Review TRENDS in Biochemical Sciences Vol.31 No.12

Centromeres put epigenetics in the driver's seat

R. Kelly Dawe[1,2] and Steven Henikoff[3]

[1] Department of Plant Biology, University of Georgia, Athens, GA 30602, USA
[2] Department of Genetics, University of Georgia, Athens, GA 30602, USA
[3] Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, WA 98109-1024, USA

A defining feature of chromosomes is the centromere, the site for spindle attachment at mitosis and meiosis. Intriguingly, centromeres of plants and animals are maintained by both sequence-specific and sequence-independent (epigenetic) processes. Epigenetic inheritance might enable kinetochores (the structures that attach centromeres to spindles) to maintain an optimal size. However, centromeres are susceptible to the evolution of 'selfish' DNA repeats that bind to kinetochore proteins. We argue that such sequence-specific interactions are evolutionarily unstable because they enable repeat arrays to influence kinetochore size. Changes in kinetochore size could affect the interaction of kinetochores with the spindle and, in principle, skew Mendelian segregation. We propose that key kinetochore proteins have adapted to disrupt such sequence-specific interactions and restore epigenetic inheritance.

The centromeric state

Every chromosome requires a kinetochore (see Glossary), the proteinaceous structure that forms the interface between centromeric DNA and the microtubules that pull the chromosomes to the poles at mitosis. The kinetochore–DNA interaction is mediated by a small group of proteins that builds the kinetochore from the DNA up [1]. Recent data show that both the DNA and proteins that comprise plant and animal kinetochores are rapidly evolving, despite the fact that the positions of centromeres are maintained for millions of years. Moreover, human centromeres can form spontaneously in chromosomal regions that lack centromere-specific sequences, suggesting that centromeres are maintained epigenetically. Here, we discuss how the evolutionary forces that act on centromeres might have resulted in their peculiar sequence organization.

The foundation of the kinetochore is the centromere-specific nucleosome [2], which differs from bulk nucleosomes by the presence of a histone H3 variant. The first example of a centromere-specific H3 variant, centromere protein (CENP)-A, was described as the smallest of three centromere-specific proteins (CENP-A, -B and -C) detected using anti-scleroderma antibodies [3,4]. CENP-A was later shown to be a core histone that replaces ordinary histone H3 in centromeric nucleosomes [5,6]. In yeast, all known components of the kinetochore are dependent on the presence of a centromere-specific nucleosome [1] and the available data from animals and plants indicate that the same is true in complex eukaryotes. Other conserved kinetochore proteins, CENP-C, Mis12 and CENP-H, depend on the presence of centromeric nucleosomes for localization [7–13]. These 'foundation' proteins [9] interact in an as yet unknown way to form a higher-order chromatin complex that, in turn, recruits dozens of transitory kinetochore proteins that are required to ensure accurate chromosome segregation.

The high degree of divergence of CENP-A is surprising in light of the absolutely conserved function of centromeric nucleosomes and the near invariance of H3 in bulk chromatin. CENP-A is significantly more closely related to H3 than to centromeric histones from distant clades, which

Glossary

α satellite: The name of the 171-base-pair human centromeric repeat.
Centromeres: DNA sequences that interact with the kinetochore.
CEN chromatin: A term used here to describe the mixture of histone H3 and CENH3 that underlies the kinetochore. The histone variant H3.3 might also be present in CEN chromatin.
Cohesin: The protein complex that mediates sister-chromatid cohesion.
Epigenetics: Heritable changes in phenotype that are not caused by changes in DNA sequence.
Gametogenesis: The formation of gametes, starting from meiosis and proceeding through the differentiation of eggs and sperm.
Kinetochore: The proteinaceous structure that connects the centromere to the spindle.
Kinetochore-foundation proteins: The small set of kinetochore proteins that are intimately associated with centromeric DNA. Foundation proteins are constitutive, that is, detectable at centromeres throughout the cell cycle. The major (evolutionarily conserved) foundation proteins are CENH3, CENP-C, Mis12 and CENP-H. Kinetochore-foundation proteins recruit outer kinetochore proteins.
Meiotic drive: The preferential segregation of one chromosome over another. The term is usually used to describe selfish DNA that accumulates by circumventing Mendel's rules.
Neocentromeres: Centromeres that arises spontaneously in a new position, often in an area that does not contain repeat arrays.
Outer kinetochore proteins: A large collection of proteins that mediate, facilitate or regulate the interaction of the kinetochore with the spindle. Outer kinetochore proteins tend to be present only during cell division.
Pericentromeric heterochromatin: The deeply staining chromatin that flanks CEN chromatin on either side. Unlike CEN chromatin, pericentromeric heterochromatin does not have defined boundaries.
Repeat arrays: Long series of short DNA sequences arranged in tandem. Centromeric repeat arrays are usually 150–180 base pairs and might extend uninterrupted for hundreds of kilobases.
Selfish DNA: DNA that accumulates in the genome without affecting the reproductive success of the organism.
Sister-chromatid cohesion: A term used to describe the tight association of sister chromatids during mitosis and meiosis. At mitotic metaphase, sister-chromatid cohesion is largely limited to pericentromeric regions.

Corresponding author: Dawe, R.K. (kelly@plantbio.uga.edu).
Available online 30 October 2006.

