 by Mph1 kinase maintains the spindle checkpoint. *Curr. Biol.* **22**, 891–899.
262. London N, Ceto S, Ranish J a., Biggins S. 2012 Phosphoregulation of Spc105 by Mps1 and PP1 regulates Bub1 localization to kinetochores. *Curr. Biol.* **22**, 900–906.
263. Yamagishi Y, Yang C-H, Tanno Y, Watanabe Y. 2012 MPS1/Mph1 phosphorylates the kinetochore protein KNL1/Spc7 to recruit SAC components. *Nat. Cell Biol.* **14**, 746–752.
264. Janicki SM *et al.* 2004 From silencing to gene expression: Real-time analysis in single cells. *Cell* **116**, 683–698.
265. Petrovic A *et al.* 2010 The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *J. Cell Biol.* **190**, 835–852.
266. Nijenhuis W *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.
267. Santaguida S, Vernieri C, Villa F, Ciliberto A, Musacchio A. 2011 Evidence that Aurora B is implicated in spindle checkpoint signalling independently of error correction. *EMBO J.* **30**, 1508–1519.
268. Saurin AT, van der Waal MS, Medema RH, Lens SM a, Kops GJPL. 2011 Aurora B potentiates Mps1 activation to ensure rapid checkpoint establishment at the onset of mitosis. *Nat. Commun.* **2**, 316.
269. Schittenhelm RB, Chaleckis R, Lehner CF. 2009 Intrakinetochore localization and essential functional domains of Drosophila Spc105. *EMBO J.* **28**, 2374–2386.
270. Johnson VL, Scott MIF, Holt S V, Hussein D, Taylor SS. 2004 Bub1 is required for kinetochore localization of BubR1, Cenp-E, Cenp-F and Mad2, and chromosome congression. *J. Cell Sci.* **117**, 1577–1589.
271. Hegemann B *et al.* 2011 Systematic Phosphorylation Analysis of Human Mitotic Protein Complexes. *Sci. Signal.* **4**, rs12-rs12.
272. London N, Biggins S. 2014 Signalling dynamics in the spindle checkpoint response. *Nat. Rev. Mol. Cell Biol.* **15**, 736–47.
273. Bolanos-Garcia VM, Blundell TL. 2011 BUB1 and BUBR1: multifaceted kinases of the cell cycle. *Trends Biochem. Sci.* **36**, 141–150.
274. Tang Z, Bharadwaj R, Li B, Yu H. 2001 Mad2-Independent inhibition of APC Cdc20 by the mitotic checkpoint protein BubR1. *Dev Cell* **1**, 227–237.
275. Han JS, Holland AJ, Fachinetti D, Kulukian A, Cetin B, Cleveland DW. 2013 Catalytic Assembly of the Mitotic Checkpoint Inhibitor BubR1-Cdc20 by a Mad2-Induced Functional Switch in Cdc20. *Mol. Cell* **51**, 92–104.
276. Vleugel M, Tromer E, Omerzu M, Groenewold V, Nijenhuis W, Snel B, Kops GJPL. 2013 Arrayed BUB recruitment modules in the kinetochore scaffold KNL1 promote accurate chromosome segregation. *J. Cell Biol.* **203**, 943–955.
277. Krenn V, Overlack K, Primorac I, van Gerwen S, Musacchio A. 2014 KI motifs of

human Knl1 enhance assembly of comprehensive spindle checkpoint complexes around MELT repeats. *Curr. Biol.* **24**, 29–39.
278. Björklund ÅK, Ekman D, Elofsson A. 2006 Expansion of protein domain repeats. *PLoS Comput. Biol.* **2**, 0959–0970.
279. O'Leary M a *et al.* 2013 The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–7.
280. Misof B *et al.* 2014 Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767.
281. Yang Z. 2007 PAML Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 496–503.
282. Panhuis TM, Clark NL, Swanson WJ. 2006 Rapid evolution of reproductive proteins in abalone and Drosophila. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361**, 261–268.
283. Oliver PL *et al.* 2009 Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. *PLOS Genet.* **5**, e1000753.
284. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014 An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242–245.
285. Bennett SM, Noor M a F. 2009 Molecular evolution of a Drosophila homolog of human BRCA2. *Genetica* **137**, 213–219.
286. Lou DI, McBee RM, Le UQ, Stone AC, Wilkerson GK, Demogines AM, Sawyer SL. 2014 Rapid evolution of BRCA1 and BRCA2 in humans and other primates. *BMC Evol. Biol.* **14**, 155.
287. Yang C, Stiller JW. 2014 Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5920–5.
288. Ellegren H. 2004 Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**, 435–445.
289. Chmátal L *et al.* 2014 Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr. Biol.* **24**, 2295–2300.
290. Dumont J, Desai A. 2012 Acentrosomal spindle assembly and chromosome segregation during oocyte meiosis. *Trends Cell Biol.* **22**, 241–249.
291. Sasao T, Itoh N, Takano H, Watanabe S, Wei G, Tsukamoto T, Kuzumaki N, Takimoto M. 2004 The protein encoded by cancer/testis gene D40/AF15q14 is localized in spermatocytes, acrosomes of spermatids and ejaculated spermatozoa. *Reproduction* **128**, 709–716.
292. Werren JH. 2011 Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl**, 10863–10870.
293. Finn RD, Clements J, Eddy SR. 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37.
294. Baron AP *et al.* 2016 Probing the catalytic functions of Bub1 kinase using the small molecule inhibitors BAY-320 and BAY-524. *Elife* **5**.
295. Sudakin V, Chan GKT, Yen TJ. 2001 Checkpoint inhibition of the APC / C in HeLa cells. *Cell.* **154**, 925–936.
296. Lischetti T, Zhang G, Sedgwick GG, Bolanos-Garcia VM, Nilsson J. 2014 The internal Cdc20 binding site in BubR1 facilitates both spindle assembly checkpoint signalling and silencing. *Nat. Commun.* **5**, 5563.
297. King EMJJ, van der Sar SJAA, Hardwick KG. 2007 Mad3 KEN Boxes Mediate both Cdc20 and Mad3 Turnover, and Are Critical for the Spindle Checkpoint. *PLoS One* **2**, e342.
298. Murray AW. 2012 Don't Make Me Mad, Bub! *Dev. Cell* **22**, 1123–1125.
299. Taylor SS, Ha E, McKeon F. 1998 The human homologue of Bub3 is required for

