Genetic conflicts during meiosis and the evolutionary origins of centromere complexity

H.S. Malik*[1] and J.J. Bayes*†

*Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, A1-162, Seattle, WA 98109, U.S.A., and †Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, U.S.A.

Abstract

Centromeric DNA evolves rapidly, ranging in size and complexity over several orders of magnitude. Traditional attempts at studying centromeres have left unexplained the causes underlying this complexity and rapid evolution. Instead of directly studying centromeric DNA sequence, our approach has been to study the proteins that epigenetically determine centromere identity. We have discovered that centromeric histones (CenH3s) have evolved under positive selection in multiple lineages, suggesting an involvement in recurrent genetic conflict. Our hypothesis is that 'centromere-drive' is the source of this conflict. Under this model, centromeres compete via microtubule attachments for preferential transmission in female meioses occurring in animals and plants. Since only one of four meiotic products will become the egg, this competition confers a selfish advantage to chromosomes that can make more microtubule attachments, resulting in runaway expansions of centromeric satellites. While beneficial to the 'driving' chromosome, these expansions can have deleterious effects on the fitness of an organism and of the species. CenH3s as well as other heterochromatin proteins have evolved under positive selection to suppress the deleterious consequences of 'centromere-drive' by restoring meiotic parity.

Unexplained centromere sequence complexity

Centromeres are typically visualized as the primary constriction point of chromosomes, laying the foundation for the kinetochore complex and the recruitment of microtubules. Thus centromeres provide an absolutely fundamental function for the faithful segregation of chromosomes at each cell division. However, despite this essential role, centromeres range in size and complexity from the 125 bp point centromeres in *Saccharomyces cerevisiae* [1] to the hundreds of kilobases of satellite repeat arrays that constitute the complex centromeres of plants and animals [2,3]. The centromeres of holokinetic organisms such as *Caenorhabditis elegans* are even more complex; they comprise centromeric determinants dispersed throughout the length of the chromosome that coalesce at metaphase. In such instances, the centromere runs the entire length of the chromosome [4]. Generally speaking, satellite repeat arrays appear to be important for the function of complex centromeres. However, this simple relationship is challenged by human neocentromeres, which appear to lack any tandemly repetitive sequence whatsoever [5]. Furthermore, in *Drosophila*, centromeric satellites can be found in distal blocks from the centromeres, some of which have weak centromeric activity [6] and others not. In the best-studied *Drosophila* centromere, centromeric and heterochromatic sequences are almost indistinguishable [7]. Adding to the complexity of centromeric regions within a species is the finding that satellite DNA sequences can change quite rapidly between closely related species. For instance, there is very little overlap between the centromeric satellite sequences of *Drosophila melanogaster* and *D. simulans*, in spite of the fact that many satellites are shared between the two species [8]. Thus their locations in the genome have changed dramatically in the 2.5 million year divergence between these two *Drosophila* species. Similarly, the human X centromeric satellite appears to be only as old as the great apes [2]. In several instances, homologous chromosomes in closely related primate species bear different, non-orthologous α-satellite sequence variants [9,10]. Thus centromeric regions evolve rapidly both within and between species.

Painstaking sequencing and assembly efforts have made some progress in describing centromeric DNA complexity in diverse organisms. The 420-kb-long Dp1187 minichromosome in *D. melanogaster* [7], the 750 kb centromere on rice chromosome 8 [11] and the human X centromere [2] are examples of assembly efforts that have led to a detailed picture of the heterochromatin–centromere boundary in complex centromeres. The assembly of the human X centromere indicated a highly homogeneous region of α-satellite repeats at the 'core' of centromeres flanked by satellite repeats with a gradient of heterogeneity (accumulated mutations) and transposon insertions away from the centromere. A recent retrotransposon insertion in the flanking region allowed investigators to conclude that the extant X centromere α-satellite was young and probably arose only in the great

apes [2]. These findings support the simple mutation–recombination balance model where recombination (either unequal crossing over or gene conversion) is the underlying force that homogenizes centromeric repeats in the middle of an array, balanced by mutation and transposition in the flanks [12,13].