 0968-0004/$ – see front matter doi:10.1016/j.tibs.2006.10.004

reflects multiple evolutionary origins from a canonical Histone 3. Evidently, these 'CENH3' proteins in different lineages have converged for centromeric function (Box 1). CENH3s are homologous to histone H3 within the histone core domain except for a small region that is required for targeting CENH3 [14,15]. Another region of particular interest is the N-terminal tail, which, at least on H3, provides chromatin with a distinct identity that influences gene expression. The N-terminal tail of CENH3 is strikingly different from that of other H3s and displays almost no homology across species. A tail is necessary for proper centromere function [16] but, as yet, there is no evidence that specific tail sequences are required (for instance, the N-terminal tail of CENP-A can be replaced with unrelated sequence [17]). Paradoxically, the CENH3 N-terminal tail shows evidence of accelerated evolution, as if it were under strong positive selection for an important function [18]. Similarly, CENP-C is evolving at remarkable rates and shows evidence of adaptive evolution across all lineages studied [19].

Box 1. Evolutionary convergence of histones

Phylogeny is a highly successful predictor of function. For example, kinesins comprise a family of microtubule-based motor proteins that have diverged to acquire numerous roles in cellular processes; by organizing kinesins according to subfamilies, the roles of otherwise unknown kinesin family members have been successfully predicted. When kinesins were renamed by the community to enforce common usage, it made sense to classify them by the subfamily, regardless of their various historical designations [81]. By contrast, the family of proteins related to histone H3 is unique in that function is not predicted by phylogeny because multiple lineages of histones have converged to the same function, rather than diverging from a single ancestor [18]. This discrepancy between phylogeny and function raises questions as to how these histones should be classified and named.

In most eukaryotic lineages, there are three ancestrally related forms of Histone 3 [82] with H3 being the usual designation for the major form, H3.3 for the replacement form and CENP-A for the centromeric form. Grouping all H3 histones together and all H3.3 histones together is convenient, but it violates phylogeny as there seems to be at least four independent origins of the protein pair (in animals and fungi, plants, ciliates and apicomplexans) [18]. Similarly, there is evidence that centromeric forms have evolved multiple times from an H3 or H3.3 ancestor [18]. For example, the original name for the *Arabidopsis* centromeric histone was HTR12 (for histone three number 12), assigned by the Plant Chromatin project, and HTR12 was retained when its centromeric localization was demonstrated [83]. It was clear, however, that a name that implies centromeric function is preferred, and CENH3 was suggested as a suitable 'umbrella' term [83]. Unlike CENP-A, the CENH3 term does not imply a family relationship. CENH3 retains the first syllable of CENP-A (for function and localization) and unites it with H3 (for phylogeny), thus it follows the precedence of using a prefix before the histone name to describe unusual histones, as was done for the histone H2A variant, macroH2A. CENH3 has since become the generally adopted term for plants [84], ciliates [85] and apicomplexans [86]; however, historical usage prevails for budding yeast, where the designation, Cse4 is usually retained in the literature [1].

Therefore, we aim to promote the adoption of CENH3 in the field while recognizing that, when searching the literature, researchers should still be able to find articles using the alternative and historical names for these proteins. To that end, the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) has implemented an automatic link between the terms CENP-A and CENH3 so that searching PubMed using either returns all references for both.

Why do eukaryotes tolerate (or favor) key kinetochore proteins that change so rapidly? As we argue here, the sequence variation in key kinetochore proteins is the outcome of a complex interplay between histone deposition, selfish DNA, and meiotic drive that enables the organism to maintain Mendelian segregation of the chromosomal DNA of the organism.

The unstable, transient nature of centromeric DNA

Simple repeats are the primary sequence of all centromeres in complex eukaryotes. The α satellite (a 171-base-pair tandemly repeated sequence) in humans is a well-studied example, but repeats of approximately the same size can be found throughout the animal and plant kingdoms. The basic structure of the repeat arrays is similar among species but the base-pair level sequences differ [20]. Thousands of simple repeats can evolve in unison at rates that exceed those observed in non-coding portions of genes [20,21]. This seems to be a dynamic process because the lengths of the repeat arrays on a single centromere can vary naturally by as much as tenfold [22,23]. The most striking examples come from the rice lineage, where entire centromere families have been erased and replaced in unexpectedly short time frames [24,25]. The primary mechanisms of change are presumed to be unequal recombination and gene conversion, which can cause the generation and spread of new repeats [26,27] in addition to the loss of repeat arrays [28]. Most centromere sequence data have been interpreted in the context of unequal recombination [29–32], although direct measurements of the rates and impact of recombination within centromeres are as yet unavailable.

Although megabase-sized satellite sequence arrays are characteristic features of higher eukaryotic centromeres, they are not necessarily required for centromere function. In human cells, new centromeres ('neocentromeres') can appear sporadically at seemingly random sites that lack satellite repeats [33], and, in other species, centromeres move laterally to new sites that have little or no sequence similarity [34,35]. Centromere plasticity is also revealed by overexpressing human CENH3 (CENP-A), which leads to expansion of centromeric chromatin [36]. Expansion can occur over both higher-order α-satellite repeats and active genes, indicating that there is no sequence or compositional specificity requirements for CENH3-containing (CEN) chromatin formation.

Spatial differentiation within CEN chromatin

The distinction between CEN chromatin that encompasses the kinetochore-forming region and the surrounding pericentromeric heterochromatin represents the first level of chromatin differentiation at the centromere [37] (Figure 1). At most centromeres, this chromatin distinction seems to correspond to a differentiation between young and old DNA sequences. The DNA sequences in CEN chromatin are young, consisting of nearly identical copies of satellite repeats that are constantly being homogenized by genetic exchange [30,38]. CENH3 is usually found within the most homogeneous satellite sequences [20], whereas the DNA in pericentromeres is much less uniform and contains ancient diverged satellite arrays that are peppered by transposons