kinetochore localization of Bub1 and a Mad3/Bub1-related protein kinase. *J. Cell Biol.* **142**, 1–11.
300. Kang J, Yang M, Li B, Qi W, Zhang C, Shokat KM, Tomchick DR, Machius M, Yu H. 2008 Structure and Substrate Recruitment of the Human Spindle Checkpoint Kinase Bub1. *Mol. Cell* **32**, 394–405.
301. Moyle MW, Kim T, Hattersley N, Espeut J, Cheerambathur DK, Oegema K, Desai A. 2014 A Bub1-Mad1 interaction targets the Mad1-Mad2 complex to unattached kinetochores to initiate the spindle checkpoint. *J. Cell Biol.* **204**, 647–657.
302. Suijkerbuijk SJE, Van Osch MHJ, Bos FL, Hanks S, Rahman N, Kops GJPL. 2010 Molecular causes for BUBR1 dysfunction in the human cancer predisposition syndrome mosaic variegated aneuploidy. *Cancer Res.* **70**, 4891–4900.
303. Ogo N, Oishi S, Matsuno K, Sawada J, Fujii N, Asai A. 2007 Synthesis and biological evaluation of L-cysteine derivatives as mitotic kinesin Eg5 inhibitors. *Bioorg. Med. Chem. Lett.* **17**, 3921–4.
304. Elowe S, Dulla K, Uldschmid A, Li X, Dou Z, Nigg EA. 2010 Uncoupling of the spindle-checkpoint and chromosome-congression functions of BubR1. *J. Cell Sci.* **123**, 84–94.
305. Lara-Gonzalez P, Scott MIF, Diez M, Sen O, Taylor SS. 2011 BubR1 blocks substrate recruitment to the APC/C in a KEN-box-dependent manner. *J. Cell Sci.* **124**, 4332–4345.
306. Darriba D, Taboada GL, Doallo R, Posada D. 2011 ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165.
307. Akiyoshi B, Gull K. 2013 Evolutionary cell biology of chromosome segregation: insights from trypanosomes. *Open Biol.* **3**, 130023.
308. Keeling PJ *et al.* 2014 The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.* **12**, e1001889.
309. Ovchinnikov S, Kamisetty H, Baker D. 2014 Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **3**, e02030.
310. Marks DS, Colwell LJ, Sheridan R, Hopf T a., Pagnani A, Zecchina R, Sander C. 2011 Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, e28766.
311. Ovchinnikov S, Kinch L, Park H, Liao Y, Pei J, Kim DE, Kamisetty H, Grishin N V, Baker D. 2015 Large-scale determination of previously unsolved protein structures using evolutionary information. *Elife* **4**, e09248.
312. Ovchinnikov S, Park H, Varghese N, Huang P-S, Pavlopoulos GA, Kim DE, Kamisetty H, Kyrpides NC, Baker D. 2017 Protein structure determination using metagenome sequence data. *Science* **355**, 294–298.
313. Elias M, Brighouse A, Gabernet-Castello C, Field MC, Dacks JB. 2012 Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J. Cell Sci.* **125**, 2500–2508.
314. Veretnik S, Wills C, Youkharibache P, Valas RE, Bourne PE. 2009 Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. *PLoS Comput. Biol.* **5**, e1000315.
315. Páez-Pereda M, Arzt E. 2015 Function and structure of the RWD domain. *J. Biol. Chem.* **290**, 20627.
316. 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.
317. Rakesh R, Krishnan R, Sattlegger E, Srinivasan N. 2017 Recognition of a structural

domain (RWDBD) in Gcn1 proteins that interacts with the RWD domain containing proteins. *Biol. Direct* **12**, 12.
318. Hodson C, Cole AR, Lewis LPC, Miles JA, Purkiss A, Walden H. 2011 Structural analysis of human FANCL, the E3 ligase in the fanconi anemia pathway. *J. Biol. Chem.* **286**, 32628–32637.
319. 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.
320. Tagaya M, Arasaki K, Inoue H, Kimura H. 2014 Moonlighting functions of the NRZ (mammalian Dsl1) complex. *Membr. Traffic*
321. Drechsler H, McAinsh AD. 2012 Exotic mitotic mechanisms. *Open Biol.* **2**, 120140.
322. Raikov IB. 1994 The diversity of forms of mitosis in protozoa: a comparative review. *Eur. J. Protistol.* **30**, 253–269.
323. Azimzadeh J. 2014 Exploring the evolutionary history of centrosomes. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130453.
324. Doerder FP. 2014 Abandoning sex: multiple origins of asexuality in the ciliate Tetrahymena. *BMC Evol. Biol.* **14**, 112.
325. Cervantes MD, Xi X, Vermaak D, Yao M-C, Malik HS. 2006 The CNA1 Histone of the Ciliate Tetrahymena thermophila Is Essential for Chromosome Segregation in the Germline Micronucleus. *Mol. Biol. Cell* **17**, 485–497.
326. Francia ME, Striepen B. 2014 Cell division in apicomplexan parasites. *Nat. Rev. Microbiol.* **12**, 125–136.
327. Brooks CF, Francia ME, Gissot M, Croken MM, Kim K, Striepen B. 2011 Toxoplasma gondii sequesters centromeres to a specific nuclear region throughout the cell cycle. *Proc. Natl. Acad. Sci.* **108**, 3767–3772.
328. Verma G, Surolia N. 2013 Plasmodium falciparum CENH3 is able to functionally complement Cse4p and its, C-terminus is essential for centromere function. *Mol. Biochem. Parasitol.* **192**, 21–29.
329. Farrell M, Gubbels MJ. 2014 The Toxoplasma gondii kinetochore is required for centrosome association with the centrocone. *Cell. Microbiol.* **16**, 78–94.
330. Gornik SG, Ford KL, Mulhern TD, Bacic A, McFadden GI, Waller RF. 2012 Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Curr. Biol.* **22**, 2303–2312.
331. Spector DL, Triemer RE. 1981 Chromosome structure and mitosis in the dinoflagellates: An ultrastructural approach to an evolutionary problem. *Biosystems* **14**, 289–298.
332. Fernández-Robledo JA, Lin Z, Vasta GR. 2008 Transfection of the protozoan parasite Perkinsus marinus. *Mol. Biochem. Parasitol.* **157**, 44–53.
333. Perkins FO. 1996 The structure of Perkinsus-marinus (Mackin, Owen and Collier, 1950) Levine, 1978 with comments on taxonomy and phylogeny of Perkinsus spp. *J. Shellfish Res.* **15**, 67–87.
334. Ribeiro KC, Pereira-Neves A, Benchimol M. 2002 The mitotic spindle and associated membranes in the closed mitosis of trichomonads. *Biol. Cell* **94**, 157–172.
335. Zubácová Z, Hostomská J, Tachezy J. 2012 Histone H3 Variants in Trichomonas vaginalis. *Eukaryot. Cell* **11**, 654–661.
336. LaFountain JR, Davidson LA. 1979 An analysis of spindle ultrastructure during prometaphase and metaphase of micronuclear division in Tetrahymena. *Chromosoma* **75**, 293–308.
337. Brugerolle G. 1975 Etude de la cryptopleuromitose et de la morphogènèse de division chez Trichomonas vaginalis et chez plusieurs genres de Trichomonadines primitives. *Protistologica* **11**, 457–468.