Studies on centromeric DNA paint a highly dynamic picture of centromere evolution, but they do not provide a selective rationale for this rapid evolution and large-scale accumulation of satellite repeats. Indeed, several theoretical studies have pointed out the inadequacy of mutation and recombination alone to explain increased array sizes, suggesting that selection must play a role in their evolution [14–17]. Moreover, it has been demonstrated that pericentric satellites can contribute to a fitness difference between *D. melanogaster* strains [18]. Another line of evidence for the selective consequences of pericentric satellites comes from studies that implicate a newly arisen satellite repeat in the hybrid inviability seen in *D. simulans/D. melanogaster* interspecific hybrids [19,20]. One form of selective constraint that acts on centromeric satellites could be simply purifying selection to maintain an uninterrupted, homogeneous array of a minimum size to form a functional centromere. Another selective force may be the transmission advantages of larger centromeres in female meiosis, which we suggest may play a more profound role in dictating array length of centromeric satellites [13,21].

Asymmetric female meioses provides an opportunity for 'centromere-drive'

In most plants and animals, the process of female meiosis is asymmetric. Out of four haploid products, only one will have a chance at evolutionary success as it will be chosen to become the egg, while the other three products degenerate into evolutionary dead-ends. Why such asymmetry evolved is itself an intriguing evolutionary question that has led to some debate. Nevertheless, it is clear that this asymmetric nature of female meiosis can lead to genetic elements subverting this process for their own advantage. The knob elements from maize provide one such example [22]. Knobs are blocks of heterochromatic satellite DNA that are always found distally from the centromere. If a pair of chromosomes is heterozygous, i.e. only one contains a knob, then crossing over can occur between the knob and centromere during female meiosis. Under the appropriate genetic background, knobs bind microtubules and knob-bearing chromatids are pulled towards the outermost megaspores during meiosis II. One of these outermost megaspores will become the gametophyte and produce gametes [23]. By virtue of this favourable orientation, instead of a 50% expected ratio of transmission in a heterozygote, knob transmission in female meiosis varies from 59 to 82% correlated with the size of the satellite array [24]. Thus the 'selfish' knobs exploit the inherently non-Mendelian nature of female meiosis for their survival.

A transmission advantage in female meiosis may also account for high rates of non-disjunction in *Drosophila* females [25]. A sensitized assay found a large range of non-disjunction frequencies among X chromosomes. This variation in non-disjunction correlated significantly with two variants of the *nod* chromokinesin, which were found to be present at intermediate frequencies in natural populations. The *nod* chromokinesin is required for proper achiasmate segregation [26–28], yet apparently deleterious alleles have thrived in *Drosophila* populations. These findings led to the ootid-competition model. This model posits that polymorphic alleles of loci involved in segregation of ootids during female meiosis were likely to provide multiple opportunities for competitive interactions among ootids, since only one ootid is included in the pronucleus [25]. Thus female meiotic drive could result in the sponsoring of otherwise defective alleles, as a balance is struck between the competitive advantages conferred on to this allele in female meiosis and the cost in causing high rates of non-disjunction. This model also predicted that centromeres and other chromosomal elements could compete directly in this manner. Specifically, centromeres would competitively orient towards the preferred pole during meiosis I, whereas telomeres and other distal elements would do so later in female meiosis (such as the knob elements in maize). This model serves as the basis of the 'centromere-drive' model that we have proposed to explain the evolution of centromeres [29,30].

Under the 'centromere-drive' model (Figure 1), the first stage is a satellite expansion that leads to a centromere with enhanced microtubule binding abilities, which can result in its transmission advantage in female meiosis. An extreme case of transmission distortion that is likely to be the result of centromere-drive was recently elegantly demonstrated in *Mimulus* (monkeyflower) species [31]. In crosses of F1 hybrids of *Mimulus nasutus* and *M. guttatus*, there was such a strong transmission bias against the *M. nasutus* allele that instead of the Mendelian expectation of 1:2:1 (MN/MN, MN/MG, MG/MG, where MN is *M. nasutus* and MG is *M. guttatus*), a ratio of 0:2:2 was observed. Using a series of experiments designed to rule out alternate hypotheses, a strong case for centromere-mediated transmission distortion of 100% could be made [31]. Indeed, such a strong transmission distortion implies that it must happen early during female meiosis, at meiosis I [25], leaving the centromere as the only possible suspect [32]. While the satellite DNA configurations of the MN and MG centromeres have yet to be discovered, this study highlights what a potent force 'centromere-drive' can be in natural populations.