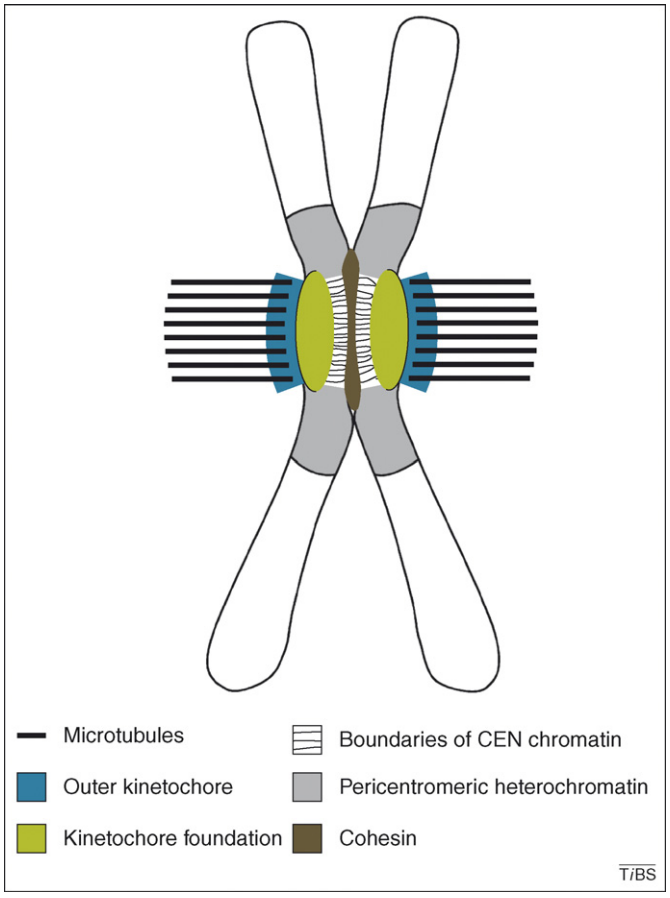

Figure 1. Metaphase kinetochores in their chromosomal context. The larger centromere–kinetochore domain includes pericentromeric heterochromatin and cohesin. The CEN chromatin, which is poorly understood, is composed of a mixture of histone H3 and CENH3 (and maybe histone H3.3). The relationship between histone H3 and cohesion in plants and animals is unclear. There is at least a spatial correlation between heterochromatin and cohesion, but it is not known whether cohesin interacts with CEN chromatin.

[30,39]. For example, all human centromeres are embedded within homogeneous 171-base-pair α-satellite repeats and *Arabidopsis* centromeres are similarly embedded within a different 178-base-pair satellite repetitive array [21,40]. Pericentromeric regions tend to be heterochromatic and enriched in cohesin, which holds chromatids together until anaphase [41] (Figure 1). The heterochromatic state might be required to recruit pericentromeric cohesin in some species [41,42], but active genes (presumably within euchromatin) can also exist within centromeric and pericentromeric domains [43,44].

Recent work has highlighted the fact that H3-containing nucleosomes are dispersed within the centromere core domains [45], although the modifications on these specialized histone H3s are not typical of heterochromatin [37,46]. CENH3 and H3 seem to be dispersed in long alternating arrays. A combination of fluorescent *in situ* hybridization and antibody staining suggests that the sizes of the individual CENH3- and H3-containing nucleosomal arrays are ~15–40 kb in *Drosophila* [45]. CENH3-containing arrays are thought to face outwards towards microtubules, whereas H3-containing arrays comprise the internal regions where sister-chromatid cohesion takes place [45,47] (Figure 1). Chromatin immunoprecipitation analyses of two centromeres (the 10q25 human neocentromere and rice centromere 8) have provided important confirmation of the idea of alternating CENH3 and H3 arrays [44,48]. In the case of the human 10q25 neocentromere, the alternating arrays are ~10–50 kilobases, matching prior estimates from cytological analysis [48].

Just how alternating histone patterns are maintained in centromeres is unknown, although it is interesting to consider the possibility that they are laid down during nucleosome assembly. Like centromeres, chromosome arms are differentiated by alternative H3 variants, which are deposited by distinct nucleosomes-assembly pathways [49,50]. Histone H3 is exclusively deposited at replication, whereas H3.3 is deposited at transcriptionally active genes throughout the cell cycle. H3 is frequently methylated at Lys9 and Lys27, whereas H3.3 is usually acetylated at lys9 and methylated at Lys4 [51–54]. It is not known whether the centromeric arrays of canonical nucleosomes contain H3 or H3.3, although the fact that *Schizosaccharomyces pombe* encodes only the H3.3 form [55] and their centromeres show enrichment of dimethyl-Lys4 [56] indicates that it could be the H3.3 replication-independent form that separates CENH3 arrays. If so, then centromeric chromatin would consist entirely of nucleosomes that are deposited independently of replication (containing CENH3 and H3.3), whereas the surrounding heterochromatin consists of nucleosomes that are deposited during replication (containing H3).

Dynamics of CEN chromatin

Biochemical purification of the soluble CENH3 nucleosome pre-assembly complex from *Drosophila* has yielded a simple trimolecular complex that consists of CID (the *Drosophila* CENH3), histone H4 and RbAp48 (a histone chaperone protein found also in the H3 and H3.3 assembly complexes) [57]. The lack of any centromere-specific protein in the soluble complex that deposits CENH3, except for CENH3 itself, suggests either or both of two possible explanations: (i) CENH3 could be deposited wherever there is pre-existing CENH3 on chromatin, and/or (ii) CENH3 could deposit promiscuously but can only gain access for assembly onto DNA at centromeres. There is evidence that is consistent with both of these possibilities. On the one hand, biochemical fractionation of human centromeric chromatin (as opposed to the soluble pre-assembly complex) using either CENH3 (CENP-A) or CENP-H reveals complex sets of centromere-specific proteins [11,58], which could create a structure that would enable recognition by the soluble CENP-A-containing assembly complex. On the other hand, overexpression of CENP-A or CID causes promiscuous assembly throughout euchromatin [49,59,60], which might suggest a non-sequence-specific gap-filling process. A promiscuous gap-filling process could also explain the *de novo* formation of artificial chromosomes [61] and neocentromeres [33–35]. It seems likely that both mechanisms are used to assure the faithful assembly of CENH3-containing nucleosomes [60].