Samenvatting in het Nederlands

DNA, genen en eiwitten

Evolutie van het leven op aarde behelst het behoud en de overdracht van het continu veranderende erfelijk materiaal aan volgende generaties. Net als bij dieren, planten, schimmels, eencellige algen, bacteriën en al het andere leven, bestaat ons erfelijk materiaal uit lange strengen opgebouwd uit slechts vier chemische bouwstenen: adenine (A), thymine (T), guanine (G) en cytosine (C), ook wel bekend als DNA. Verborgen in de kluwen van ATCG's (het genoom) liggen kleine eilandjes van orde (genen), die samen de code bevatten voor de productie van alle componenten die nodig zijn voor het functioneren van de basiseenheid van alle levende wezens: de cel. Het grootste deel van de functies in een cel wordt uitgevoerd door complexe moleculaire 'machineonderdelen' die opgebouwd zijn uit twintig verschillende aminozuren en in het Nederlands eiwitten worden genoemd. Als zodanig bevatten de meeste genen de code voor de precieze opbouw van één specifiek eiwit. Het menselijk genoom (3.000.000.000 ATCG's) bevat bijvoorbeeld ongeveer 20.000 genen en codeert dus voor ongeveer evenzoveel functionele eiwitten die tezamen in verschillende hoeveelheden en in verscheidene celtypen de mens in al haar biologische facetten vormt en onderscheidt van al het andere leven.

Vergelijkende genoomanalyse

Omdat de blauwdruk van de diversiteit van het leven zich dus schuil houdt in het DNA, zou inzicht in de volgorde van alle ATCG's van alle soorten en de locatie van alle genen, in principe kunnen leiden tot een volledige reconstructie van de afstamming van alle soorten (fylogenie) en ons mogelijk inzicht kunnen geven in de functionele overeenkomsten en verschillen tussen eiwitten en cellen van bijvoorbeeld de mens, de bakkersgist en de fruitvlieg. Na decennia van arbeidsintensief laboratoriumwerk hebben recente technologische ontwikkelingen het nu mogelijk gemaakt om in relatief korte tijd de ATCG-volgorde (sequentie) van het genoom van organismen van uiteenlopende complexiteit te bepalen. Met de gestaag groter wordende collectie van genomen en genen was het met de hand vergelijken van de DNA- of aminozuursequenties al snel ondoenlijk geworden en zo ontstond het onderzoeksveld van de vergelijkende genoomanalyse. Binnen deze wetenschappelijke discipline worden bioinformatische technieken en modellen ontwikkeld om de (co-)evolutie van genomen, genen en eiwitten tot in detail te reconstrueren en in kaart te brengen.

Een opmerkelijke uitkomst uit onderzoek van de afgelopen twintig jaar is dat in tegenstelling tot de intuïtie van velen, de cellulaire complexiteit van verschillende vormen van leven voorouderlijk is. Dit betekent dat bijvoorbeeld veel eigenschappen van menselijke cellen zeer waarschijnlijk al aanwezig waren in

de gemeenschappelijke voorouder van alle dieren en niet geleidelijk geëvolueerd zijn. Een ander interessant patroon is dat een groot aantal genen en zelfs hele genomen tijdens de evolutie dupliceren (verdubbelen), wat ook wel geassocieerd wordt met het ontstaan van veel nieuwe soorten. Hoewel veel van deze duplicaten ook weer verloren zijn gegaan, vormen deze extra kopieën mogelijkheden voor het evolueren van nieuwe functies en blijkt uit verschillende reconstructies dat een deel van de cellulaire complexiteit is ontstaan door de duplicatie van 'oercomponenten'. Als laatste heeft de ontwikkeling van het concept en de vaststelling van homologie tussen genen (gemeenschappelijke afstamming) ertoe bijgedragen dat veel van de functionele gegevens die beschikbaar zijn voor een scala van modelorganismen, zoals bijvoorbeeld de fruitvlieg, bakkersgist en een microscopische worm, nu ook gebruikt kunnen worden om de functie van eiwitten uit nieuwe genomen te annoteren. Al met al biedt de grote hoeveelheid genoominformatie en een goed gevulde gereedschapskist met adequate analysetechnieken de hedendaagse (evolutionaire) celbioloog een grote hoeveelheid mogelijkheden om zowel de evolutie alsook de functie van cellulaire systemen te onderzoeken.

Celdeling in eukaryoten

Om voortbestaan van het leven te garanderen, delen cellen. Cruciaal voor dit proces is de coördinatie van de celdeling met de verdubbeling en juiste verdeling van het genoom. Afwijkingen in dit proces kunnen leiden tot disfunctionerende cellen, celdood of bijdragen aan het ontstaan van verschillende aandoeningen zoals bijvoorbeeld kanker in de mens. Grofweg bestaan er twee systemen in de natuur om cellen en het genoom te verdelen, die een reflectie zijn van de evolutie van het leven. In cellen zonder een celkern (prokaryoten), zoals bijvoorbeeld in veel bacteriën, bestaat het genoom uit een groot rond DNA-molecuul. Alles in deze zeer kleine cellen (ongeveer 1 micrometer) is erop gericht om zo snel mogelijk te delen en daarom vinden de verdubbeling en verdeling van de cel en het genoom in essentie op hetzelfde moment plaats. Bij eukaryoten (cellen met een celkern), waartoe bijvoorbeeld de mens, planten, schimmels en veel eencelligen (protisten) behoren, zit het genoom opgesloten in de veilige omgeving van de celkern en is het opgedeeld in verschillende grote stukken DNA die ook wel chromosomen worden genoemd. Door deze ruimtelijke ordening en omdat ze doorgaans veel groter zijn dan prokaryoten (in orde van 10 micrometer), bezitten eukaryoten allerhande complexe moleculaire regulatiesystemen om het verdubbelen en verdelen van chromosomen af te stemmen met de celdeling. Tijdens dit proces gaan eukaryote cellen door een cyclus van 4 fasen: (1) groeifase 1, waarin de cel bepaalt of er genoeg bouwstoffen aanwezig zijn om daadwerkelijk over te gaan tot de totale verdubbeling - (2) tijdens de replicatiefase worden alle chromosomen in de celkern afzonderlijk gekopieerd en aan elkaar gekoppeld (chromosoom-

paar) - (3) in groeifase 2 checkt de cel of er geen fouten zijn gemaakt tijdens het verdubbelen van de chromosomen en neemt het celvolume verder toe - (4) in mitose krijgt de celdelingsmachinerie in de vorm van het spoelfiguur toegang tot de chromosoomparen. Het spoelfiguur heeft twee uiteinden (centrosomen) en bestaat uit dynamische kabels (microtubuli) die uiteindelijk de chromosoomparen uit elkaar zullen trekken om zo een gelijke verdeling van het aantal chromosomen over de twee te vormen dochtercellen te bewerkstelligen. Omdat het DNA en de microtubuli niet zomaar aan elkaar kunnen binden wordt op elk chromosoom een speciale ankerplaats aangelegd die bekend staat als het kinetochoor.