Arresting 'centromere-drive' by restoring meiotic parity

A number of negative effects can be associated with a sweep of a 'selfish centromere', including the fixation of linked deleterious mutations. The effects of a driving centromere

Figure 1 | Centromere-drive and its suppression

We highlight the two steps of the centromere-drive model using the X–Y chromosomes as an example. In the first stage, a satellite expansion leads to a centromere with enhanced microtubule binding abilities, which can result in a transmission advantage in female meiosis. This can lead to deleterious effects, including enhanced non-disjunction between the X–Y chromosomes in male meiosis (as diagrammed here). In the second stage, a suppressor allele in CenH3 (or any other satellite-binding protein) that can restore meiotic parity will be selectively favoured because it alleviates the deleterious effects of centromere-drive. This can be done in two ways: either (i) by expanding CenH3 binding and increasing microtubule attachments on the Y centromere (as shown) or (ii) by restricting CenH3 binding and reducing microtubule binding by the driving X centromere expansion (not shown). Repeated episodes of centromere-drive followed by the fixation of suppressing CenH3 alleles will lead to rapid expansions of centromeric satellites and the rapid fixation of non-synonymous nucleotide substitutions in genes encoding CenH3s (referred to as positive selection). While the X–Y pair of chromosomes is shown here for illustrative purposes, similar arguments would apply to both Z–W sex chromosomes as well as autosomes.
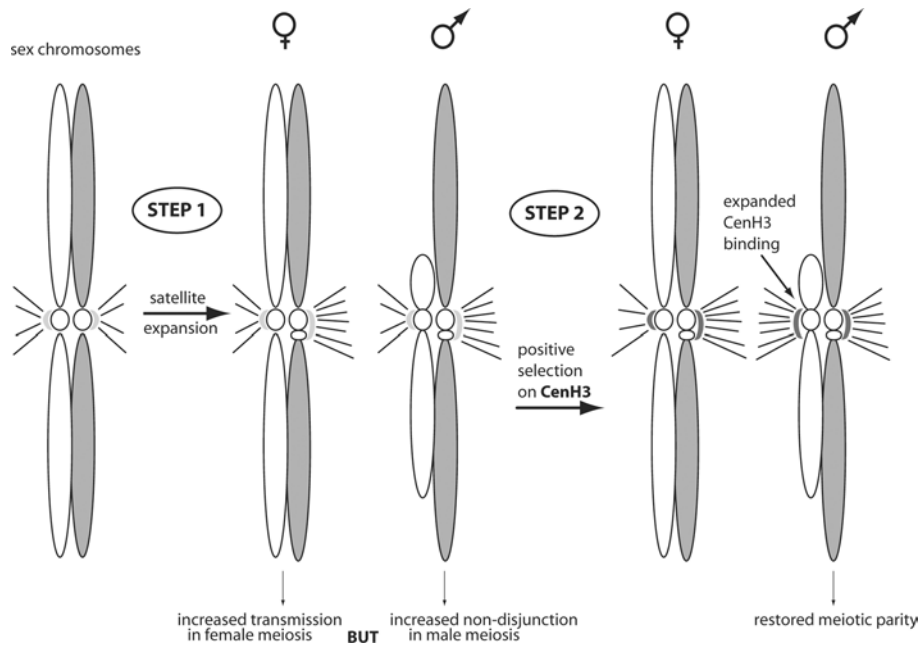
would be even more pronounced in the case of the sex chromosomes. For instance, in the case of ZW heterogametic systems (birds and lepidopterans), competition between the sex chromosomes for inclusion into the egg would lead to skewed sex ratios and threaten the population. In the case of the XY males (mammals and flies), competition between the X chromosomes would lead to 'stronger X centromeres' emerging via selective advantage. However, in XY male meiosis, which relies on symmetry, this would lead to greater non-disjunction, and in extreme instances, sterility (due to recurrent meiotic checkpoint-induced apoptosis) [33,34]. For example, Robertsonian fusions that result from the fusion of two acrocentric chromosomes have a differential advantage through female but not male meiosis in many vertebrates, including humans. In humans, the Robertsonian fusions are preferentially transmitted through female meiosis [35,36], which may partly explain why a significant proportion (0.12%) of the human population are carriers of a Robertsonian translocation [37]. There are no reports of any somatic (mitotic) effects but three-quarters

of male carriers of Robertsonian fusions appear to suffer deleterious fertility consequences [38]. This sterility probably results from a male meiotic checkpoint that monitors tension of microtubule attachment in mice [33] and may occur in *Drosophila* as well [34]. Thus female meiotic success can be balanced by the high cost to male fertility.

