## Electronic Supplementary Material

| Chapter      | Page |
|--------------|------|
| Chapter 2    | 2    |
| Chapter 3    | 17   |
| Chapter 4    | 20   |
| Chapter 5    | 29   |
| Chapter 6    | 44   |
| Supplementary References | 56   |

Files indicate with an * could not be represented in this file due to the nature of their contents and can be downloaded as separate .zip files per chapter.
CHAPTER 2

The Kinetochore Network in Eukaryotes as Revealed by Comparative Genomics

EMBO reports; 2017

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

*Sequence File S1 Orthologs for all proteins used in this study.

*Matrix S1 Pairwise correlations of the phylogenetic profiles for all proteins used in this study.

*Tree S1 Tree file of the phylogenetic analysis of Histon 3-like proteins in all euakaryotes. Associated to Figure S4.

Table S1 Measures of protein diversity in the set of kinetochore and APC/C proteins. Scores present the average across the proteins, except in the case of completeness (average across species). Statistical validity was assessed by performing unpaired, two-sided $t$-test.

Table S2 Sources of proteomes

Table S3 Phylogenetic profiles of the kinetochore proteins were compared to (sub)families defined using PANTHER10 [19]. The PANTHER10 (sub)families that have similar phylogenetic profiles as many kinetochore proteins (as measured by the Pearson correlation coefficient, indicated by their frequency in the top 30 of each kinetochore protein) are shown here.

Figure S1 Anaphase-promoting complex/cyclosome (APC/C) subunits across 90 eukaryotic lineages. Presences and absences (“phylogenetic profiles”) of APC/C subunits in 90 eukaryotic species. Top: Phylogenetic tree of the species in the genome set, with colored areas for the eukaryotic supergroups. Left side: APC/C proteins clustered by average linkage based on the pairwise Pearson correlation coefficients of their phylogenetic profiles. The orthologous sequences are available as fasta files in Sequence File 1 allowing full usage of our data for further evolutionary cell biology investigations.

Figure S2 Loss frequencies and sequence evolution of kinetochore and APC/C proteins. Scatter plots for loss frequencies and dN/dS values (A) and percent identity (B) of human–mouse orthologs for the kinetochore and APC/C proteins that were inferred to have been present in LECA. Loss frequencies and dN/dS values positively correlate ($P = 3.9e-5$, Spearman correlation), whereas loss frequencies and percent identity negatively correlate ($P = 0.0005$, Spearman correlation).

Figure S3 Copy numbers of kinetochore proteins. Heatmap indicating the copy numbers of each kinetochore protein in the 90 eukaryotic lineages. Please note that
these copy numbers might contain some over- and underestimates due to unpredicted or imperfectly predicted genes and database errors.

**Figure S4 Gene phylogeny of histone H3 homologs.** To find the putative orthologs of CenpA, we first aligned candidate orthologous sequences, which were experimentally identified centromeric H3 variants in divergent species (indicated with a pink branch in this phylogeny). From this alignment, we constructed a profile HMM and performed multiple HMM searches through our local proteome database. From these searches, we selected 831 sequences (belonging to the histone H3 family), aligned these and constructed the gene phylogeny, which is presented in this figure (see also Material and Methods). We rooted the phylogeny on the cluster that contained all of these experimentally identified centromeric H3 variants and some additional sequences that, based on best blast hits, were also likely to be orthologous to CenpA. The cluster did not contain the candidate orthologs in *Toxoplasma gondii* [1]. We do not know whether this is due to an error in the gene phylogeny, or to parallel invention of a centromeric H3 variants in this species, which would mean that it is not orthologous to CenpA. Nevertheless, we included these sequences in the orthologous group. The candidate centromeric H3 variants that are part of the CenpA cluster include sequences from all five eukaryotic supergroups: *Homo sapiens* [2], *Saccharomyces cerevisiae* [3], *Drosophila melanogaster* [4], *Caenorhabditis elegans* [5], *Schizosaccharomyces pombe* [6] (Opisthokonta), *Dictyostelium discoideum* [7] (Amoebozoa), *Arabidopsis thaliana* [8] (Archaeplastida), *Tetrahymena thermophila* [9], *Plasmodium falciparum* [10] (SAR), *Giardia intestinalis* [11] and *Trichomonas vaginalis* [12] (Excavata). The original gene tree in newick format is provided (Tree S1).

**Figure S5 Evolution of the Mad2-interacting motif (MIM) in green plants and co-occurrences of Mad2 with the MIM under a less strict motif definition.** (A) Viridiplantae (green plants) phylogeny [13] and the occurrences of the canonical MIM or the “land plant” MIM in Mad1 orthologs of the associated species. Asterisk (*) indicates species lacking an aligned MIM, possibly caused by incomplete gene prediction of Mad1 orthologs. (B) The sequence logos of the MIMs of Mad1 (upper panel) and Cdc20 (lower panel) based on the alignments of the motifs present in the right-sided panels of (C and D). Below is indicated the required amino acid sequence of the MIM (+: positive residue, Φ: hydrophobic residue, P: proline). In contrast to Figure 6, the MIM is considered present if it agrees with the pattern [ILV](2)X(3,7)P or [RK][ILV](2), in order that the land plant motif suffices. (C and D) Left side: Numbers of presences and absences of Mad2 in 90 eukaryotic species and its interaction partners Mad1 (C) and Cdc20 (D). Right side: Frequencies of Mad2 and MIM (according to definition in B) occurrences in species having Mad1 (C) or Cdc20 (D), respectively. Also the Pearson correlation coefficients (r) for the corresponding phylogenetic profiles are shown.
Figure S6 Performance of various measures that compare phylogenetic profiles in predicting physically interacting proteins. Various metrics quantify the similarity between phylogenetic profiles, such as Pearson correlation coefficient, hamming distance, chance co-occurrence probability distribution, jaccard index, mutual information [14,15] and various phylogeny-sensitive measures such as those based on Dollo parsimony [16]. We compared these metrics by assessing how well they return known physically interacting genes. A set of physically interacting proteins was obtained for our proteins of interest (kinetochore proteins) using the BioGRID [17]. For each metric, we calculated the enrichment of these confirmed interacting protein pairs among pairs having a given phylogenetic profile similarity score (converted into the coverage of all possible protein pairs at that similarity score). Across most scores, the Pearson correlation coefficient returns the highest number of interacting pairs. For this Pearson correlation coefficient ($r$), the threshold $t$ was set at the $r$ value that yields 6-fold enrichment of interacting pairs relative to pairs for which no interaction is observed.

Figure S7 $t$-SNE map of kinetochore proteins. The kinetochore proteins were visualized using a Barnes-Hut implementation of $t$-Distributed Stochastic Neighbor Embedding ($t$-SNE) [18] based on their pairwise distances measured by the Pearson correlation coefficient of the phylogenetic profiles. The protein names are colored according to their complex memberships, identical to Figure 1.

Figure S8 Establishing homology between Zwint-1, Sos7 and Kre28. Sequences that link Sos7 (Schizosaccharomyces pombe), Kre28 (Saccharomyces cerevisiae) and Zwint-1 (Homo sapiens) with homology searches (arrows) and corresponding e-values (searches against UniProtKB database performed online on October 1 2015, http://www.ebi.ac.uk/Tools/hmmer/), indicated by species and UniProt IDs. Colors represent to which protein that sequence is most similar. Of the sequences indicated here, in addition to Sos7, Kre28 and Zwint-1 also the hit in Capsaspora owczarzaki is in the proteome database used in this study. We used sequences in additional species to connect sequences in the database, as indicated by this scheme.
Table S1

| Diversity feature (p-value kinetochore vs. APC/C)                                    | Kinetochore (average) | APC/C (average) |
|---------------------------------------------------------------------------------------|-----------------------|-----------------|
| Frequency (p=0,0045)                                                                 | 0,460                 | 0,689           |
| Entropy (p=0,0248)                                                                   | 0,731                 | 0,578           |
| Pearson correlation coefficient (p=0,0006)                                             | 0,219                 | 0,267           |
| Completeness (p=3,66e-13)                                                            | 0,481                 | 0,701           |
| Loss (Dollo parsimony, p=0,140)                                                      | 16,5                  | 13,1            |
| Transitions (p=0,636)                                                                | 0,172                 | 0,184           |
| % Identity (human-mouse, p=0,0016)                                                   | 74.8%                 | 89.2%           |
| dN/dS (human-mouse, p=1,80e-5)                                                       | 0,245                 | 0,059           |
| Table S2 | Description | Summary (SI) | Detailed (SI) | References |
|---------|-------------|--------------|--------------|------------|
| S2 | | | | |

**Column headers**
- **Description**: Description of the table contents.
- **Summary (SI)**: Summary of the table in the supplementary information.
- **Detailed (SI)**: Detailed information in the supplementary information.
- **References**: References for the table.