Kinetochoor: functie en evolutie

Kinetochoren hebben een centrale coördinerende rol in het proces van mitose. Ten eerste zijn ze verantwoordelijk voor het onderhouden van verbindingen tussen microtubuli van het spoelfiguur en het DNA van de chromosomen. Daarnaast communiceert het kinetochoor met de celcyclusmachinerie over de status van de chromosoom-microtubuli verbindingen. Indien de chromosomen binnen de chromosoomparen niet elk apart een verbinding zijn aangegaan met microtubuli van een van de twee uiteinden van het spoelfiguur, zendt het kinetochoor een signaal uit naar de rest van de cel om te wachten met delen. Dit fenomeen staat ook wel bekend als het 'mitotisch checkpoint'. Pas zodra alle chromosoomparen geordend en gerangschikt in het midden van de cel liggen (metafase), stoppen de kinetochoren met signaleren. De banden die de chromosoomparen bij elkaar houden worden doorgeknipt en microtubuli van het spoelfiguur trekken de chromosomen uiteen om ze uiteindelijk te laten belanden in een van de twee reeds vormende dochtercellen.

Om het DNA en microtubuli op een gecontroleerde manier te verbinden bestaan kinetochoren in bijvoorbeeld mens uit ongeveer 80 eiwitten in verschillende hoeveelheden met ieder een specifieke taak. Hoewel we al aardig wat weten over de functie van deze eiwitten in een aantal modelorganismen (fruitvlieg, bakkersgist, mens, muis en de zandraket (plantje)), zijn er nog veel vragen. Daarnaast is er van een groot gedeelte van de eukaryoten nagenoeg niets bekend over hun kinetochoren en zijn er indicaties dat deze anders zijn opgebouwd en mogelijk ook anders functioneren. Door de revolutie in het bepalen van de genoominformatie is de DNA-sequentie van een steeds groter aantal eukaryote soorten beschikbaar gekomen. Het onderzoek dat beschreven staat in dit proefschrift had daarom als doel om in kaart te brengen hoe elk van de 80 bekende onderdelen van het kinetochoor zich evolutionair gedraagt, binnen een set van ongeveer 100 genomen die een weerslag zijn van de evolutie van eukaryoten tijdens de afgelopen anderhalf miljard jaar. Door gebruik te maken van verschillende concepten en technieken die ontwikkeld zijn

binnen de vergelijkende genoomanalyse, is een reconstructie gemaakt van de evolutie van het kinetochoor sinds de laatste gemeenschappelijke voorouder van alle eukaryoten (LECA). Op basis van deze reconstructies zijn vervolgens hypotheses opgesteld over de functie van enkele kinetochooreiwitten die in het laboratorium getest zijn in menselijke cellen met behulp van verschillende microscopische en biochemische technieken.

Overzicht van dit proefschrift

Hoofdstuk 1 bevat een korte beschrijving van de ontwikkelingen binnen het onderzoeksveld van de vergelijkende genoomanalyse. Daarnaast wordt uitgebreid ingegaan op wat er bekend is over de moleculaire functie van de 80 eiwitten die samen het kinetochoornetwerk vormen. In **hoofdstuk 2** wordt de basis gelegd voor de rest van dit proefschrift met een uitgebreide analyse van de af- en aanwezigheid van kinetochooreiwitten in een grote diversiteit van eukaryote genomen. Reconstructies laten zien dat LECA zeer waarschijnlijk al een zeer complexe kinetochoor bezat en dat de kinetochoren van hedendaagse eukaryote soorten sterk veranderd zijn, door snelle evolutie van de aminozuursequentie, het verlies van genen, meerdere (onafhankelijke) genduplicaties, nieuwe functionele eiwitcomplexen en zelfs vervanging van complete functionele systemen. Opvallend is dat eiwitten die dezelfde functie uitvoeren, co-evolueren, wat betekent dat ze tegelijk aan- of afwezig zijn in de genomen van de soorten die onderzocht zijn. In **hoofdstuk 3** wordt voortgebouwd op de vindingen uit het voorgaande hoofdstuk en getest hoe robuust de bioinformatische methoden zijn om genen en eiwitten in de genomen van eukaryoten te lokaliseren. Daarnaast wordt binnen dit hoofdstuk een workflow ontwikkeld (ConFeaX) om te bepalen welke stukjes van de aminozuursequentie (motieven) hetzelfde zijn (geconserveerd) binnen eiwitten die een gemeenschappelijke voorouder hebben (homologen) en of deze co-evolueren met andere componenten van het kinetochoor. Doorgaans wordt aangenomen dat als deze aminozuren geconserveerd zijn over een langere periode van evolutie, ze erg belangrijk zijn voor de functie van de eiwitten. Als zodanig is het mogelijk om hypotheses te formuleren over motieven binnen (verschillende) eiwitten die samen een functie uitvoeren op het kinetochoor.

In **hoofdstuk 4-6** wordt gebruik gemaakt van de ConFeaX methode om de evolutie en functie van eiwitfamilies onder de loep te nemen die betrokken zijn bij het mitotische checkpoint en chromosoomoriëntatie. In **hoofdstuk 4 en 5** komt de functie en opmerkelijke evolutie van Knl1 (eiwitnamen zijn vaak 3 letters en 1-2 cijfers) aan bod. Knl1 bestaat uit repeterende stukjes aminozuursequentie die als landingsmodule kunnen dienen voor drie BUB-eiwitten (Bub1, BubR1 en Bub3). Deze BUB-eiwitten zijn betrokken bij de generatie (Bub1) en zelf onderdeel (BubR1 en Bub3) van het checkpointsignaal dat door