In situations where meiotic drivers have thrived in a population but cannot drive to fixation, theory predicts that suppressor alleles will arise to alleviate the effects of the drive or to eliminate the drive itself [39]. Furthermore, these suppressor alleles would have to be unlinked from the drive locus so as to not reap the 'benefits' of the drive [40]. Therefore we propose that in the second stage of our 'centromere-drive' model (Figure 1), a suppressor allele in CenH3 or any other satellite-binding protein can restore meiotic parity. This can be done in two ways: either (i) by increasing microtubule binding by other centromeres (as shown in Figure 1) or (ii) by reducing microtubule binding by the driving centromere expansion (not shown). Such suppressor alleles will be selectively favoured because of their

ability to alleviate the deleterious effects of centromere-drive. Thus genetic conflict between centromeres and suppression of this competition can drive centromeres to become larger and CenH3s to be under positive selection. Success of the suppressor alleles can lead to the degeneration of the drive system when the transmission advantage is no longer present. Subsequently, the suppressor will degenerate, leading to the presence of cryptic drive-suppressor systems [41]. This model can explain why extant centromeric satellites in complex centromeres tend to be flanked by closely related pericentric satellites, which are probably evolutionary remnants of previous centromere expansions that have lost their transmission advantage, and are therefore evolving neutrally. Typically meiotic drivers and their suppressors are neomorphs [42] and neither is essential for an organism. In the unusual scenario when essential elements (centromeres and CenH3s) act as drivers or suppressors, we could only uncover a cryptic genetic conflict by observing episodes of positive selection within them [21].

Indeed, positive selection (an excess of replacement over synonymous changes) has been documented for the centromeric histone genes in both *Drosophila* and *Arabidopsis* species [29,43], but not in mammals [44]. In addition, another ubiquitously conserved centromeric protein, CENP-C, has been shown to have evolved under positive selection in a variety of plant species as well as mammalian lineages [44]. It must be emphasized that both CENP-A and CENP-C are predominantly single copy genes that are absolutely essential for chromosome segregation; therefore it is rather simple to rule out scenarios in which brief periods of relaxed selection after gene duplication might lead to the artefactual findings of positive selection. It is difficult to imagine scenarios in which such innovation would be selectively sponsored in essential genes in the absence of genetic conflict.

While a strong circumstantial case can be made that female meiotic asymmetry is the cause for runaway centromere complexity seen in some species, this question needs to be revisited in taxa that have chosen alternate paths for their meiotic programmes. Some, such as flies, plants and mammals, choose to employ both 'female' (or asymmetric) and 'male' (or symmetric) meioses. Others such as budding yeasts (including *S. cerevisiae*) only employ 'male' meiosis. We have argued that the absence of a transmission advantage associated with asymmetric female meiosis has allowed the centromeric sequences in budding yeasts to become 'simple' [13]. Furthermore, lack of female meiosis may have allowed budding yeast CenH3s to become optimally suited for their centromeres, with no evidence of positive selection. Other interesting counterpoints are provided by ciliate protozoans such as *Tetrahymena thermophila*, which only undergo 'female meiosis'. The cytological investigations of centromeric complexity and evolution of genes encoding their centromeric proteins [45] will provide both a valuable test of the 'centromere-drive' hypothesis and a suitable framework for future studies addressing the selective consequences of satellite sequences in the functioning of eukaryotic genomes.