**Table content**

| Description | Summary (SI) | Detailed (SI) | References |
|-------------|--------------|--------------|------------|

**Notes**
- SI: Supplementary Information
- References: Additional details or links provided within the supplementary information.
| PANTHER10 (sub)family | Frequency (top 30) | Pearson correlation coefficient \( (r, \text{average}) \) | Protein | Information from |
|------------------------|-------------------|---------------------------------|--------|-----------------|
| PTHR11444:SF3IARGINOSUCCINATE LYASE | 12 | 0,4914568 | ARGINOSUCCINATE LYASE | human |
| PTHR28080:IFAMILY NOT NAMED | 12 | 0,485799234 | Pex3: Peroxisomal Biogenesis Factor 3 | human |
| PTHR12309:SF12ICENTROMERE PROTEIN N | 12 | 0,770296195 | CenpN | human |
| PTHR14582:IFAMILY NOT NAMED | 11 | 0,682881415 | CenpO | human |
| PTHR23342:SF0IN-ACETYLGLUTAMATE SYNTHASE, MITOCHONDRIAL | 11 | 0,620553285 | NAGS:N-Acetylglutamate Synthase | human |
| PTHR28262:IFAMILY NOT NAMED | 10 | 0,773155189 | Spc19 | human |
| PTHR128566TRANSCRIPTION INITIATION FACTOR IIH-RELATED | 10 | 0,567686548 | GTF2H1: General Transcription Factor IIH Subunit 1 | human |
| PTHR14401:IFAMILY NOT NAMED | 10 | 0,714825979 | CenpK | human |
| PTHR10606:SF39i6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2,6-BISPHOSPHATASE YLR345W-RELATED | 10 | 0,773155189 | Similar to 6-phosphofructo-2-kinase enzymes | yeast |
| PTHR14778:IFAMILY NOT NAMED | 10 | 0,642404326 | Dsn1 | human |
| PTHR31749:IFAMILY NOT NAMED | 10 | 0,6242374 | Nsl1 | human |
| PTHR34832:IFAMILY NOT NAMED | 10 | 0,635973745 | CenpW | human |
| PTHR10555:SF136IVACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN 17 | 10 | 0,773155189 | Vps17 | yeast |
| PTHR28017:IFAMILY NOT NAMED | 10 | 0,773155189 | Dad3 | yeast |
| PTHR212866NUCLEAR PORE COMPLEX PROTEIN NUP160 | 10 | 0,533128741 | Nup160 | human |
| PTHR31382:SF4IN(+)/H(+)/ANTIPORTER | 9 | 0,764295654 | NHA1: Na+/H+ antiporter | yeast |
| PTHR24343:SF137ISERINE/THREONINE-PROTEIN KINASE RTK1-RELATED | 9 | 0,76076875 | RTK1 | yeast |
| PTHR11689:SF93IANION/PROTON EXCHANGE TRANSPORTER GEF1 | 9 | 0,76076875 | Gef1 | yeast |
| PTHR11266:SF8MPV17-LIKE PROTEIN 2 | 9 | 0,592282748 | MPV17L2: Mitochondrial Inner Membrane Protein Like 2 | human |
| PTHR31740:SF2ICENTROMERE PROTEIN L | 9 | 0,667189989 | CenpL | human |
| PTHR28051:SF1IRESISTANCE TO GLUCOSE REPRESSION PROTEIN 1 | 9 | 0,709057063 | REG1: Regulatory subunit of type 1 protein phosphatase Glc7p | yeast |
| PTHR28113:IFAMILY NOT NAMED | 9 | 0,799278566 | Dam1 | yeast |
| PTHR23139:SF60ICENTROMERE PROTEIN I | 9 | 0,661327982 | CenpI | human |
| PTHR12064:SF29IPROTEIN MAM3 | 9 | 0,756454928 | Mam3 | yeast |
| PTHR28036:IFAMILY NOT NAMED | 9 | 0,787260404 | Dad2 | yeast |
| PTHR231686MITOTIC SPINDLE ASSEMBLY CHECKPOINT PROTEIN MAD1 MITOTIC ARREST DEFICIENT-LIKE PROTEIN 1 | 9 | 0,550060967 | Mad1 | human |
| PTHR28662:IFAMILY NOT NAMED | 9 | 0,654806603 | CenpH | yeast |
| PTHR28077:IFAMILY NOT NAMED | 9 | 0,731332783 | Kei1: Kex2-cleavable | yeast |
| PTHR11365:SF11SUBFAMILY NOT NAMED | 9 | 0,731332783 | OXP1: OxoProlinase | yeast |
**Figure S5**

A) Mutation of canonical MIM

B) MIM definition

C) Mad2  Mad1

D) Mad2  Cdc20

- Canonical MIM
- Land plant MIM
- ‘Transition’ MIM

- Protein absent

---

**Legend:**
- Canonical MIM
- Land plant MIM
- ‘Transition’ MIM

**Species:**
- Arabidopsis thaliana
- Aquilegia coerulea
- Oryza sativa
- Amborella trichopoda
- Selaginella moellendorffii
- Physcomitrella patens
- Klebsormidium flaccidum
- Choroskybus atmosphyticus
- Chlamydomonas reinhardtii
- Volvox carteri
- Coccomyxa subellipsoidea
- Chlorella variabilis*
- Ostreococcus lucimarinus
- Bathycoccus prasinos
- Micromonas species*

**MIM Definition:**
- Or
- P

**Correlation Coefficients:**
- $r = 0.492$
- $r = 0.609$
- $r = 0.440$
- $r = 0.519$
Figure S6

Measure

Phylogeny-insensitive
- Chance co-occurrence probability distribution
- Jaccard index
- Mutual information
- Pearson correlation coefficient

Phylogeny-sensitive (Dollo parsimony)
- Dollo Fisher's exact
- Differential Dollo
- Dollo overall
Figure S8

Schizosaccharomyces pombe (Sos7, U3H042)

1.1e-19

Bipolaris oryzae (W6YUM4)

0.0019

Wickerhamomyces ciferrii (K0KQQ5)

2nd iteration, 0.0016

Zygosaccharomyces rouxii (C5DZ48)

7.8e-33

Saccharomyces cerevisiae (Kre28, Q04431)

1.1e-09

Strongylocentrotus purpuratus (W4Z8J2)

0.00018

Capsaspora owczarzaki (A0A0D2WQB7)

0.0054

Oncorhynchus mykiss (A0A060WEI3)

Homo sapiens (Zwint, O95229)
CHAPTER 3
Kinetochore evolution deconstructed: mapping motifs and domains

Manuscript in preparation

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

*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 Zwint1 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

Journal of Cell Biology; 2013

Mathijs Vleugel, Eelco Tromer, Manja Omerzu, Vincent Groenewold, Wilco Nijenhuis, Berend Snel & Geert JPL Kops
Supplementary Files

Figure S1 The N-terminal MDLT-KI module in KNL1 independently recruits BUB-proteins. (A) Schematic of U2OS-LacO cells and LacI-LAP-KNL1 fusion proteins. Bars, 5µm. (B) Quantification of Figure 1a as BUB1/GFP and BUBR1/GFP (n=8). (C) Schematic representation of synthetic LAP-KNL1-FL and -NC 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, N-terminal MELT-like sequence (MDLT) is indicated in blue. (D) Quantitative Mass. Spectrometry (SILAC) immunoprecipitation comparing LAP-KNL1-FL (light label) vs. LAP-KNL1-NC (heavy label). NC/FL ratio is set to 1.00, NDC80-complex subunits are in blue, MIS12-complex subunits in green, BUB-proteins in orange. (d-e) Representative images (E) and quantification (F) 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, MPS1 in red, centromeres (CREST) in blue and DNA (DAPI) in white. Bars, 5µm. Quantification in (E) shows total kinetochore signal intensity (+SD) of LAP-KNL1 and MPS1 over CREST. Data are from >15 cells and representative of 3 experiments. Levels of kinetochore MPS1 in control cells and of kinetochore LAP-KNL1 in KNL1-FL-expressing cells are set to 1. Representative images (G) and quantification (H) of MAD1 kinetochore levels as done for MPS1 in (E/F).