kinetochoren wordt uitgezonden om de celdeling een halt toe te roepen als chromosomen nog niet goed aan microtubuli vastzitten. Gemuteerde varianten van Knl1 waarbij alleen enkele van deze repeterende modules aanwezig zijn, functioneren net zo goed als bij een intact Knl1. Door de aminozuursequentie van verschillende Knl1 homologen met elkaar te vergelijken, blijkt daarnaast ook dat de BUB-bindende modules zeer snel evolueren, niet alleen wat betreft de aminozuurvolgorde, maar ook in aantal. Vergelijkbare patronen van snelle evolutie zijn ook gevonden voor eiwitten die betrokken zijn bij de competitie van spermacellen voor een eicel of voor het immuunsysteem dat zich verweert tegen allerhande indringers, zoals virussen. Wat de precieze reden is voor deze snelle evolutie en wat de implicaties zijn voor Knl1 functie in andere eukaryote cellen, moet verder onderzoek uitwijzen. Naast Knl1 evolueren ook Bub1 en BubR1 op een zeer spectaculaire wijze. In voorgaand onderzoek was alreeds vastgesteld dat Bub1 en BubR1 tot dezelfde evolutionaire eiwitfamilie behoren en een voorouderlijk eiwit met functionaliteiten van beide eiwitten maar liefst negen keer onafhankelijk is gedupliceerd. Het laatste experimentele hoofdstuk, **hoofdstuk 6**, breidt deze analyses verder uit met nieuwe genoominformatie en laat zien dat deze duplicatie 15 keer heeft plaatsgevonden tijdens de evolutie van hedendaagse eukaryoten. Met behulp van ConFeaX worden geconserveerde elementen in beide duplicaten gelokaliseerd. Na duplicatie wordt in nagenoeg alle gevallen hetzelfde patroon van verlies en behoud van functionele elementen waargenomen. Bub1-achtige eiwitten behouden functies die een rol spelen op het kinetochoor, terwijl BubR1-achtige eiwitten juist co-evolutie laten zien van aminozuurmotieven die betrokken zijn bij het remmen van de celcyclusmachinerie. Aan twee van deze motieven (ABBA) was nog geen functie toegekend. Met verschillende experimenten wordt aannemelijk gemaakt dat de twee ABBA-motieven van BubR1 een belangrijke rol spelen bij het in stand houden van het mitotisch checkpointsignaal in menselijke cellen.

In **hoofdstuk 7** worden de experimentele en bioinformatische vindingen van dit proefschrift besproken in het licht van relevante literatuur. In dit laatste hoofdstuk breekt de promovendus een lans voor het onderzoek naar atypische kinetochoren in niet-modelorganismen, zoals de ciliaat *Tetrahymena thermophila*, de intracellulaire parasiet *Toxoplasma gondii* en een merkwaardige groep organismen die dinoflagellaten wordt genoemd. Daarnaast ziet hij mogelijkheden om te reconstrueren hoe het kinetochoornetwerk in eukaryoten mogelijk ontstaan is in één van de gemeenschappelijke voorouders van alle eukaryoten. Al met al laat het onderzoek in dit proefschrift zien dat een zeer essentiële structuur in cellen van alle eukaryoten merkwaardig snel evolueert. Een interessante hypothese die daarom opkomt uit dit werk is dat de kinetochoor mogelijk zelf één van de drijvende krachten is achter de evolutie van soorten. Meer onderzoek is nodig om dit idee en de observaties uit dit proefschrift verder te onderbouwen.

Curriculum Vitae

Elbert Cornelis (Eelco) Tromer werd op 23 April 1986 geboren te Soest. In 2004 behaalde hij het VWO diploma aan het van Lodenstein College te Amersfoort met het profiel Natuur & Gezondheid. In september van hetzelfde jaar begon hij met de Bacheloropleiding Biomedische Wetenschappen aan de Universiteit van Utrecht die hij aanvulde met vakken uit de opleiding Farmacie en in 2007 afrondde. Eelco vervolgde zijn opleiding aan de Universiteit van Utrecht met een Masteropleiding Drug Innovation die hij voltooide in 2011. Gedurende deze opleiding heeft hij 9 maanden onderzoek gedaan naar de moleculaire basis voor geneesmiddel-geneesmiddel interacties in het laboratorium van dr. Irma Meijerman op de afdeling Biomedische Analyse van het departement Farmaceutische Wetenschappen. Vervolgens deed Eelco zijn tweede onderzoekstage bij de vakgroep van dr. Rob Wolthuis aan het Nederlands Kanker Instituut in Amsterdam en onderzocht hij lineaire aminozuurmotieven in eiwitten die een rol spelen bij de regulatie van de celcyclus. In juli 2012 is Eelco gestart met zijn promotieonderzoek in de laboratoria van Prof. dr. Geert Kops en Prof. dr. Berend Snel met als hoofdonderwerp de evolutie en functie van het kinetochoornetwerk, met een speciale focus op het mitotische checkpoint. De resultaten van dit onderzoek staan beschreven in dit proefschrift. Vanaf april 2018 hoopt Eelco zijn wetenschappelijke carrière voort te zetten als post-doctoraal onderzoeker in het laboratorium van dr. Ross Waller aan de Universiteit van Cambridge in het Verenigd Koninkrijk.

List of publications

Jolien JE van Hooff, **Eelco Tromer**, Leny M van Wijk, Berend Snel & Geert JPL Kops (2017) Evolutionary dynamics of the kinetochore network as revealed by comparative genomics *EMBO reports*

Eelco Tromer, Debora Bade, Berend Snel & Geert JPL Kops (2016) Phylogenomics-guided discovery of a conserved cassette of short linear motifs in BubR1 essential for the spindle checkpoint *Open Biology*

Eelco Tromer, Berend Snel & Geert JPL Kops (2015) Widespread recurrent patterns of rapid repeat evolution in the kinetochore scaffold KNL1 *Genome biology and evolution*

Mathijs Vleugel, Tim A Hoek*, **Eelco Tromer***, Tale Sliedrecht, Vincent Groeneveld, Manja Omerzu, Geert JPL Kops (2015) Dissecting the roles of human BUB1 in the spindle assembly checkpoint *Journal of Cell Science*
*equal contribution as second author

Mathijs Vleugel, **Eelco Tromer**, Manja Omerzu, Vincent Groeneveld, Wilco Nijenhuis, Berend Snel, Geert JPL Kops (2013) Arrayed BUB recruitment modules in the kinetochore scaffold KNL1 promote accurate chromosome segregation *Journal of Cell Biology*

Wilco Nijenhuis, Eleonore von Castelmur, Dene Litter, Valerie De Marco, **Eelco Tromer**, Mathijs Vleugel, Maria HJ van Osch, Berend Snel, Anastassis Perrakis, Geert JPL Kops (2013) A TPR domain–containing N-terminal module of MPS1 is required for its kinetochore localization by Aurora B *Journal of Cell Biology*

Dankwoord

Hoewel schrijven niet mijn favoriete hobby is ben ik dan nu toch eindelijk aan het einde van dit proefschrift gekomen. Af en toe was het een (on)behoorlijke strijd, maar ik ben blij en misschien zelfs wel opgelucht dat ik hier nu ben aanbeland. De afgelopen vijf jaren zijn voor mij een zeer vormende ervaring geweest op zowel persoonlijk als wetenschappelijk vlak. Ik beschouw het dan ook als een groot voorrecht onderdeel uit te maken van een wereldwijde wetenschappelijke gemeenschap. Door ideeën uit te wisselen en telkens maar weer kritische vragen te stellen heb ik (hopelijk) geleerd om op een open manier in het leven en de wetenschap te staan. Wat dat betreft wens ik een ieder een keer een werkbespreking met wetenschappers toe. Wetenschap is mijn grootste passie en ik hoop haar kunsten nog vele jaren te mogen beoefenen.