It is now realized that chromatin is far more dynamic than once thought. Heterochromatin-associated proteins are in constant flux [62,63] and heterochromatin formation, at least in *S. pombe*, requires cycles of transcription

followed by small interfering RNA (siRNA)-directed silencing [64]. Similarly, maize centromeric repeats are transcribed and the resulting centromeric transcripts are bound to CENH3 [65]. When CENH3 is overexpressed, centromeric chromatin can expand and invade surrounding heterochromatin [36]; and, vice versa, when CENH3 is underexpressed, heterochromatin invades centromeres [45]. Expansion and contraction of epigenetic states might also occur in normal centromeric chromatin because both CENH3 and H3 can be detected on the same single-copy DNA sequences in rice centromere 8 [66]. These data indicate that the deposition machineries for CENH3-containing and canonical nucleosomes compete with each other during assembly. A built-in competition between CENH3 and H3 (or H3.3) would cause regular jostling of the two forms of nucleosomes during deposition that would tend to break up large blocks of CENH3 (Figure 2). The requirement for kinetochore attachment on one face of the centromere and cohesion on the other [41,67–70] should lead to a natural parity between CENH3- and H3-containing arrays and strong selection against either assembly process dominating.

Evolution of centromere repeats through epigenetic selection

Given that tandem-repeat arrays are common at nearly all centromeres, it seems likely that their size, sequence or arrangement contributes in some manner to centromere function. In human cells, arrays of α satellites are sufficient to organize kinetochores within artificial chromosomes [61,71]. The α satellite contains a key sequence known as the CENP-B box, which binds in a sequence-specific manner to the CENP-B protein and facilitates kinetochore formation [72]. However, CENP-B (and its DNA-binding motif) is absent at many human, plant and numerous animal centromeres [73,74]. On CENP-B-free human centromeres, and in species such as flies, nematodes and flowering plants, the available data are consistent with an epigenetic mode of centromere specification. What drives the evolution of centromere-specific repeats when epigenetics is the determining factor?

Most centromere repeats are approximately the length of DNA that wraps around a single nucleosome (150–180 base pairs), suggesting that satellites might have evolved to facilitate regular nucleosome packaging. Another possibility is that nucleosomes limit unequal recombination to small regions of the packaged chromatin [64,75] such that the repeat unit is approximately equivalent to a recombination unit. However, although such models help to explain the size of the repeats, they do not offer an explanation for the uniformity of the sequences, that is, the fact that the same repeats are usually found among all centromeres within a species (stochastic models such as those proposed by Smith [26] only explain the spread of repeats within centromeres).

The model we propose is that centromeric DNA behaves selfishly and adapts to the local chromatin environment. Species-specific differences in kinetochore-foundation proteins would be expected to present unique DNA-binding faces and the opportunity for evolving to better fit those faces. Furthermore, there should be strong competition among repeats (or arrays of repeats) to adapt more efficiently because the repeats that bind tightest are most likely to be connected to the spindle apparatus. Because all kinetochores within a species will have the same protein features, such a competition mechanism would be expected to drive centromeric repeat arrays to uniformity. In this model, the CENH3-centered epigenetic state is the first stage in centromere evolution and the formation and spread of centromere-repeat arrays is a consequence [25,66].