References

1 Fitzgerald-Hayes, M., Clarke, L. and Carbon, J. (1982) Cell **29**, 235–244
2 Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K. and Willard, H.F. (2001) Science **294**, 109–115
3 Copenhaver, G.P., Nickel, K., Kuromori, T., Benito, M.I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D. et al. (1999) Science **286**, 2468–2474
4 Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B. and Henikoff, S. (1999) Nature (London) **401**, 547–548
5 Lo, A.W., Craig, J.M., Saffery, R., Kalitsis, P., Irvine, D.V., Earle, E., Magliano, D.J. and Choo, K.H. (2001) EMBO J. **20**, 2087–2096
6 Platero, J.S., Ahmad, K. and Henikoff, S. (1999) Mol. Cell **4**, 995–1004
7 Sun, X., Le, H.D., Wahlstrom, J.M. and Karpen, G.H. (2003) Genome Res. **13**, 182–194
8 Lohe, A.R. and Brutlag, D.L. (1987) J. Mol. Biol. **194**, 161–170
9 Haaf, T. and Willard, H.F. (1997) Chromosoma **106**, 226–232
10 Samonte, R.V., Ramesh, K.H. and Verma, R.S. (1997) Genetica **101**, 97–104
11 Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R. and Jiang, J. (2004) Nat. Genet. **36**, 138–145
12 McAllister, B.F. and Werren, J.H. (1999) J. Mol. Evol. **48**, 469–481
13 Malik, H.S. and Henikoff, S. (2002) Curr. Opin. Genet. Dev. **12**, 711–718
14 Charlesworth, B., Sniegowski, P. and Stephan, W. (1994) Nature (London) **371**, 215–220
15 Stephan, W. (1989) Mol. Biol. Evol. **6**, 198–212
16 Stephan, W. and Cho, S. (1994) Genetics **136**, 333–341
17 Walsh, J.B. (1987) Genetics **115**, 553–567
18 Wu, C.I., True, J.R. and Johnson, N. (1989) Nature (London) **341**, 248–251
19 Sawamura, K., Yamamoto, M.T. and Watanabe, T.K. (1993) Genetics **133**, 307–313
20 Sawamura, K. and Yamamoto, M.T. (1993) Mol. Gen. Genet. **239**, 441–449
21 Henikoff, S. and Malik, H.S. (2002) Nature (London) **417**, 227
22 Rhoades, M. (1942) Genetics **27**, 395–407
23 Dawe, R.K. and Cande, W.Z. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 8512–8517
24 Buckler, 4th, E.S., Phelps-Durr, T.L., Buckler, C.S., Dawe, R.K., Doebley, J.F. and Holtsford, T.P. (1999) Genetics **153**, 415–426
25 Zwick, M.E., Salstrom, J.L. and Langley, C.H. (1999) Genetics **152**, 1605–1614
26 Zhang, P., Knowles, B.A., Goldstein, L.S. and Hawley, R.S. (1990) Cell **62**, 1053–1062
27 Hawley, R.S., Irick, H., Zitron, A.E., Haddox, D.A., Lohe, A., New, C., Whitley, M.D., Arbel, T., Jang, J., McKim, K. et al. (1992) Dev. Genet. **13**, 440–467
28 Karpen, G.H., Le, M.H. and Le, H. (1996) Science **273**, 118–122
29 Malik, H.S. and Henikoff, S. (2001) Genetics **157**, 1293–1298
30 Henikoff, S., Ahmad, K. and Malik, H.S. (2001) Science **293**, 1098–1102
31 Fishman, L. and Willis, J.H. (2005) Genetics **169**, 347–353
32 Malik, H.S. (2005) Trends Ecol. Evol. **20**, 151–154
33 Eaker, S., Pyle, A., Cobb, J. and Handel, M.A. (2001) J. Cell Sci. **114**, 2953–2965
34 McKee, B.D., Wilhelm, K., Merrill, C. and Ren, X. (1998) Genetics **149**, 143–155
35 Pardo-Manuel de Villena, F. and Sapienza, C. (2001) Mamm. Genome **12**, 331–339
36 Pardo-Manuel de Villena, F. and Sapienza, C. (2001) Genetics **159**, 1179–1189

37 Nielsen, J. and Wohlert, M. (1991) Hum. Genet. **87**, 81–83
38 Daniel, A. (2002) Am. J. Med. Genet. **111**, 450–452
39 Sandler, L. and Novitski, E. (1957) Am. Nat. **41**, 105–110
40 Hartl, D.L. (1975) Theor. Popul. Biol. **7**, 168–174
41 Tao, Y., Hartl, D.L. and Laurie, C.C. (2001) Proc. Natl. Acad. Sci. U.S.A. **98**, 13183–13188
42 Merrill, C., Bayraktaroglu, L., Kusano, A. and Ganetzky, B. (1999) Science **283**, 1742–1745
43 Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L. and Henikoff, S. (2002) Plant Cell **14**, 1053–1066
44 Talbert, P.B., Bryson, T.D. and Henikoff, S. (2004) J. Biol. **3**, 18
45 Cervantes, M.D., Xi, X., Vermaak, D., Yao, M.C. and Malik, H.S. (2006) Mol. Biol. Cell **17**, 485–497