Figure S2 The N-terminal MDLT-KI module in KNL1 is sufficient to support SAC activity but not chromosome biorientation (A) Still images corresponding to Figure 2a. Bars, 10µm. (B) Time-lapse analysis of Flp-in HeLa cells expressing LAP-KNL1 variants, transfected with siLUC or siKNL1, and treated with nocodazole. Data (n=40 representative of three independent experiments) indicate cumulative fraction of cells that exit from mitosis (as scored by cell morphology using DIC) at the indicated time after NEB. (C) 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 three independent experiments) indicate cumulative fraction of cells that exit from mitosis (as scored by cell morphology using DIC) at the indicated time after NEB. Bars, 10µm. (D) Time-lapse analysis of Flp-in HeLa cells expressing LAP-KNL1 variants and transfected with siLUC or siKNL1. Bars, 5µm. Data (n=40 representative of three independent experiments) indicate cumulative fraction of cells that exit from mitosis at the indicated time after metaphase (as scored by GFP-H2B). (E) Still images corresponding to Figure 2e. (F) Representative images of LAP-KNL1 expressing Flp-in HeLa cells transfected with siLUC (control) or siKNL1 and treated with nocodazole. LAP-KNL1 is shown in green, BUB1 in red, centromeres (CREST) in blue and DNA (DAPI). Bars, 5µm. (G) As in Supplemental Figure 2b, except with full-length KNL1 mutated at indicated positions. (H) Quantification of chromosome alignment in Flp-in HeLa cells express-
ing LAP-KNL1 variants, transfected with siLUC or siKNL1, and treated with MG132 for 45 minutes. The data shown are from a single representative experiment out of three repeats. For the experiment shown, n=40. (I) Time-lapse analysis of Flp-in HeLa cells expressing LAP-KNL1 variants and transfected with siLUC or siKNL1. Data (n=40 representative of three independent experiments) indicate cumulative fraction of cells that exit from mitosis at the indicated time after NEB (as scored by GFP-H2B). (J) As in (I) but treated with 8.3nM nocodazole.

**Figure S3** KNL1 contains multiple independent BUB-recruitment modules. (A) Immunolocalisation of BUBR1 (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. Inserts show magnifications of the boxed regions. Bars: 5µm and 0.5µm (inserts). Table indicates the ability (- or +) to recruit BUBR1 by the indicated KNL1 fragments (see also Figure 3b). (B) Quantification of Figure 3b and Supplemental Figure 3a as BUB1/GFP and BUB1/GFP (n=8). (C) Amino acid substitutions of KNL1818-1061 mutants described in Figure 3e. (D) Quantification of Figure 3e as BUB1/GFP (n=8). (E) Immunolocalisation of BUB1 (red) in nocodazole-treated U2OS-LacO cells transfected with LacI-LAP-KNL11-261 wild-type and TΩΩ mutant. LacI-LAP-KNL1-fragments are shown in green, centromeres (CREST) in blue and DNA (DAPI) in white. Bars, 5µm and 0.5µm (inserts). Inserts show magnifications of the boxed regions. (F) Quantification of (E) as BUB1/GFP (n=8).

**Figure S4** Engineered KNL1 proteins reveal differential requirements for Ω-MELT modules in the SAC and chromosome biorientation. Representative images (A) and quantification (B) 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, BUBR1 in red, centromeres (CREST) in blue and DNA (DAPI) in white. Bars, 5µm. Quantification in b shows total kinetochore signal intensity (+SD) of LAP-KNL1 and BUB1 over CREST. Data are from >15 cells and representative of 3 experiments. Levels of kinetochore BUB1 in control cells and of kinetochore LAP-KNL1 in KNL1-FL-expressing cells are set to 1. Representative images (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, BUB1 in red, centromeres (CREST) in blue and DNA (DAPI) in white. Bars, 5µm. Quantification in b shows total kinetochore signal intensity (+SD) of LAP-KNL1 and BUB1 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. (E) Quantification of centromeric H2A-Thr120 phosphorylation in Flp-in HeLa cells expressing LAP-KNL1 variants and transfected with siLUC or siKNL1. pH2A-Thr120 is quantified
over CREST (n=10 representative of three independent experiments). Representative images (F) and quantification (G) of the indicated constructs, as in (C/D).

**Figure S5** TΩ-MELT modules in KNL1 are redundant and exchangeable. (A) Representative images 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, BUBR1 in red, centromeres (CREST) in blue and DNA (DAPI) in white. Bars, 5µm. Quantification is shown in Figure 5b. (B) Immunostaining and quantification of centromeric H2A-Thr120 phosphorylation in Flp-in HeLa cells expressing LAP-KNL1 variants and transfected with siLUC or siKNL1. T120 is quantified over CREST (n=10 representative of three independent experiments).
Figure S1

A

B

C

D

E

F

G

H

LacI-LAP-KNL1

256xLacO

Relative levels of LacO

0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
BUB1/GFP
BUBR1/GFP

PP1 & MT-binding site

MELT

KT-recruitment domain

KNL1-WT

1-261

1834

2342

KNL1-N1-261C1834-2342

D

E

F

G

H

RNase Protection

Protein
LAP-KNL1
NUF2
NDC80
SPC25
SPC24
NSL1
MIS12
PMF1
DSN1
BUBR1
BUB1
BUB3

Ratio
NC/FL
1.00*
0.75
0.75
0.74
0.73
0.68
0.68
0.68
0.67
0.17
0.14
0.13

Coverage
49.0%
73.3%
70.2%
71.9%
81.2%
65.8%
81.5%
74.6%
87.4%
11.6%
12.5%
39.3%

Bait

LAP-KNL1
MPS1
CREST
DAPI
MERGE

siLUC

siKNL1

B

G

H

KNL1-FL

KNL1-C

∆M

3

4

siLUC

siKNL1

KNL1-NC

KNL1

ΔM

3

siLUC

siKNL1

siLUC

siKNL1

siLUC

siKNL1

*Ratio KNL1-NC/KNL1-FL = set to 1.00
Figure S2

(A) Images showing entry and exit of cells during mitosis with and without Nocodazole treatment.

(B) Cumulative fraction of cells in mitosis over time with Nocodazole treatment.

(C) Cumulative fraction of cells in mitosis over time with Nocodazole and Reversine (250nM) treatment.

(D) Cumulative fraction of cells in mitosis over time without treatment.

(E) Images showing NEB during mitosis with and without Nocodazole treatment.

(F) Fluorescent images showing LAP-KNL1, BUB1, CREST, MERGE, and DAPI with different treatments.

(G) Cumulative fraction of cells in mitosis over time with Nocodazole and Reversine (250nM) treatment with fluorescence imaging.

(H) Bar graph showing the fraction of cells with misaligned chromosomes.

(I) Cumulative fraction of cells in mitosis over time without treatment with fluorescence imaging.

(J) Cumulative fraction of cells in mitosis over time with 8.3nM Nocodazole treatment with fluorescence imaging.