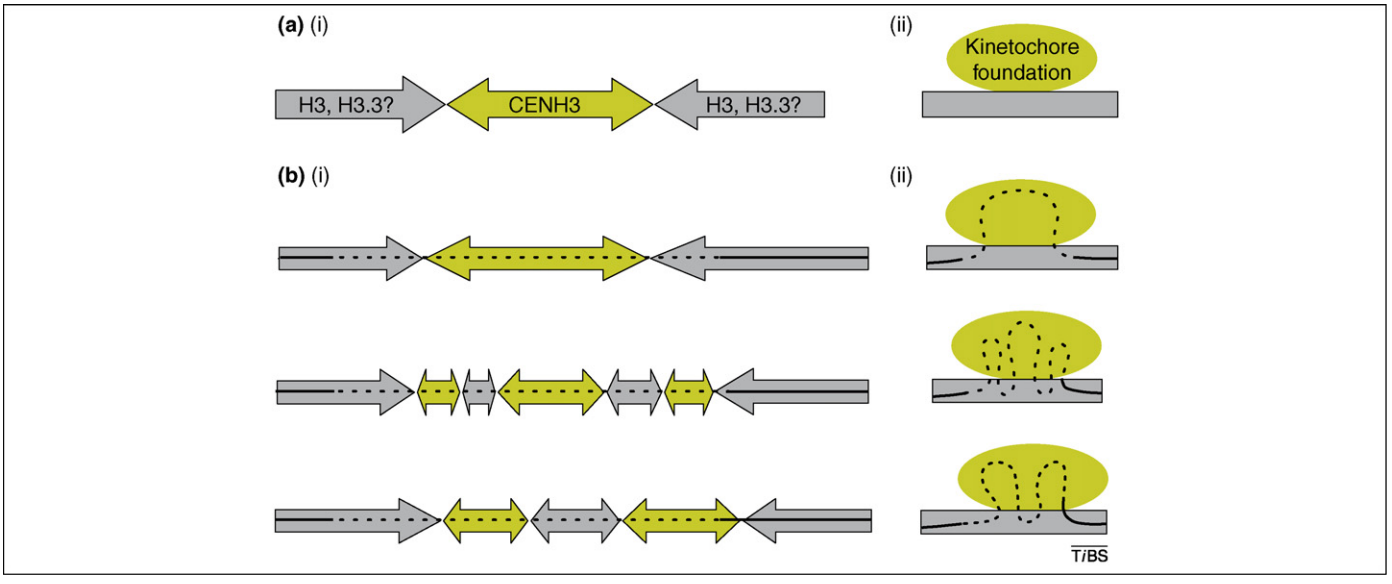

Figure 2. An epigenetic-competition model for centromere organization. **(a)** (i) We envision a situation whereby CENH3- and H3-containing nucleosomes compete or 'jostle' for placement over large expanses of similar centromere repeats. Arrowheads indicate that the two forms of chromatin often invade each other's territory. (ii) During metaphase, the CENH3-containing portion interacts with kinetochore-foundation proteins (right). **(b)** Three arrangements of chromatin over the same DNA sequence (i), together with representations of how the chromatin might fold into metaphase kinetochores in each case (ii). Dots represent satellite repeat arrays and solid lines represent the older, more heterogeneous sequences typical of pericentromeric regions.

Meiotic drive promotes the epigenetic mode of centromere inheritance

Why did this odd epigenetic mechanism evolve in complex eukaryotes but not in small genome species such as *S. cerevisiae* and *S. pombe*? We suspect that the answer lies in the mechanics of gametogenesis. In fungi, all four products of meiosis are functional, whereas in animals and plants only one product of female meiosis survives to make an egg. Meiotic inequality presents the opportunity for the centromeres to be 'abused' by meiotic drive [20,76,77]. In principle, increases in the size of centromere-repeat arrays could lead to increases in kinetochore size, increased capacity of the kinetochores to interact with microtubules and an increased likelihood that a chromosome is segregated to the egg (Figure 3a). Such deviation from Mendelian inheritance limits the capacity of a species to adapt to different environments.

Under conditions that enable centromere-mediated meiotic drive (Figure 3), the organism might find itself with a bad (unfit) chromosome that it cannot get rid of (segregate away). Mutations in key foundation proteins such as CENH3 or CENP-C could relieve the imbalance between strong and weak centromeres. The example illustrated in Figure 3 shows a mutated CENP-C that binds to existing centromere sequence in addition to a new (neocentromeric) sequence near the normal centromere (Figure 3b). The neocentromere sequence might be a slightly different variant of the native centromeric array, a dissimilar array that has transposed from another chromosome [78], or even a completely novel satellite repeat, such as is found at *Drosophila* centromeres [79]. The combined pull of the normal centromere and the nearby neocentromere would restore Mendelian segregation and provide a fitness advantage for both the neocentromere-containing chromosome and the mutant foundation protein. Both the new centromere and mutant kinetochore protein would then rapidly drive to fixation.

The mutation that best fits this scenario is one that reduces sequence specificity of a kinetochore-foundation protein, so that the protein then binds to both the normal centromere and the neocentromere. Any reduction in sequence specificity will effectively loosen the influence of DNA on kinetochore size and favor the epigenetic mode that promotes uniform kinetochores (Figure 2). CENP-C is a prime candidate for mediating this process because it is a DNA-binding protein [80] that shows evidence of recurrent positive selection over most of its length in both plants and animals [19]. In organisms in which all four products of meiosis are functional, genetically specified centromeres have evolved and the associated foundation proteins show strong purifying selection. However, where centromere drive poses the threat of extinction, any tendency towards genetic specification would be counteracted by the need to

