



---

The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA

Author(s): Steven Henikoff, Kami Ahmad, Harmit S. Malik

Source: *Science*, New Series, Vol. 293, No. 5532 (Aug. 10, 2001), pp. 1098-1102

Published by: American Association for the Advancement of Science

Stable URL: <http://www.jstor.org/stable/3084372>

Accessed: 05/11/2009 18:21

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=aaas>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



American Association for the Advancement of Science is collaborating with JSTOR to digitize, preserve and extend access to *Science*.

<http://www.jstor.org>

10. M. Wiekowski, M. Miranda, M. L. DePamphilis, *Dev. Biol.* **159**, 366 (1993).
11. F. Aoki, D. M. Worrad, R. M. Schultz, *Dev. Biol.* **181**, 296 (1997).
12. P. G. Adenot, Y. Mercier, J. P. Renard, E. M. Thompson, *Development* **124**, 4615 (1997).
13. D. M. Worrad, P. T. Ram, R. M. Schultz, *Development* **120**, 2347 (1994).
14. A. Razin, R. Shemer, *Hum. Mol. Genet.* **4**, 1751 (1995).
15. S. K. Howlett, W. Reik, *Development* **113**, 119 (1991).
16. J. P. Sanford, H. J. Clark, V. M. Chapman, J. Rossant, *Genes Dev.* **1**, 1039 (1987).
17. J. Sanford, L. Forrester, V. Chapman, A. Chandley, N. Hastie, *Nucleic Acids Res.* **12**, 2823 (1984).
18. P. M. Warnecke, S. J. Clark, *Mol. Cell. Biol.* **19**, 164 (1999).
19. T. Kafri *et al.*, *Genes Dev.* **6**, 705 (1992).
20. R. Shemer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11300 (1991).
21. W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **403**, 501 (2000).
22. M. Monk, M. Boubelik, S. Lehnert, *Development* **99**, 371 (1987).
23. Y. K. Kang *et al.*, *Nature Genet.* **28**, 173 (2001).
24. R. Daniels, V. Hall, A. O. Trounson, *Biol. Reprod.* **63**, 1034 (2000).
25. S. M. Tilghman, *Cell* **96**, 185 (1999).
26. M. Constanica, B. Pickard, G. Kelsey, W. Reik, *Genome Res.* **8**, 881 (1998).
27. K. L. Tucker *et al.*, *Genes Dev.* **10**, 1008 (1996).
28. T. Tada *et al.*, *Dev. Genes Evol.* **207**, 551 (1998).
29. Y. Kato *et al.*, *Development* **126**, 1823 (1999).
30. T. Kono, Y. Obata, T. Yoshimzu, T. Nakahara, J. Carroll, *Nature Genet.* **13**, 91 (1996).
31. J. B. Gurdon, R. A. Laskey, E. M. De Robertis, G. A. Partington, *Int. Rev. Cytol. Suppl.* **9**, 161 (1979).
32. N. Kikyo, P. A. Wade, D. Guschin, H. Ge, A. P. Wolffe, *Science* **289**, 2360 (2000).
33. K. E. Latham, J. I. Garrels, D. Solter, *Dev. Biol.* **163**, 341 (1994).
34. J. Kanka *et al.*, *Mol. Reprod. Dev.* **43**, 135 (1996).
35. Q. A. Winger *et al.*, *Mol. Reprod. Dev.* **56**, 458 (2000).
36. R. A. DePinho, *Nature* **408**, 248 (2000).
37. J. B. Gurdon, *Semin. Cell Dev. Biol.* **10**, 239 (1999).
38. M. F. Lyon, *Curr. Biol.* **9**, R235 (1999).
39. K. Eggan *et al.*, *Science* **290**, 1578 (2000).
40. P. G. Shiels *et al.*, *Nature* **399**, 316 (1999).
41. D. Betts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1077 (2001).
42. X. C. Tian, J. Xu, X. Yang, *Nature Genet.* **26**, 272 (2000).
43. R. P. Lanza *et al.*, *Science* **288**, 665 (2000).
44. T. Wakayama *et al.*, *Nature* **407**, 318 (2000).
45. C. M. Counter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14723 (1998).
46. I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind, K. H. Campbell, *Nature* **385**, 810 (1997).
47. J. R. Hill *et al.*, *Theriogenology* **51**, 1451 (1999).
48. L. E. Young, K. D. Sinclair, I. Wilmut, *Rev. Reprod.* **3**, 155 (1998).
49. K. J. McCreath *et al.*, *Nature* **405**, 1066 (2000).