Naast deze korte reflectie is dit dankwoord natuurlijk vooral bedoeld om een ieder te bedanken die een belangrijke bijdrage heeft geleverd aan dit proefschrift.

Allereerst heb ik het getroffen met mijn promotores, Prof. dr. Geert Kops en Prof. dr. Berend Snel. Geert, bedankt dat je mij de kans en het vertrouwen hebt gegeven om binnen jouw groep het evolutieproject verder uit te breiden, we staan eigenlijk nog maar aan het begin! Ondanks dat het afronden van projecten niet altijd even soepel ging ben je niet afgehaakt waar anderen dat misschien al wel gedaan zouden hebben. De vrijheid die jij de mensen in je lab gunt om hun eigen vragen te stellen heeft mij erg geïnspireerd en het laat denk ik iets zien van jouw medemenselijkheid en interesse in fundamentele vragen, deugden die bij groepsleiders in het huidige wetenschappelijke klimaat helaas nog wel eens gemist worden. Daarnaast heb ik grote bewondering voor jouw zeer efficiënte werkstijl, vermogen om op een heldere manier kritiek te leveren en je brede wetenschappelijke en maatschappelijke interesse. Het is een voorrecht om in jouw lab als wetenschapper te mogen rond lopen, ik heb hoop op een mooie toekomst.

Berend, hoe jij soms in een paar woorden complexe problemen en ideeën kan samenvatten en helder weet neer te zetten is wat mij betreft ongeëvenaard. De keren dat ik dacht: 'had ik dit maar opgenomen', kan ik niet meer op twee handen tellen. Dankjewel voor alles wat je mij geleerd hebt over evolutie, bioinformatica, hoe je een goed verhaal vertelt en over hoe het leven in elkaar zit. Naast dat je ontzettend slim bent en heel erg veel weet, heb ik altijd veel plezier beleeft aan jouw manier van doen. Soms kwam je zo maar het kantoor in lopen en zonder iets te zeggen was je ook weer weg. Jouw eigenheid, maar ook soms extreme zelftwijfel, zijn vreemd genoeg een erg grote inspiratie voor mij geweest. Mede hierdoor was er wellicht altijd wel even tijd om bij je binnen te lopen voor een praatje, een ik-weet-het-ook-even-niet gesprek of een mening over een van mijn nieuwe ondoordachte ideeën. Ik heb mij altijd heel erg thuis gevoeld in jouw groep; sorry dat ik niet wat meer met je over voetbal kon praten. Veel succes met het uitvoeren van al je nieuwe onderzoeksplannen en ideeën om de FECA-to-LECA-transitie te

reconstrueren, patronen van parallele evolutie te vinden en al die andere verhalen die je nog wil vertellen. We houden contact.

De mensen in de promotiecommissie, leescommissie en mijn OIO-commissie: Prof. dr. Martijn Huynen, Prof. dr. Sander van de Heuvel, Prof. dr. René Medema, dr. John van Dam, Prof. dr. Anna Akhmanova en Prof. dr. Frank Holstege bedank ik voor het betrokken zijn bij mijn promotietraject, het lezen en beoordelen van dit proefschrift en/of voor het nemen van zitting in de oppositie tijdens de verdediging. I want to extend my special thanks to Prof. dr. Stefan Westermann for giving a talk at the minisymposium on kinetochore function and evolution, and for taking part in the opposition during my defense.

Alles wat in dit proefschrift staat is op een of andere manier wel een keer de revue gepasseerd tijdens een van de vele werkbesprekingen in het Kruytgebouw, het Stratenum of op het Hubrecht. Als zodanig wil ik mijn mede-labgenoten, oud-labgenoten en (oud)-studenten heel erg bedanken voor al hun waardevolle input en constructieve kritiek en daarnaast natuurlijk ook voor de vele borrels, informele gesprekken en andere activiteiten waar we elkaar ontmoetten. Hoewel er zoveel mensen zijn die ik in de afgelopen vijf jaar ben tegengekomen, wil ik sommige mensen wil graag extra bedanken.

Mijn paranimfen, Bas en Carlos. Dear Carlos, the way you approach things in live and science is fresh, innovative and head-on. It has been an absolute delight to work with you over the last couple of years and it is a great privilege to call you my friend. When I think about a good experimental scientist, you are always the first that comes to mind. I really appreciate the open way we discuss our ideas during our 'weekly' bouldering evenings, when I sometimes work at the Hubrecht or just whenever. Although your time in the Netherlands has sometimes been quite challenging, I wish you all the happiness in the world and especially here in this cold and small country by the sea, you deserve it. Bas, van jou heb ik geleerd om je soms even helemaal niks aan te trekken van de wereld om je heen en vooral daar energie in te stoppen waar het nuttig is voor jezelf. De schijnbaar soepele tred waarmee jij door het leven gaat suggereert wellicht dat je alles niet zo serieus neemt, maar niets is minder waar. Jouw aandacht voor detail en organisatie (in een van je vele excelsheets) is niet gemakkelijk te evenaren en dat in combinatie met hogere filosofische redevoeringen over het al dan niet bestaan van MVA en plannen voor uitgekiende financiële investeringen, maken jou tot een persoon om rekening mee te houden. Ik vond het een plezier om met je samen te werken en ik wens je het allerbeste toe in je onderzoek en al je plannen voor het leven.

Dan het Kopslab. Als lab zijn we de afgelopen jaren een redelijk hechte club geworden en daarom ben ik ook veel mensen veel dank verschuldigd. Lieve Banafsheh, we zijn ongeveer tegelijk begonnen aan ons promotietraject en daarom promoveren we ook op dezelfde datum, met minisymposium en feest erbij, logisch toch. Omdat we qua gevoelsleven wellicht elkaars tegenpolen zijn (ligt denk ik