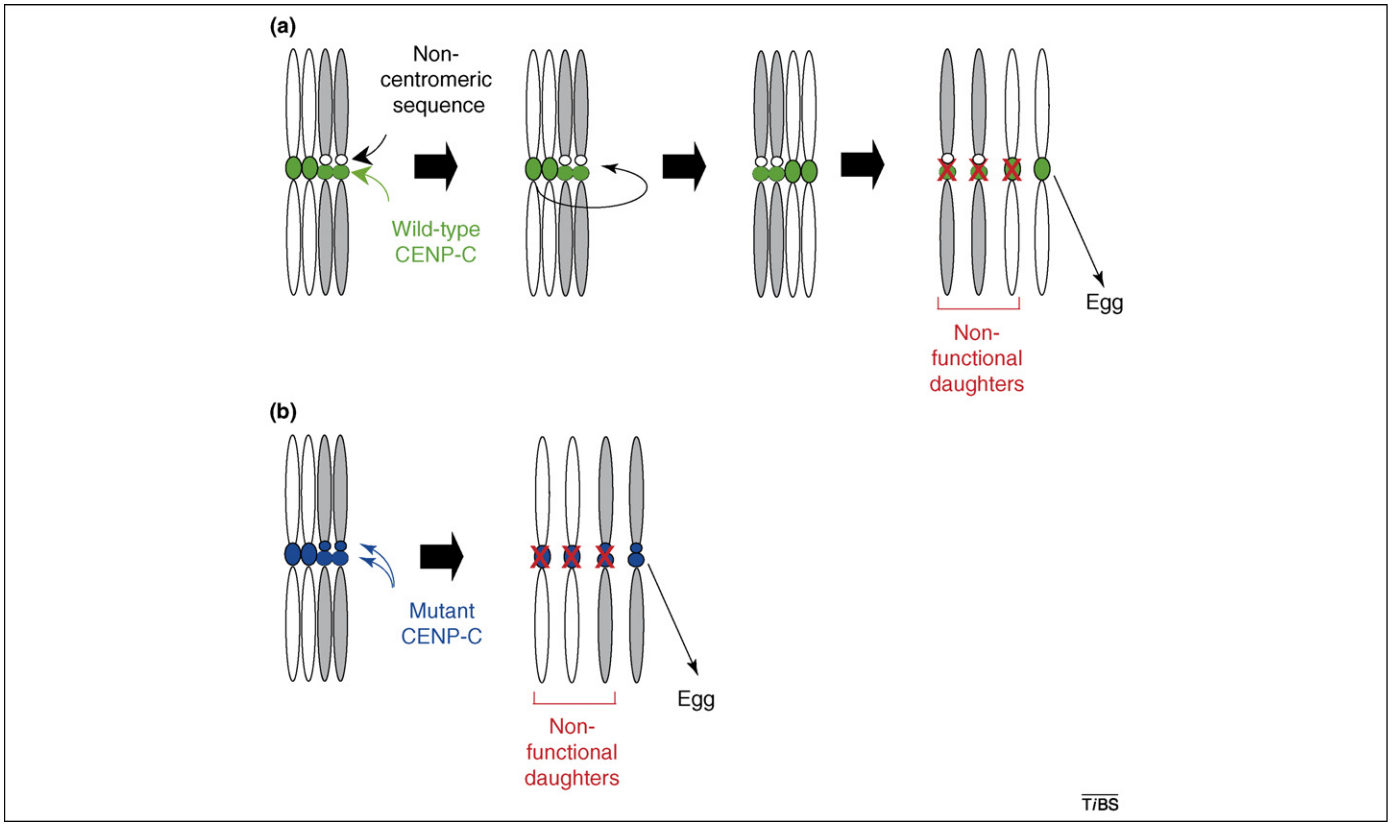

Figure 3. Meiotic drive favors epigenetic centromeres. A centromere that expands is proposed to gain an advantage at female meiosis by achieving an orientation towards the egg pole [20]. **(a)** The expanded centromere increases its representation in the population by accumulating more foundation proteins (represented by CENP-C; green) and making more microtubule connections, causing it to orientate towards the egg pole (curved arrow) even when it starts out in an unfavorable position. However, the resulting departure from Mendelian segregation would be expected to limit the capacity of the species to adapt to new environments. **(b)** Normal segregation can be restored if the driving centromere encounters a mutant CENP-C (blue) with reduced DNA-binding specificity that binds to previously non-centromeric sequence. This results in meiotic parity, providing an advantage for the gene encoding the mutant foundation protein.

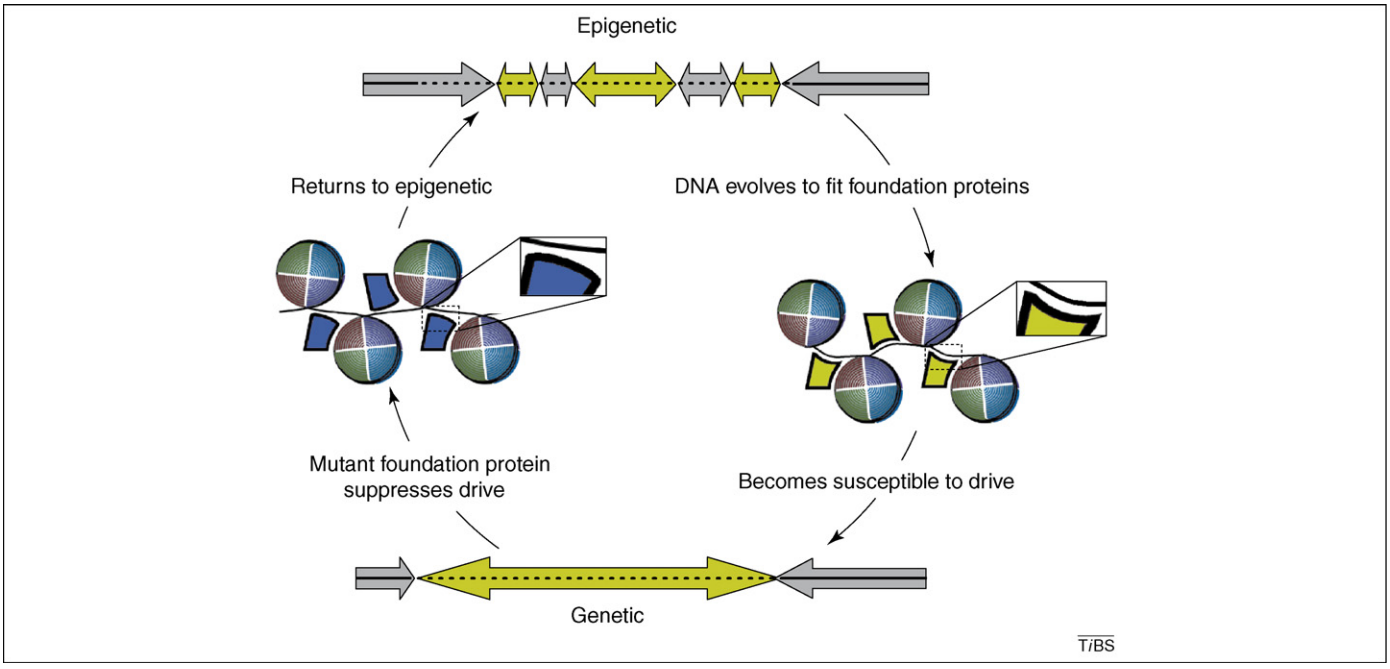

Figure 4. Epigenetics is in the driver's seat of a cyclical process. The fact that centromeres are directly attached to spindles promotes the evolution of selfish repeats that bind tightly to foundation proteins. The bent, tightly bound region of DNA is shown interacting with CENP-C as an example (green). In principle, any of the foundation proteins could be involved in this process. Such sequence-dependent (genetic) interactions between centromeric DNA and the kinetochore are susceptible to meiotic drive. A mutation that changes the structure of CENP-C (blue) would disrupt meiotic drive and return the centromeres to a more stable epigenetic mode of centromere specification. Adaptive mutations are presumed to change the shape of foundation proteins such that the selfish DNA no longer binds with high affinity. The process should be cyclical, driving the rapid evolution of both centromeric repeats and kinetochore-foundation proteins.

relieve drive, thus putting epigenetic centromeres back in the driver's seat.