Legend:
- siLUC
- siKNL1
- KNL1-FL
- KNL1-NC
- KNL1-FLΔ261
Figure S3

A

LacI-LAP BUBR1 CREST MERGE DAPI

| LacI-Fragment | BUBR1 |
|---------------|-------|
| x            | -     |
| 70-261       | +     |
| 262-817      | -     |
| 818-1051     | -     |
| 1052-1292    | -     |
| 1293-1832    | -     |
| 1833-2342    | -     |

B

Relative levels at LacO (Test protein/GFP)

| LAP   | KNL1'201 | KNL1'282-817 | KNL1'818-1051 | KNL1'1052-1292 | KNL1'1293-1832 | KNL1'1833-2342 |
|-------|----------|--------------|---------------|----------------|----------------|-----------------|
| BUB1/GFP | BUB1/GFP | BUB1/GFP | BUB1/GFP | BUB1/GFP | BUB1/GFP | BUB1/GFP |

C

TΩ-motif

| M3     | 818 | 1051 |
|--------|-----|------|
|        | KIDK| TIVF |
|        | GTSE| TILY |
|        | PMDK| TVVF |
| M3-ΩA  |     |      |
|        | KIDK| AIVA |
|        | GTS | EILA |
|        | PMDK| AVVA |
| M3-MELTA |   |      |
|        | KIDK| TIVF |
|        | GTSE| TILY |
|        | PMDK| TTVF |
| A3     |     |      |
|        | KIDK| AIVA |
|        | GTS | AILA |
|        | PMDK| AVVA |
|        |     |      |

M3-MELTA

| M3-MELTA |   |      |
|          |   |      |

D

Relative levels at LacO (Test protein/GFP)

| M3     | M3-ΩA | M3-MELTA | A3  |
|--------|-------|----------|-----|
|        |       | BUB1/GFP |     |

E

LacI-KNL1'260

F

Relative levels at LacO (Test protein/GFP)

| KNL1'WT | KNL1'ΩA |
|---------|---------|
| BUB1/GFP | BUB1/GFP |

Figure S4

Panel A: Representative images showing LAP-KNL1, BUBR1, CREST, MERGE, and DAPI staining.

Panel B: Bar graph showing relative levels at kinetochores (BUBR1/CREST) for different treatments.

Panel C: Similar to panel A, with different treatments.

Panel D: Similar to panel B, with different treatments.

Panel E: Bar graph showing relative levels at kinetochores (pH2a-Thr120) for different treatments.

Panel F: Similar to panel A, with different treatments.

Panel G: Similar to panel E, with different treatments.
Figure S5

A

siLUC

| LAP-KNL1 | BUBR1 | CREST | MERGE | DAPI |
|----------|-------|-------|-------|------|
| [Image]  | [Image] | [Image] | [Image] | [Image] |

siKNL1

| KNL1-FL | KNL1^A | KNL1^A-M_3-M_3 | KNL1^A-A_3-A_3 | KNL1^A-2_3-2_3 | KNL1^A-17_3-17_3 |
|---------|--------|----------------|-----------------|-----------------|------------------|
| [Image] | [Image] | [Image] | [Image] | [Image] | [Image] |

B

Relative levels at kinetochores (pH2a-Thr120/CREST)

| pH2a-Thr120 | siLUC | siKNL1 |
|-------------|-------|--------|
| KNL1-FL     | [Image] | [Image] |
| KNL1^A      | [Image] | [Image] |
| KNL1^A-M_3-M_3 | [Image] | [Image] |
| KNL1^A-A_3-A_3 | [Image] | [Image] |
| KNL1^A-2_3-2_3 | [Image] | [Image] |
| KNL1^A-17_3-17_3 | [Image] | [Image] |
CHAPTER 5
Widespread Recurrent Patterns of Rapid Repeat Evolution in the Kinetochore Scaffold KNL1

Genome Biology and Evolution; 2015

Eelco Tromer, Berend Snel & Geert JPL Kops
Supplementary Files

*Sequence File S1* Orthologous sequences of all eukaryotic KNL1 proteins, mentioned and used in this study.

*Alignment PAML analysis* - multiple sequence alignment of concatenated repeats (xxxTxxF-MELTxSHTxxx) of 13 selected primate KNL1 sequences without gaps (PAML format, FASTA format for codons and amino acid sequences).

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

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

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

Alignment S4 Multiple sequence alignment of full-length KNL1 of various placental mammals, including mouse, rat and human. Numbers indicate the position of the repeats within the alignment (21 for human). Overall the topology of the repeat array is conserved, although mouse and rat have lost repeat 8, 9, 15 and 21 and diverged some of the other repeats as well.

Figure S1 Distinct modes of repeat evolution in different vertebrate species. Similar to Figure 3. (*) Repeats were allowed to diverge to such extent that functional residues were lost. (+) Scars of overlapping block duplications. (A) Alignment that belongs to the matrix represented in Figure 3a. (B) Alignment that belongs to the matrix in Figure 3b. (C) Homogenization of the repeat array in lamprey. The outsides of the array reveal old repeats that are in decay. (D) Repeat in zebra fish diverged to such extent that no duplication events can be inferred and the SHT motif seems to be mostly lost except for repeat 1.

Figure S2 Clustered similarity matrices for four mammal species with Turtle as outgroup. Species names correspond to the blocks under the clustered matrix, showing the number of repeats of a particular species in a cluster. Cluster colors are projected onto linear representation of repeat arrays to depict the shared duplication history of repeat units. Incomplete clustering was manually corrected (see asterisk). Repeats that did not undergo duplication are shaded to highlight the dynamic region in which many duplications have occurred. The width of the line between blocks indicates the likely order of events.
Figure S3 DNA alignment of the repeats of *Blumeria graminis* and *Petromyzon marinus*. The first panel shows all positions of the repeats. In the second panel only non-conserved positions are shown. The low amount of substitutions suggests a recent homogenization event. Note that the repeats of lamprey that are marked as being divergent (Figure S1c**'), are the most diverged repeats in the DNA alignment as well.

Figure S4 Alignment of higher order repeat block in marsupials. Alignment of larger blocks (containing multiple repeats) for opossum (*Monodelphis domestica*), Tasmanian devil (*Sarcophilus harrisii*) and platypus (*Ornithorhynchus anatinus*) reveals multiple overlapping block duplications and also ‘scars’ of this process (gaps or missing repeats in the alignment).

Figure S5 KNL1 repeat evolution in 19 drosophilid species. Similar to Figure 3c but now for 19 drosophilid species. Colors of the repeats now correspond to the consensus for all drosophilid species. Note the duplication of KNL1 in *Drosophila persimilis*, *Drosophila pseudoobscura* and *Drosophila yakuba*. The blue box indicates 1-to-1 orthology between many drosophilid species. The red box shows many multiplication events for different drosophilid species.

Figure S6 Modular evolution of the repeat consensus sequence in eukaryotes. Repeats are abstracted into four ‘slots’ similar to Figure 4b; LEGO bricks indicate conserved features (see legend). Loss, gain and other mutational events of repeat slots are projected onto the eukaryotic tree of life. Colored asterisks indicate recurrent changes to repeats in the respective species.

Figure S7 KNL1 during primate evolution. A multiple sequence alignment of concatenated repeats of KNL1 (xxTxxFxMELTxSHTxxx for each repeat, Alignment PAML analysis) in 13 selected primate species was generated using MAFFT [93]. All analyses were performed using the graphical interface PAML X [94,95]. For the tables in A and B, parameters estimates shows all the parameters estimated for each model; $\omega = dN/dS$ ratio, $p =$ proportion of sites, PSS is the number of positive selected sites under Bayes Empirical Bayes [96] with posterior probability > 0.5 and between brackets > 0.95), $l =$ the likelihood of the model given the data, $df =$ degrees of freedom, $2\Delta lnL = 2$ times the different between the likelihoods of models (Likelihood ratio test) that are compared and the $p$-value indicates the statistical significance over the difference between the two models as the LRT follows a chi-square distribution. Fits were performed under the F3x4 codon frequency model. 79 (A) Under the null model of one $dN/dS$ ratio for the whole tree, $dN/dS$ ($\omega$) was estimated 0.55, which indicates negative selection. In order to get an idea of potentially interesting branches that might be under different selective pressures we estimated rude $dN/dS$ values were calculated for each branch of the phylogenetic using the
free ratios model. Two branches, White-cheeked gibbon (blue) and the common ancestor of the Old World Monkeys + Golden Snud-nosed Monkey (red) showed a \( \omega > 1 \). To test whether these branches were under significant positive selection relative to the rest of the tree, we compared the one-ratio model with the two-ratios model for the blue and red branch. Although \( \text{dN/dS} > 1 \) for these branches, the \( p \)-values indicate that the positive selection at these branches is not significant (0.698 and 0.128 respectively). Some branches on the tree show \( \text{dS}=0 \) which means that no synonymous substitutions were estimated resulting in an unrealistically large \( \text{dN/dS} \) value. These values do not interfere with further significance testing of the models.