meer aan het gebrek ervan aan mijn kant) was het voor ons beiden volgens mij niet altijd even gemakkelijk om elkaar in complexe gesprekken te vinden. Daarom misschien wel heb ik jou als mijn beste labmaatje beschouwd. Jouw betrokkenheid en interesse voor anderen is een inspiratie en ik beleef altijd veel plezier aan onze gesprekken over hoe het leven werkt, hoe de wereld beter kan en moet en waarom mensen zo complex zijn (inclusief wijzelf). Door de verwarring heen blijkt vaak dat ik weer wat van je heb geleerd. Daarnaast ben je een creatieve wetenschapper die niet zomaar tevreden is met een antwoord en op zoek gaat waar anderen al hopeloos hebben opgegeven. Bedankt ook dat je me hebt gepusht om te gaan boulderen; sorry dat ik nog steeds niet mee doe met de yogalessen. Ik zie uit naar onze promotiedag! Nannette, dank voor al je scherpzinnige opmerkingen, eigenheid en voor het zijn van een gezellige kamergenoot. Het was interessant om als leek de afgelopen jaren meer te leren over muizenmodellen, kanker en CIN. Ajit, jouw gevoel voor mens en dier is heel bijzonder. Zonder schijnbare moeite weet jij saaie groepen om te toveren in dansende massa's en er is geen dier bang voor jou. Veel plezier met het muizenwerk en doe vooral rustig aan. Wilma, leven als Brabantse boven de rivieren suggereert dat je goed weet wat je wil. Je draait je hand niet meer om voor grote muizenproeven en rare blauwe kleuringen van de muizendarm. Veel succes met het afronden van je promotieonderzoek en ook met je keuze voor toekomstige avonturen. Antoinette, het Kopslab bestieren is geen gemakkelijke taak, zoals je ook wel hebt ondervonden. Ik heb diep respect voor alle dingen die jij doet om het lab draaiende te houden en ik wens je veel plezier en succes om dit ook de komende jaren te doen. Vincent, zonder jou geen massaspectrometrie. Met jouw unieke talent bent je goud waard voor het lab, dat je het even weet. Debora, heel erg bedankt voor de fijne samenwerking en al het biochemische werk dat je hebt gedaan voor hoofdstuk 6. Veel plezier en geluk bij het opvoeden van jullie kleine. Richard, we hebben heel wat biertjes achterover getikt de afgelopen jaren, mooi dat we elkaar op dit vlak mochten vinden. Met jouw mild-agressieve stijl van discussiëren en zachtaardige karakter heb je geen versterker nodig om je imago als woest aantrekkelijke bassist hoog te houden. Onze trip naar Spetses zal ik denk ik niet snel vergeten. Bedankt voor je interesse in mijn onderzoek, het ga je goed! Timo, als een van de meest technisch begaafde experimentalisten die ik ken, kijk ik met enige jaloezie naar jouw werk met de MPS1-FRET-probe; overigens ook naar jouw lunch, die zag er altijd heerlijk uit. Dank voor al je technische adviezen tijdens mijn gestuntel in het lab, of het nu met de om kloneringen, microscopen of andere zaken ging, je had altijd wel even tijd om mij te helpen. Ana, your arrival in the Kops lab was very interesting. From the moment you started you were determined to do so many things that your schedule did not have room for it. I really admire your determination to get things done and the enthusiasm about your project. Thank you for being the co-instigator of the SKAvonden. I wish you good luck in all your endeavours. Spiros, obviously you have had one of the hardest projects in the lab and I really admire your resolve to see it to an end. Thank you for not immediately taking the things I say during our work discussions for granted and for asking critical questions that reveal gaps in our knowledge. Your keen eye for incorrect reasoning is a quality you should cherish. Success with finishing your

PhD. Claudia, Xiaorong, Jing Chao, Sjoerd and Joris, all the best with your projects, I hope to see a lot of it. Mathijs, bedankt voor de samenwerking op het KNL1 en BUB1 project en ook voor het wegwijs maken op het lab. Veel succes met je carrière in de wondere wereld van de science journals. Tale, Wilco, Saskia en Manja bedankt voor de samenwerking voor langere en kortere tijd en het overerven van zeer interessante projecten.

Naast het Kopslab ben ik ook al die tijd onderdeel geweest van het Snellab. Beste Jolien, zonder jou was dit proefschrift lang niet zo diepgaand en uitgebreid geweest. Hoewel het niet zo lijkt, zit er heel veel werk in alle figuren en tekst van hoofdstuk 2 en ik wil je heel erg bedanken voor de samenwerking die wij de afgelopen jaren hebben gehad op het KT evolutieproject. Ik heb erg veel bewondering voor jouw snelle, efficiënte en intelligente manier van werken en ik heb mede door discussies met jou ook veel geleerd over hoe je grootschalige analyses in de bioinformatica aanpakt en interpreteert. Hopelijk kunnen we ons project over de pre-LECA evolutie van het kinetochoor nog van de grond krijgen, zou leuk zijn. Wat voor plannen je ook hebt voor de toekomst, ik weet zeker dat je met jouw kwaliteiten je plek zal vinden. Michael, I owe you a debt of gratitude as you have helped me a lot in figuring out how to code in Perl. I was delighted to hear that you managed to get an independent position at the University of Wageningen, well deserved! I wish you all the best in your career in science and perhaps one day you might be up for a rematch? Lidija, you made me feel very welcome in the Snel lab and you have been a wonderful roommate to me. Your aptitude for clear reasoning is very impressive and was at times equally helpful. I really enjoyed our conversation on all sorts of topics and I hope you will have a wonderful future in Wageningen together with Michael. Alessia, it already feels like ages since you have finished. It is very satisfying to see that you have pursued your dreams and have left the frustrating world of bioinformatics and bad datasets behind you. Thank you for all the fun discussions we had on life, science and use of it all. Good luck in all your adventures! Leny, het was interessant om samen met jou, als iemand met een refo-achtergrond, in een lab te werken dat onderzoek doet naar de evolutie van het leven. Onze recente lange treinreizen hebben mij weer doen beseffen hoe moeilijk ik het eigenlijk vind om zowel gelovige als wetenschapper te zijn. Dankjewel voor alle leuke en open gesprekken over werkwijzen, kinderen en levensfilosofieen. Jammer dat je nu moet stoppen met het kinase project nu het net lekker begint te lopen. Succes met het afronden van je proefschrift en we lopen elkaar vast nog wel eens tegen het lijf. John, hoewel je nog niet eens zo heel lang in Utrecht bent, voelt het al heel vertrouwd om met jou in een kantoor te werken. Jouw grondige aanpak is een voorbeeld voor menig bioinformaticus en ik hoop dat ook de funding agencies gaan inzien wat voor talenten jij in huis hebt. Hopelijk kunnen we ook in de toekomst nog enkele samenwerkingen smeden en meer licht werpen op de evolutie van cellulaire systemen in het algemeen. Succes en tot ziens.