The process should be cyclical (Figure 4). The most stable state is presumably epigenetic, but the opportunity always exists for selfish repeats to evolve within the centromeric chromatin environment. Competition among selfish DNAs will push the system towards genetic specification and again make centromeres susceptible to meiotic drive. Recurrent cycles of drive and suppression should cause the rapid divergence of both centromeric sequence arrays and foundation proteins. Such an evolutionary interplay between epigenetic and genetic states can explain most of the available data on centromeres. However, many of the underlying assumptions have yet to be rigorously tested. Future work will be focused on showing that chromosome transmission can be influenced by inequities in kinetochore size, that gene conversion and/or unequal recombination occur at high frequencies within centromeric repeat arrays, and that some kinetochore-foundation proteins (in addition to CENP-B) have sequence-specific binding affinities for centromeric repeats.

Acknowledgements

We thank David Landsman and Kathi Canese at the NCBI for their assistance in getting modifications described in Box 1 made to the PubMed query engine.

References

1 Collins, K.A. *et al.* (2005) *De novo* kinetochore assembly requires the centromeric histone H3 variant. *Mol. Biol. Cell* 16, 5649–5660
2 Palmer, D.K. and Margolis, R.L. (1985) Kinetochore components recognized by human autoantibodies are present on mononucleosomes. *Mol. Cell. Biol.* 5, 173–186
3 Guldner, H.H. *et al.* (1984) Human anti-centromere sera recognise a 19.5 kD non-histone chromosomal protein from HeLa cells. *Clin. Exp. Immunol.* 58, 13–20
4 Earnshaw, W.C. and Rothfield, N. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with schleroderma. *Chromosoma* 91, 313–321
5 Palmer, D.K. *et al.* (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* 104, 805–815
6 Palmer, D.K. *et al.* (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3734–3738
7 De Wulf, P. *et al.* (2003) Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. *Genes Dev.* 17, 2902–2921
8 Goshima, G. *et al.* (2003) Human centromere chromatin protein hMis12, essential for equal segregation, is independent of CENP-A loading pathway. *J. Cell Biol.* 160, 25–39
9 Amor, D.J. *et al.* (2004) Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol.* 14, 359–368
10 Kline, S.L. *et al.* (2006) The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. *J. Cell Biol.* 173, 9–17
11 Okada, M. *et al.* (2006) The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat. Cell Biol.* 8, 446–457
12 Emanuele, M.J. *et al.* (2005) Measuring the stoichiometry and physical interactions between components elucidates the architecture of the vertebrate kinetochore. *Mol. Biol. Cell* 16, 4882–4892
13 Meraldi, P. *et al.* (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol.* 7, R23
14 Vermaak, D. *et al.* (2003) Maintenance of chromatin states: an open-and-shut case. *Curr. Opin. Cell Biol.* 15, 266–274
15 Black, B.E. *et al.* (2004) Structural determinants for generating centromeric chromatin. *Nature* 430, 578–582
16 Van Hooser, A.A. *et al.* (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *J. Cell Sci.* 114, 3529–3542