**B** To test whether certain sites are under selection, we fitted the MSA to models allowing for differences in the \( \omega \) among sites (including explicit proportions of sites evolving under positive selection). Three test comparing two models are available: M0 vs. M3 (one ratio versus three ratios) M1a vs. M2a (negative/neutral versus negative/neutral/positive selection), M7 vs. M8 (\( \omega \) follows a beta distribution between 0 and 1 vs. M7 + extra class where \( \omega > 1 \)). Likelihood ratio tests (LRT) indicated that models for positive selection (M2a and M8) do not fit the data significantly better than that of neutral/negative selection (M1a and M7) \( p \)-value is 0.364 and 0.0829 respectively. Comparison of the one ratio model (model 0) and 3-classes model (model 3) shows a class of sites with \( \omega = 1.35 \), which significantly fits the data better then a one-ratio model (\( p \)-value = 1.16e-07). The Naïve Empirical Bayes (BEB is not estimated) values for model 3 are not reliable given the low amount of sites and number of taxa so these are not included. Given that only one of the three tests for positive selection (M0 vs. M3) is significant, we cannot robustly show positive selection impinging on KNL1 repeat array.
**Figure S1**

**A Species:** Human  
**Type:** single block duplication

---

**B Species:** Platypus (Ornithorhynchus anatinus)  
**Type:** duplication multiplication

---

**C Species:** Lamprey (Petromyzontiformes marinus)  
**Type:** homogenization

---

**D Species:** Danio rerio (Zebrafish)  
**Type:** repeat decay / loss of SHT except for repeat 1
Figure S2

Similarity to consensus

Corrected pairwise alignment score

repeat number 1,4 and 2,5 have a common ancestor

pattern of duplication?

Assigned Clusters

T. chinensis
S. harrisii
M. domestica
O. anatinus
P. sinensis

1-to-1 orthologous
dynamic region

repeat consensus sequence
Figure S7

A Branch-models

- Olive Baboon
- Rhesus Macaque
- Crab-eating Macaque
- Green Monkey
- Golden Snub-nosed Monkey
- White-cheeked gibbon
- Sumatran Orangutan
- Chimpanzee
- Human
- Black-capped Squirrel Monkey
- Common Marmoset
- Philippine Tarsier
- Galago

| Branch       | Model         | dN/dS (ω) | Parameter estimates | l   | df | 2ΔlnL | p-value |
|--------------|---------------|-----------|---------------------|-----|----|-------|---------|
| all          | 0: one ratio  | 0.55      | ω₀ = 0.55           | -4410.69 | - | - | - |
|              | 2: two ratio  | 0.53, 1.03| ω₀=0.53, ω₁=1.03    | -4409.53 | 1 | 2.32 | 1.28e-01 |
|              | 2: two ratio  | 0.53, 3.84| ω₀=0.53, ω₁=3.84    | -4407.66 | 1 | 3.74 | 5.31e-02 |
|              | all free-ratio| see tree  | see tree            | -4391.64 | 23 | 38.1 | 6.98e-01 |

B Sites-models

| Model       | dN/dS | Parameter estimates | PSS | l   | df | 2ΔlnL | p-value |
|-------------|-------|---------------------|-----|-----|----|-------|---------|
| 0: one ratio| 0.55  | ω₀ = 0.55           | -   | -4410.69 | 4 | 37.92 | 1.16e-07 |
| 3: discrete | 0.61  | p₀=0.24, p₁=0.42, p₂=0.34, ω₀=0.24, ω₁=1.35, ω₂=1.35 | -   | -4391.73 | 2 | 2.02 | 3.64e-01 |
| M1a: neutral| 0.56  | p₀=0.53, p₁=0.47, ω₀=0.17, ω₁=1 | -   | -4392.74 | 2 | 2 | 3.64e-01 |
| M2a: selection | 0.61 | p₀=0.66, p₁=0.34, ω₀=0.24, ω₁=1, ω₂=1.35 | 21(0) | -4391.73 | | |
| M7: beta    | 0.56  | p=0.34, q=0.27      | -   | -4394.24 | 2 | 4.98 | 8.29e-02 |
| M8: beta β₀ | 0.61  | p₀=0.67, p₁=0.33, p=32.14, q=99.00, ω=1.35 | 36(0) | -4391.75 | | |
CHAPTER 6
Phylogenomics-guided discovery of a novel conserved cassette of short linear motifs in BubR1 essential for the spindle checkpoint

Open Biology; 2016

Eelco Tromer, Debora Bade, Berend Snel & Geert JPL Kops
Procedures and Discussion

Phylogenomic analyses
Evolutionary relevant and divergent eukaryotic species were selected, in addition to our previously used set [97], based on their position relative to duplications and the inclusion of newly sequenced domains of the eukaryotic tree of life. Since the TPR domain is shared by all MadBub orthologs, we used HMMsearch [98] to capture and align the TPR domains in our dataset containing 152 genes in total. We selected only those columns of the multiple sequence alignment that had an occupancy of 80% or higher. RAxML [99] was used to perform phylogenetic analysis on 129 positions (2.1% gaps). A phylogenetic tree was estimated using the evolutionary model, selected by ProtTest [100] (LG + G). Parameters were set to be estimated, where possible. Confidence for the resulting maximum likelihood tree was assessed by a bootstrap analysis (1000 replicates). A schematic representation of the phylogenetic tree can be found in Figure 1b and Figure S1b based on our reconciliation of the maximum likelihood gene tree with the known species tree (Figure S1a, for species names see Table S1), species taxonomy (Uniprot and newly published) and motif/domain content.

Similar to our previous findings, the maximum likelihood tree topology is fully inconsistent with a single duplication explaining all events in the MadBub gene family but neither are all independent duplications unambiguous and with maximal support present in the gene tree [97]. We could however still infer a manually reconciled tree based on the following pieces of information: First in some cases known whole genome duplications and their syntenic conservation provide unambiguous phylogenetic timing of duplication despite poorly supported or wrong topology in the gene tree. Second the presence of a single full-length MadBub protein in the genomes of closely-related species that branch of just before a species where a BUB and a MAD protein are both present in the tree, even if they are not precisely inferred where they should be. There can be many reasons why such a short piece of sequence would hinder the correct inference of the evolutionary history of this gene family besides general lack of phylogenetic signal in so few amino acids. In some cases these inconsistencies are likely explained by the increased rate of evolution of one of the paralogs after duplication (mostly BUB). Low bootstrap support values furthermore signified the incorrect placement of a number of species, relative to the species tree (although most of the major eukaryotic supergroups were recovered). And thus in general if we would perform strict tree reconciliation on our full gene tree, unrealistic losses and or duplications had to be inferred. In any case, our current analysis strengthens our previous conclusion on recurrent independent duplications of an ancestral MadBub gene. We therefore focused on the duplication events in specific
taxa and chose to guide our reconciliation by the motif/domain content if applicable. Relevant new evidence for duplication events is discussed in short below.