Naast begeleid worden, heb ik ook zelf 3 studenten mogen begeleiden. Andrew, bedankt voor al je energie voor het visualiseren en begrijpen van EVfold. Jammer

genoeg kwam er niks uit maar ik ben er van overtuigd dat in de nabije toekomst jouw inspanningen nog van pas gaan komen. Laura, ik denk dat ik bij het begeleiden van jou, minstens zo veel van jou heb geleerd, als jij van mij. Met al jouw ervaring op het lab was het soms lastig voor mij, als labamateur, om je van goed advies te voorzien en liep het project hier en daar wat stroef. Het hielp ook niet echt dat we midden in een verhuizing zaten. Mooi om te zien dat je het zo naar je zin hebt in het Hecklab en ik wens je daar heel veel succes met je promotieonderzoek. Joris, je hebt een van mijn projecten overgenomen en je bent voortvarend begonnen. Leuk om te zien dat je zoveel plezier hebt in je werk en met gezonde nieuwsgierigheid aan je PhD bent begonnen. Ik ben heel benieuwd wat je allemaal gaat doen de komende jaren. Veel succes en tot de volgende werkbespreking J.

Graag wil ik ook Michiel Boekhout en Rob Wolthuis bedanken voor hun bezielende begeleiding tijdens mijn laatste masterstage. Dank jullie wel voor de vrijheid die jullie mij hebben gegeven om mijn eigen onderzoek op te zetten. Ik heb heel veel van jullie geleerd en mijn tijd op het NKI is de basis geweest voor mijn verdere ontwikkeling als wetenschapper. Succes in jullie verdere carrières!

Voor het COMPAKIN project zijn we verschillende samenwerkingen begonnen met (inter)nationale onderzoeksgroepen waarvan de resultaten het niet hebben gehaald in dit proefschrift. Toch wil ik ook hiervoor een aantal mensen bedanken: Arjan, Ineke en Douwe, bedankt voor het opstarten en opzetten van taggings voor kinetochoor eiwitten in de amoebozoa *Dictyostelium* en leuk dat ik een keer bij jullie het lab in Groningen een paar dagen kon meelopen. Although the pulldown protocols still need some optimization, I want to thank Elysa, Han and Tijs from the University of Wageningen for helping us out with growing transgenic *Physcomitrella*. In addition I am really grateful for the help we were offered by the lab of Katerina Bisova for trying to set up gene taggings in green algae *Chlamydomonas*. The work on *Tetrahymena* is kindly supported by Josef Loidl and Rachel Howard-Till from Vienna – your help is much appreciated.

Omdat ik op drie verschillende plaatsen heb gewerkt, wil ik graag de Universiteit van Utrecht, het Hubrecht Instituut en het Universitair Medisch Centrum van Utrecht bedanken voor het beschikbaar stellen van hun faciliteiten om onderzoek te doen. Aan de mensen van TBB: dank voor de fijne sfeer en het adopteren van mij als gastmedewerker. In het bijzonder wil ik nog Bas, Bastiaan en Ksenia bedanken voor hun bijdragen tijdens de maandagmiddagwerkbesprekingen. Duizendmaal dank ben ik verschuldigd voor al het werk dat Jan Kees de afgelopen jaren heeft verzet om het mogelijk te maken voor mij om bioinformatische analyses te draaien op het altijd perfect werkende netwerk. De eerste jaren van mijn PhD onderzoek hebt ik gewerkt bij de afdeling Molecular Cancer Research op het UMC. In het bijzonder wil ik alle leden van de Lensgroep bedanken voor hun input op mijn projecten en de goede samenwerking op het lab. Susanne, Sanne, Martijn, Michael, Ingrid, Amanda en Sippe: bedankt en succes met al jullie projecten, we moesten maar weer eens een goede gezamenlijke werkbespreking plannen. Livio,

we missen je nog steeds op het lab. Jouw energie bracht altijd veel vrolijkheid met zich mee. Dank voor het organiseren van het lab, instructies op de microscopen en je attente mailtjes als er weer een interessante meeting aan zat te komen over evolutie. Succes daar op het UMC. With the move to the Hubrecht, the lab, and me included, were in unknown territory. I therefore want to specifically thank Xavi, PieterJan, Wim and Saman for organizing the weekly futsal hour, which allowed me to get to know some of the Hubrecht people. Furthermore, I want to extend my thanks to all the groups of the second floor of the new Hubrecht building for the nice atmosphere, good working environment and great pie time on Friday afternoon. To all members of the de Rooij, Tanenbaum, Knipscheer and Kind groups: thank you!

Wie kan er leven zonder vrienden? Hoewel jullie dan geen fysieke bijdrage leverden, hebben jullie de afgelopen 5 jaar wel mijn geklaag, gesteun en overenthousiaste verhalen over kinetochoren aan moeten horen. Martijn, Teus en Aldert, bedankt voor het eens in de zoveel tijd afspreken om een boek en het leven te bespreken. Onmisbaar in deze lijst zijn ook mijn vrienden van uit mijn middelbare schooltijd. Anneke, Hans, Henk Jan, Carla, Teunis, Femke, Lydia, Willem, Steveline, Arco, Hadassa, Remmelt, Klaas, Marianne, Ben en Albertina, ieder jaar zie ik weer uit naar de zomervakantie. Een week lang relaxen en bijpraten was een welkome rust tussen alle deadlines van papers, proeven, presentaties en een proefschrift in. Dank voor jullie betrokkenheid op mijn leven en wetenschappelijke avonturen. Veel dank ben ik ook aan de Epifanen verschuldigd. Hoewel jullie mijn genoom ooit voor zwartgallig hebben uitgemaakt is het altijd een klein feestje om jullie ontmoeten. Ik zie dan ook weer uit naar onze ontmoeting aanstaande december als we weer voltallig hopen te zijn.

Mijn familie wil ik bedanken voor hun steun en liefde in alle dingen van het leven. Sorry dat ik op menig moment een verjaardag of ander festiviteit moest afzeggen omdat ik een experiment moest doen of dat mijn planning volledig uit de hand was gelopen. Pa en ma, leuk dat jullie een keer zijn komen kijken op het lab. Hoewel jullie misschien niet heel veel hebben meegekregen van wat ik nu eigenlijk allemaal uitspookte op de Uithof in Utrecht, heb ik mij altijd gesteund gevoeld en hebben jullie mij alle kansen gegeven om mij als zelfstandig mens te ontwikkelen. Ooit hoop ik het allemaal nog eens heel goed uit te leggen.

Lieve Gertine, jij hebt van dichtbij gezien wat het mij voor mij betekend heeft om dit proefschrift te schrijven en een promotietraject af te leggen. Tijdens al mijn geworstel met schrijven en euforie over nieuwe ontdekkingen en rare ideeën, was jij erbij. Ik ben je eeuwig dank verschuldigd voor je geduld als ik weer eens te laat thuis kwam en je liefdevolle aansporingen als ik een schop onder de kont nodig had. Ik bewonder hoe jij in het leven staat als mens, als vrouw en als pastor en ik zou me geen leven kunnen voorstellen zonder jou. Volgend jaar op avontuur naar Cambridge, ik heb er zin in! Dankjewel dat je mijn leven mooier maakt. Ik hou van jou.