17 Wieland, G. *et al.* (2004) Functional complementation of human centromere protein A (CENP-A) by Cse4p from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 24, 6620–6630
18 Malik, H.S. and Henikoff, S. (2003) Phylogenomics of the nucleosome. *Nat. Struct. Biol.* 10, 882–891
19 Talbert, P.B. *et al.* (2004) Adaptive evolution of centromere proteins in plants and animals. *J. Biol.* 3, 18
20 Henikoff, S. *et al.* (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293, 1098–1102
21 Hall, S.E. *et al.* (2003) Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* 13, 195–205
22 Mahtani, M.M. and Willard, H.F. (1990) Pulsed-field gel analysis of alpha-satellite DNA at the human X chromosome centromere: high-frequency polymorphisms and array size estimate. *Genomics* 7, 607–613
23 Kato, A. *et al.* (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13554–13559
24 Lee, H.R. *et al.* (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11793–11798
25 Dawe, R.K. (2005) Centromere renewal and replacement in the plant kingdom. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11573–11574
26 Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191, 528–535
27 Schindelhauer, D. and Schwarz, T. (2002) Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous α-satellite DNA array. *Genome Res.* 12, 1815–1826
28 Walsh, J. (1987) Peristence of tandem arrays: implications for satellite and simple-sequence DNAs. *Genetics* 115, 553–567
29 Warburton, P.E. and Willard, H.F. (1992) PCR amplification of tandemly repeated DNA: analysis of intra- and interchromosomal sequence variation and homologous unequal crossing-over in human α satellite DNA. *Nucleic Acids Res.* 20, 6033–6042
30 Schueler, M.G. *et al.* (2001) Genomic and genetic definition of a functional human centromere. *Science* 294, 109–115
31 Hall, S.E. *et al.* (2005) Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. *Genetics* 170, 1913–1927
32 Ma, J. and Bennetzen, J.L. (2006) Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. *Proc. Natl. Acad. Sci. U. S. A.* 103, 383–388
33 Warburton, P.E. (2004) Chromosomal dynamics of human neocentromere formation. *Chromosome Res.* 12, 617–626
34 Maggert, K.A. and Karpen, G.H. (2001) The activation of a neocentromere in *Drosophila* requires proximity to an endogenous centromere. *Genetics* 158, 1615–1628
35 Nasuda, S. *et al.* (2005) Stable barley chromosomes without centromeric repeats. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9842–9847
36 Lam, A.L. *et al.* (2006) Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4186–4191
37 Sullivan, B.A. and Karpen, G.H. (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.* 11, 1076–1083
38 Roizes, G. (2006) Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. *Nucleic Acids Res.* 34, 1912–1924
39 Copenhaver, G.P. *et al.* (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286, 2468–2474
40 Hosouchi, T. *et al.* (2002) Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. *DNA Res.* 9, 117–121
41 Bernard, P. *et al.* (2001) Requirement of heterochromatin for cohesion at centromeres. *Science* 294, 2539–2542
42 Nakashima, H. *et al.* (2005) Assembly of additional heterochromatin distinct from centromere-kinetochore chromatin is required for *de novo* formation of human artificial chromosome. *J. Cell Sci.* 118, 5885–5898
43 Saffery, R. *et al.* (2003) Transcription within a functional human centromere. *Mol. Cell* 12, 509–516
44 Yan, H. *et al.* (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. *Plant Cell* 17, 3227–3238
45 Blower, M. *et al.* (2002) Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* 2, 319–330
46 Shi, J. and Dawe, R.K. (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics* 173, 1571–1583
47 Zinkowski, R. *et al.* (1991) The centromere-kinetochore complex: a repeat subunit model. *J. Cell Biol.* 113, 1091–1110
48 Chueh, A.C. *et al.* (2005) Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Hum. Mol. Genet.* 14, 85–93
49 Ahmad, K. and Henikoff, S. (2002) Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6477–6484
50 Tagami, H. *et al.* (2004) Histone H3. 1 and H3. 3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116, 51–61
51 Waterborg, J.H. (1990) Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa. *J. Biol. Chem.* 265, 17157–17161
52 Johnson, L. *et al.* (2004) Mass spectrometry analysis of *Arabidopsis* histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res.* 32, 6511–6518
53 McKittrick, E. *et al.* (2004) Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1525–1530
54 Hake, S.B. *et al.* (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J. Biol. Chem.* 281, 559–568
55 Choi, E.S. *et al.* (2005) Dynamic regulation of replication independent deposition of histone H3 in fission yeast. *Nucleic Acids Res.* 33, 7102–7110
56 Cam, H.P. *et al.* (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.* 37, 809–819
57 Furuyama, T. *et al.* (2006) Chaperone-mediated assembly of centromeric chromatin *in vitro*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6172–6177
58 Foltz, D.R. *et al.* (2006) The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* 8, 458–469
59 Shelby, R.D. *et al.* (1997) Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. *J. Cell Biol.* 136, 501–513
60 Heun, P. *et al.* (2006) Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev. Cell* 10, 303–315
61 Harrington, J.J. *et al.* (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nat. Genet.* 15, 345–355
62 Cheutin, T. *et al.* (2003) Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* 299, 721–725
63 Festenstein, R. *et al.* (2003) Modulation of heterochromatin protein 1 dynamics in primary mammalian cells. *Science* 299, 719–721
64 Verdel, A. *et al.* (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303, 672–676
65 Topp, C.N. *et al.* (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15986–15991
66 Nagaki, K. *et al.* (2004) Sequencing of a rice centromere reveals active genes. *Nat. Genet.* 36, 138–145
67 Wines, D.R. and Henikoff, S. (1992) Somatic instability of a *Drosophila* chromosome. *Genetics* 131, 683–691
68 Hoque, M.T. and Ishikawa, F. (2002) Cohesin defects lead to premature sister chromatid separation, kinetochore dysfunction, and spindle-assembly checkpoint activation. *J. Biol. Chem.* 277, 42306–42314
69 Shibata, F. and Murata, M. (2004) Differential localization of the centromere-specific proteins in the major centromeric satellite of *Arabidopsis thaliana*. *J. Cell Sci.* 117, 2963–2970

70 Zhang, X. *et al.* (2005) Phosphoserines on maize centromeric histone H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. *Plant Cell* 17, 572–583
71 Larin, Z. and Mejia, J.E. (2002) Advances in human artificial chromosome technology. *Trends Genet.* 18, 313–319
72 Masumoto, H. *et al.* (2004) The role of CENP-B and α-satellite DNA: *de novo* assembly and epigenetic maintenance of human centromeres. *Chromosome Res.* 12, 543–556
73 Earnshaw, W.C. *et al.* (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* 98, 1–12
74 Saffery, R. *et al.* (2000) Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. *Hum. Mol. Genet.* 9, 175–185
75 Baumann, M. *et al.* (2003) Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. *EMBO J.* 22, 5197–5207
76 Dawe, R.K. and Hiatt, E.N. (2004) Plant neocentromeres: fast, focused, and driven. *Chromosome Res.* 12, 655–669
77 Burt, A. and Strivers, R. (2006) Female drive. In *Genes in Conflict: The Biology of Selfish Genetic Elements*, Harvard University Press, pp. 301–321
78 Puechberty, J. *et al.* (1999) Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen. *Genomics* 56, 274–287
79 Lohe, A.R. *et al.* (1993) Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* 134, 1149–1174
80 Sugimoto, K. *et al.* (1994) Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. *J. Biochem. (Tokyo)* 116, 877–881
81 Lawrence, C.J. *et al.* (2004) A standardized kinesin nomenclature. *J. Cell Biol.* 167, 19–22
82 Luger, K. *et al.* (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260
83 Talbert, P.B. *et al.* (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14, 1053–1066
84 Houben, A. and Schubert, I. (2003) DNA and proteins of plant centromeres. *Curr. Opin. Plant Biol.* 6, 554–560
85 Cervantes, M.D. *et al.* (2006) The CNA1 histone of the ciliate *Tetrahymena thermophila* is essential for chromosome segregation in the germline micronucleus. *Mol. Biol. Cell* 17, 485–497
86 Miao, J. *et al.* (2006) The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene* 369, 53–65