Our current analysis corroborated the 10 independent duplications previously described (vertebrates {#6,7}, diptera {#9}, nematodes {#8}, two in land plants {#13-15}, saccharomycetaceae {#2}, schizosaccharomyces {#3}, Laccaria bicolor {#5, agaricomycetes}, Phycomyces blakesleeanus {#1, mucorales} and Naegleria gruberi {#16} see Figure S1a,b [97]. By the addition of sequences of recently sequenced genomes, we aimed to time the duplications more accurately and determine whether patterns of subfunctionalization were consistent between species after duplication. Unfortunately, the increased number of species did not aid to resolve the uncertain placement (see * in Figure S1a) of duplications for schizosaccharomyces {#3} and the mucorales {#1}. In addition, we could not detect any excavate MadBub-like sequence, other then Naegleria gruberi {#16}.

The presence of a single MadBub homolog in the early-branching (proto-) vertebrate Petromyzon marinus (lamprey) with an intact kinase domain, suggested that the duplication in vertebrates occurred after the divergence of lamprey. In our tree however, pmMadBub groups with the MAD paralog (RAxML 20) and it lacks the CMI motif (lost in vertebrate MAD, although present in Callorhinchus milli MAD). Furthermore, the placement of lamprey relative to the major whole genome duplication in vertebrates is currently still under debate [101]. Consistent with the hypothesis of another whole genome duplication (3R) at the base of teleost fish, we find a third MadBub homolog in Danio rerio. We could however not include this gene in our maximum likelihood analysis, since it lacked a N-terminal TPR domain. Strikingly, in most other teleost lineages this extra homolog has disappeared and only MAD A has remained (which lost its C-terminus), illustrating the potential fate of the MAD (BUBR1) paralog in other vertebrate lineages. Increased branch-lengths for the nematode and to a lesser extent diptera MadBub paralogs are reflected by the motif loss (CMI and KARD (only nematodes)) and degenerate nature of the motifs and domains in general (loss of KEN2 in nematodes), indicating extensive rewiring of SAC signaling after duplication in these taxa.

To elucidate the intriguing consecutive duplications in plants, we added MadBub orthologous sequences of early-branching plants (Klebsormidium flaccidum and Marchantia polymorpha) and a number of flowering plants (Amborella trichopoda, Aquilegia coerulea and Oryza sativa japonica). Reconciliation, taking into account the number of MadBub homologs, suggested a duplication in embryophytes (Physcomitrella patens + Selaginella moellendorfii – low support and unclear topology) followed by a duplication of the BUB-like paralog in magnoliaphytes (RAxML 48). However, careful consideration of the motif and domain content of these sequences allowed for a more parsimonious explanation (from the domain/motif
perspective): (1) independent duplication in *Physcomitrella patens* (both paralogs have a GLEBS), (2) duplication in tracheophytes (loss of GLEBS, subfunctionalization into MAD (KEN1-ABBA1-KEN2-ABBA2-MadA) and BUB (CMI-ABBA-CDII-kinase) and (3) duplication in magnoliophytes (consecutive subfunctionalization of BUB into CDII+kinase (BUB A) and CMI+ABBA (BUB B)).

The four duplications we previously found in fungi, urged us to extend our search for duplications in newly sequenced species. We could more accurately time a previously found duplication in edible basidiomyceteous fungi (#5) of the agaricomycetes: early-branching lineages were found to only have a single MadBub homolog, containing all functional features of MAD and BUB (*Exidia glandulosa* and *Rhizoctonia solani*). Searching for additional MAD and BUB sequence we found an independent duplication in the basal basidiomycete fungi clade of the pucciniomycetes (#4), *Melampsora larici-populina*, RAxML 22). (for the phylogeny of basidiomycetes see [102]).

Strikingly, we found evidence for three additional duplications in stramenopile species of the SAR super group (*Albugo laibachii* #10, *Ectocarpus siliculosus* #11 and *Aureococcus anophagefferens* #12). Although, the duplication of *Aureococcus anophagefferens* (RAxML 33) is not well supported and *Ectocarpus siliculosus* MAD and BUB are grouped in different parts of the tree, the presence of an ancestral MadBub gene in a number of stramenopile lineage (diatoms and oomycetes), do not support a common ancestral duplication to have given rise to MAD and BUB in stramenopiles.

**Detailed discussion of BUB-related motifs and domains**

We observed three sub-clusters for the BUB-like paralog in our conserved feature correlation analysis (Figure 1b, Figure 2a): (1) The CDII-kinase ($r = 0.94$) cluster represents the most coherent BUB-associated cluster, having the highest anti-correlation score with the predominant MAD-associated features (-0.7 < $r$ < -0.64). Strikingly, the CDII and the kinase domain are occasionally lost in a number of SAR and Archeaplastida species (e.g. stramenopiles, green –and red algae species, see Figure S1b, Sequence File S1, Table S2). Although loss may reflect the common problem of gene prediction programs to correctly predict either amino –or carboxy terminal regions, the parallel nature of this events in different lineages advocates true loss, and would provide an opportunity to discover co-evolving features in other SAC-related protein such as members of the chromosomal passenger complex, which are localized through the catalytic activity of the BUB-related kinase domain [103]. (2) The clustering of CMI and ‘ABBA other’ motifs ($r = 0.56$), although in close proximity in many species, is best illustrated by the secondary subfunctionalization of BUB A (TPR-CDII-kinase) and BUB B (TPR-[CMI-ABBA]²) in flowering plants (Figure
1b (15)} following a duplication the ancestral BUB (TPR-CMI-CDII-kinase; vascular plants). In addition, the recurrent loss (e.g. in mucorales {1}, diptera {9} and the proto-vertebrate lamprey) and repeated nature in distinct lineages (archeoplastids, G. theta and E. huxleyi, see Figure 1b, Figure S1b), signify a distinct role of the CMI motif (+/- ABBA motif). Recent reports suggest that the region encompassing CMI is part of the elusive MAD1 kinetochore localization module, either through phospho-regulated interaction [104], through loading [105] or in an unknown manner through the RZZ complex [106]. Given the limited phylogenetic distribution of the RZZ complex to mainly opisthokonts species [107], the latter option does not seem the most likely function of the CMI motif. We favor the recently advocated template hypothesis[108], in which the strong correlation of CMI and nearby ABBA motif signify a scaffold for MAD1 and CDC20, providing a platform for the formation of a C-MAD2-CDC20 dimer, primed to bind the MAD paralog. (3) The KARD motif was mainly detected in opisthokonts, as part of both BUB (fungi and vertebrates) and MAD (diptera and vertebrates). The presence of this motif in the rhizarium Bigelowiella natans, cryptophyte Guillardia theta and the red algae Cyanidioschyzon merolae suggested an origin in LECA, but we deem it more likely that the KARD motif evolved de novo in these lineages or are false positive hits due to the degeneracy of the motif definition (Figure S1b). The KARD - GLEBS association (r=0.43) illustrates the need for co-presence of these domains at the kinetochore. A similar pattern was observed for GLEBS and CMI (r=0.39, e.g. in MAD in Albugo laibacchii and Aureococcus anophagefferens) (Figure 1b, Figure S1b). Although crucial for proper kinetochore localization of MAD and BUB, through Bub3, the lineage –specific divergent sequences surrounding the GLEBS domain (N-terminal region of the defined feature by ConFeaX, Figure 1a) indicate plastic evolution of the GLEBS-BUB3 kinetochore interaction, reminiscent of the widespread recurrent patterns of rapid repeat evolution of its major localizing phospho-motif, termed MELT [109]. In addition we observed the loss (vascular plants {13}, MAD: schizosaccharomyces {3}, basidiomycetes {4,5}) or occasional duplication of the GLEBS domain in animal lineages (diptera {9}, Zootermopsis nevadensis, Daphnia magna, Nematostella vectensis). A recently retracted paper of Paganelli et al. [110,111] reported the novel interaction of the MAD and BUB B paralogs in Arabidopsis thaliana with MAP65-3, a member of an extensively diversified gene family in plants, which is orthologous to the human anti-parallel microtubule-crosslinking protein PRC. Interestingly, upon closer examination of this protein family, we discovered the de novo evolution of a GLEBS domain in a specific subset of the MAP65 paralogs (3 and 4). These findings suggest that MadBub kinetochore localization is regulated in a different/novel manner between species and maybe subject to forces favoring rapid evolution (positive selection).
Supplementary Files

*Sequence File S1 Full-length sequences of MadBub orthologs used in this study. Fasta file containing full-length sequences of MadBub gene family specifically selected for this study. Headers include a four-letter species code (see for ID conversion Table S1) followed by MADBUB, BUB or MAD to indicate the ancestor and the two subfunctionalized paralogs, respectively (e.g. >HSAP_BUB or DDIS_MADBUB).

*Sequence File S2 Conserved features of MadBub gene family detected by ConFeaX + TPR and kinase domain. Fasta file containing motifs and domains discovered by ConFeaX, including the TPR and kinase domain. Headers include four-letter species code, homolog type (MADBUB, BUB or MAD), domain name and position in the protein sequence (e.g. >HSAP_BUBIIkinase/777-1038).

Table S1 Species ID + full names table. This table can be used to look up species names and associated taxonomy for four letter codes used in supplementary sequences files and table II.

Table S2 Matrix of features in all MadBub orthologs. Frequency matrix of conserved features reported by our ConFeaX pipeline for the MadBub gene family. Conserved functional motifs and domains are organized and colored in similar fashion as Figure 1, 2. Names of duplicated species are in italic.

Table S3 Primers used for molecular cloning.

Figure S1 Phylogenetic analysis of the MadBub gene family. (A) Maximum likelihood tree of the TPR region of 148 MadBub sequences. Blue circles indicate bootstrap support and the dashed red lined squares and asterisk (*) indicate which clades are associated with duplications. For further discussion see Procedures and Discussion. (B) Schematic representation of our reconciliation of the tree in panel A with the eukaryotic tree of life. This shows 16 independent duplication events throughout eukaryotic evolution. Arrows indicate duplications: orange (uncertain) and red (high confidence). Question marks point out clades in which the placement of the duplications can be debated, see for Procedures and Discussion. Numbers in the tree correspond to duplications in specific taxa – containing the following species: {1}-mucorales; {2}-saccharomycetaceae; {3}-schizosaccharomyces; {4}-pucciniomycetes; {5}-agaricomycetes (excluding early-branching species); {6}-vertebrates; {7}-teleost fish; {8}-nematodes; {9}-diptera (flies); {10}-albuginaceae (oomycete); {11}-ectocarpales (brow algae); {12}-aureococcus (harmful algae bloom); {13}-bryophytes (mosses); {14}-tracheophytes (vascular plants); {15}-magnoliophytes (flowering plants); {16}-naegleria;
Figure S2 – multiple sequence alignment of ABBA1-KEN2-ABBA2 cassette in all species used for this study. Multiple sequence alignment of the ABBA1-KEN2-ABBA2 region in MADBUB and MAD paralogs of all sequences used in this study. The sequences of the alignment are grouped by relevant taxonomic levels. Colors are according to the Clustal scheme. Conservation is highlighted per group.
| mutant   | 5'-3' sequence                                      | orientation |
|----------|-----------------------------------------------------|-------------|
| dABBA1   | CCATTTCCTCAACAGATGCAAATAATAGTGCTGATGAGGCTTTCTACAGCAGAAGTTGTCT | FW          |
| dABBA1   | AGACAAACTCTGCTGTAGAAAGGCTCTACATGCAACTATTATTTCATGCTGTAGAGAATTG | RV          |
| dABBA2   | TCACTGATAGCTGTACCCGCTGCTTCCCACCTGCACAACAGCAGGCTTACGCACCACGT   | FW          |
| dABBA2   | ACATGGTGTCATAACTGGCTTTTGCTGAGTGAGGCGTACAGCTATCATGTA           | RV          |
| W1A      | CAGTCCAGCCAgCATAGCACCCCCC                             | FW          |
| W2A      | GGAGGTGCTACCGTGTCAGAAGCAAGGCCTGCTTCCAGGGCTTGAGGAA          | FW          |
| W2A      | CCTGCTCTGTGGCGAGGCCCTGCTTGC                           | RV          |
| KEN1-AAA | GATGAATGGGAACTGAGTGACAAGCTGACATACCTTTACACGTGGAAGCCAG      | FW          |
| KEN1-AAA | CCTTGCCCTAAAGGTGTACAGCTGCTGCACTGATTCCCTACATC           | RV          |
| KEN2-AAA | CCATGCCCAGCGACGCAGGCAGTGCTGAGCTGCAAGCC                 | FW          |
| KEN2-AAA | GCTTGCAAGCTCGCTGACGCCCTGCTGACATGCAAGGCAAGG             | RV          |
| dDBOX    | GAGTCTTCTGACCAAAAGAGCAAAAGGAAAAGGCAAGGCAAGGC         | FW          |
| dDBOX    | GCTGTCTTTTTCCCTTTGCTTTTTGTGTACAGAAGTGACTC           | RV          |
| dABBA3   | GGTCCCCAGTGACTTTTTCTCTCTCTCTCATGAAAAGAAAGG             | FW          |
| dABBA3   | CTTCTTTCTGAAAGAGAAGAGGTACACTGGGACC                    | RV          |
Figure S2

fungi

amoebozoa

holozoa

vertebrates

metazoa

archeplastids

SAR
**Supplementary References**

1. 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.

2. Sullivan KF, Hechenberger M, Masri K. 1994 Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol* **127**, 581–592.

3. Meluh PB, Yang P, Glowczewski L, Koshland D, Smith MM. 1998 Cse4p Is a Component of the Core Centromere of Saccharomyces cerevisiae. *Cell* **94**, 607–613.

4. Blower MD, Karpen GH. 2001 The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* **3**, 730–739.

5. Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S. 1999 Cell division: A histone-H3-like protein in C. elegans. *Nature* **401**, 547–548.

6. Takahashi K, Chen ES, Yanagida M. 2000 Requirement of Mis6 Centromere Connector for Localizing a CENP-A-Like Protein in Fission Yeast. *Science* **288**, 2215–2219.

7. Dubin M, Fuchs J, Gräf R, Schubert I, Nellen W. 2010 Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Res.* **38**, 7526–7537.

8. Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S. 2002 Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *Plant Cell* **14**, 1053–1066.

9. 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.

10. Hoeijmakers WAM et al. 2012 Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. *Cell. Microbiol.* **14**, 1391–1401.

11. Dawson SC, Sagolla MS, Cande WZ. 2007 The cenH3 histone variant defines centromeres in Giardia intestinalis. *Chromosoma* **116**, 175–184.

12. Zubáčková Z, Hostomská J, Tachezy J. 2012 Histone H3 Variants in Trichomonas vaginalis. *Eukaryot. Cell* **11**, 654–661.

13. Hori K et al. 2014 Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* **5**, 3978.

14. Wu J, Kasif S, DeLisi C. 2003 Identification of functional links between genes using phylogenetic profiles. *Bioinformatics* **19**, 1524–1530.

15. Glazko G V, Mushegian AR. 2004 Detection of evolutionarily stable fragments of cellular pathways by hierarchical clustering of phylectic patterns. *Genome Biol.* **5**, R32.

16. 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.

17. Chatr-aryamontri A et al. 2017 The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379.
18. Maaten L van der, Hinton G. 2008 Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
19. 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.
20. Gobler CJ et al. 2011 Niche of harmful alga *Aureococcus* anophagefferens revealed through ecogenomics. *Proc Natl Acad Sci U S A* 108, 4352–4357.
21. Clarke M et al. 2013 Genome of *Acanthamoeba* castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 14, R11.
22. Holt RA et al. 2002 The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149.
23. Kemen E et al. 2011 Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of *Arabidopsis thaliana*. *PLoS Biol.* 9, e1001094.
24. Srivastava M et al. 2010 The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466, 720–726.
25. *Arabidopsis Genome Initiative*. 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
26. Albert VA et al. 2013 The Amborella Genome and the Evolution of Flowering Plants. *Science* 342, 1241089–1241089.
27. Putnam NH et al. 2008 The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–1071.
28. Delnooed F et al. 2011 Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. *Genome Biol* 12, R29.
29. Ghedin E et al. 2007 Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, 1756–1760.
30. Yook K et al. 2012 WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.* 40, D735–D741.
31. Tanifuji G, Onodera NT, Brown MW, Curtis B a, Roger AJ, Ka-Shu Wong G, Melkonian M, Archibald JM. 2014 Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella* oceanica: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. *BMC Genomics* 15, 374.
32. Moreau H et al. 2012 Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol* 13, R74.
33. Chang Y et al. 2015 Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. *Genome Biol Evol* 7, 1590–1601.
34. C. elegans Sequencing Consortium. 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–8.
35. Dujon B et al. 2004 Genome evolution in yeasts. *Nature* 430, 35–44.
36. Inglis DO et al. 2012 The Candida genome database incorporates multiple Candida species: multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. *Nucleic Acids Res.* 40, D667–D674.
37. Dehal P et al. 2002 The draft genome of *Ciona intestinalis*: insights into
chordate and vertebrate origins. *Science* **298**, 2157–2167.

38. Matsuzaki M *et al.* 2004 Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657.

39. Loftus BJ *et al.* 2005 The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321–1324.

40. Abrahamsen MS *et al.* 2004 Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445.

41. Price DC *et al.* 2012 *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335**, 843–847.

42. Merchant SS *et al.* 2007 The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250.

43. Blanc G *et al.* 2012 The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol* **13**, R39.

44. Blanc G *et al.* 2010 The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**, 2943–2955.

45. Eichinger L *et al.* 2005 The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57.

46. Adams MD *et al.* 2000 The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

47. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ. 2015 FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. *Nucleic Acids Res*.

48. Howe K *et al.* 2013 Corrigendum: The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **505**, 248.

49. Loftus B *et al.* 2005 The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868.

50. Read BA *et al.* 2013 Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**, 209–213.

51. Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ. 2010 The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat. Commun.* **1**, 1–7.

52. Cock JM *et al.* 2010 The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621.

53. Morrison HG *et al.* 2007 Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* **317**, 1921–1926.

54. Chan CX, Bhattacharya D. 2013 Analysis of horizontal genetic transfer in red algae in the post-genomics age. *Mob. Genet. Elements* **3**, e27669.

55. Curtis B a *et al.* 2012 Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59–65.

56. Bennett SM, Noor M a F. 2009 Molecular evolution of a *Drosophila* homolog of human BRCA2. *Genetica* **137**, 213–219.

57. Ivens AC *et al.* 2005 The genome of the kinetoplastid parasite, *Leishmania* major. *Science* **309**, 436–442.

58. King N *et al.* 2008 The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788.

59. Lee S, Thebault P, Freschi L, Beaufils S, Blundell TL, Landry CR, Bolanos-Garcia VM, Elowe S. 2012 Characterization of spindle checkpoint kinase mps1
reveals domain with functional and structural similarities to tetratricopeptide repeat motifs of Bub1 and BubR1 checkpoint kinases. J. Biol. Chem. 287, 5988–6001.

60. Lee SC et al. 2014 Analysis of a food-borne fungal pathogen outbreak: virulence and genome of a Mucor circinelloides isolate from yogurt. MBio 5, e01390-14.

61. Waterston RH et al. 2002 Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

62. Worden AZ et al. 2009 Green evolution and dynamic adaptations revealed by genomes of the marine picocynaryotes Micromonas. Science 324, 268–272.

63. Galagan JE et al. 2003 The genome sequence of the filamentous fungus Neurospora crassa. Nature 422, 859–868.

64. Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC. 2012 Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropsis gaditana. Nat. Commun. 3, 686.

65. Fritz-Laylin LK et al. 2010 The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility. Curr. Biol. 140, 631–642.

66. Putnam NH et al. 2007 Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317, 86–94.

67. Palenik B et al. 2007 The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci U S A 104, 7705–7710.

68. International Rice Genome Sequencing Project. 2005 The map-based sequence of the rice genome. Nature 436, 793–800.

69. Swart EC et al. 2013 The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol 11, e1001473.

70. Gardner MJ et al. 2002 Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419, 498–511.

71. Haas BJ et al. 2009 Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461, 393–398.

72. Heidel AJ et al. 2011 Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res 21, 1882–1891.

73. Rensing SA et al. 2008 The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64–69.

74. Arnaiz O, Sperling L. 2011 ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate Paramecium tetraurelia. Nucleic Acids Res. 39, D632–D636.

75. Bowler C et al. 2008 The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature 456, 239–244.

76. Goffeau A et al. 1996 Life with 6000 genes. Science 274, 546,563-567.

77. Simakov O et al. 2015 Hemichordate genomes and deuterostome origins. Nature 527, 459–465.

78. Berriman M et al. 2009 The genome of the blood fluke Schistosoma mansoni. Nature 460, 352–358.

79. Shoguchi E et al. 2013 Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. Curr Biol 23, 1399–1408.

80. Banks JA et al. 2011 The Selaginella genome identifies genetic changes
associated with the evolution of vascular plants. Science 332, 960–963.

81. Wood V et al. 2002 The genome sequence of Schizosaccharomyces pombe. Nature 415, 871–880.

82. Fairclough SR et al. 2013 Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. Genome Biol 14, R15.

83. Srivastava M et al. 2008 The genome sequence of Schizosaccharomyces pombe. Nature 454, 955–960.

84. Berriman M et al. 2005 The genome of the African trypanosome Trypanosoma brucei. Science 309, 416–422.

85. Armbrust E V et al. 2004 The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science 306, 79–86.

86. Aparicio S et al. 2002 Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297, 1301–1310.

87. Eisen JA et al. 2006 Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol 4, e286.

88. Stover NA, Krieger CJ, Binkley G, Dong Q, Fisk DG, Nash R, Sethuraman A, Weng S, Cherry JM. 2006 Tetrahymena Genome Database (TGD): a new genomic resource for Tetrahymena thermophila research. Nucleic Acids Res 34, D500–3.

89. Carlton JM et al. 2007 Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science 315, 207–212.

90. Kamper J et al. 2006 Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature 444, 97–101.

91. Prochnik SE et al. 2010 Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science 329, 223–226.

92. Hellsten U et al. 2010 The genome of the Western clawed frog Xenopus tropicalis. Science 328, 633–636.

93. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

94. Yang Z. 2007 PAML Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24, 496–503.

95. Xu B, Yang Z. 2013 PAMLX: A Graphical User Interface for PAML. Mol. Biol. Evol. 30, 2723–2724.

96. Yang Z, Wong WSW, Nielsen R. 2005 Bayes \textsc{E}mpirical \textsc{B}ayes Inference of Amino Acid Sites Under Positive Selection. Mol. Biol. Evol. 22, 1107–1118.

97. Suijkerbuijk SJE et al. 2012 The Vertebrate Mitotic Checkpoint Protein BUBR1 Is an Unusual Pseudokinase. Dev. Cell 22, 1321–1329.

98. Finn RD, Clements J, Eddy SR. 2011 HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.

99. Stamatakis A. 2014 RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

100. Darriba D, Taboada GL, Doallo R, Posada D. 2011 ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

101. Smith JJ et al. 2013 Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat. Genet. 45, 415–421.
102. Nagy LG et al. 2016 Comparative Genomics of Early-Diverging Mushroom-Forming Fungi Provides Insights into the Origins of Lignocellulose Decay Capabilities. *Mol. Biol. Evol.* **33**, 959–70.

103. 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.

104. London N, Biggins S. 2014 Mad1 kinetochore recruitment by Mps1-mediated phosphorylation of Bub1 signals the spindle checkpoint. *Genes Dev.* **28**, 140–152.

105. 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.

106. 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.

107. Vleugel M, Hoogendoorn E, Snel B, Kops GJPL. 2012 Evolution and Function of the Mitotic Checkpoint. *Dev. Cell* **23**, 239–250.

108. Musacchio A. 2015 The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Curr. Biol.* **25**, R1002–R1018.

109. 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.

110. 2016 Retraction. *New Phytol.* **212**, 1106–1106.

111. Paganelli L et al. 2015 Retracted: Three BUB1 and BUBR1/MAD3-related spindle assembly checkpoint proteins are required for accurate mitosis in Arabidopsis. *New Phytol.* **205**, 202–215.