50. P. A. De Sousa *et al.*, *Biol. Reprod.* **65**, 23 (2001).
51. T. Wakayama, R. Yanagimachi, *Nature Genet.* **22**, 127 (1999).
52. T. Wakayama, A. C. Perry, M. Zuccotti, K. R. Johnson, R. Yanagimachi, *Nature* **394**, 369 (1998).
53. T. Wakayama, I. Rodriguez, A. C. Perry, R. Yanagimachi, P. Mombaerts, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14984 (1999).
54. W. M. Rideout III *et al.*, *Nature Genet.* **24**, 109 (2000).
55. K. Eggan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6209 (2001).
56. A. Nagy *et al.*, *Development* **110**, 815 (1990).
57. K. H. Campbell, J. McWhir, W. A. Ritchie, I. Wilmut, *Nature* **380**, 64 (1996).
58. J. B. Cibelli *et al.*, *Science* **280**, 1256 (1998).
59. W. M. Rideout III, K. Eggan, R. Jaenisch, unpublished data.
60. C. Pinto-Correia, C. R. Long, T. Chang, J. M. Robl, *Mol. Reprod. Dev.* **40**, 292 (1995).
61. Y. Kato, T. Tani, Y. Tsunoda, *J. Reprod. Fertil.* **120**, 231 (2000).
62. D. N. Wells, P. M. Misica, H. R. Tervit, *Biol. Reprod.* **60**, 996 (1999).
63. C. L. Keefer *et al.*, *Biol. Reprod.* **64**, 849 (2001).
64. I. A. Polejaeva *et al.*, *Nature* **407**, 86 (2000).
65. X. Zou *et al.*, *Cloning* **3**, 31 (2001).
66. R. Jaenisch, *Trends Genet.* **13**, 323 (1997).
67. D. Humpherys *et al.*, *Science* **293**, 95 (2001).
68. S. Khosla, W. Dean, D. Brown, W. Reik, R. Feil, *Biol. Reprod.* **64**, 918 (2001).
69. L. E. Young *et al.*, *Nature Genet.* **27**, 153 (2001).
70. A. S. Doherty, M. R. Mann, K. D. Tremblay, M. S. Bartolomei, R. M. Schultz, *Biol. Reprod.* **62**, 1526 (2000).
71. D. Humpherys, K. Eggan, R. Jaenisch, unpublished data.
72. J. R. Hill *et al.*, *Biol. Reprod.* **62**, 1135 (2000).
73. C. Kubota *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 990 (2000).
74. A. E. Schnieke *et al.*, *Science* **278**, 2130 (1997).
75. W. Dean *et al.*, *Development* **125**, 2273 (1998).
76. I. L. Weissman, *Science* **287**, 1442 (2000).
77. J. Nichols *et al.*, *Cell* **95**, 379 (1998).
78. R. Jaenisch, I. Wilmut, *Science* **291**, 2552 (2001).
79. We thank C. Beard, K. Hochedlinger, L. Jackson-Grusby, A. Wutz, and D. Biniakiewicz for helpful comments and D. Humpherys for letting us discuss his unpublished data.

## REVIEW

# The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA

Steven Henikoff,\* Kami Ahmad, Harmit S. Malik

Every eukaryotic chromosome has a centromere, the locus responsible for poleward movement at mitosis and meiosis. Although conventional loci are specified by their DNA sequences, current evidence favors a chromatin-based inheritance mechanism for centromeres. The chromosome segregation machinery is highly conserved across all eukaryotes, but the DNA and protein components specific to centromeric chromatin are evolving rapidly. Incompatibilities between rapidly evolving centromeric components may be responsible for both the organization of centromeric regions and the reproductive isolation of emerging species.

Inheritance of genetic information requires a faithful copying mechanism. DNA replication and repair provide high-fidelity inheritance over evolutionary periods, whereas epigenetic inheritance is less rigid. DNA methylation can mediate epigenetic inheritance during development and even between generations of complex organisms (1). Stable protein-based inheritance is well-established for prions (2), and chromatin-based mechanisms are thought to maintain developmental states (3). Over the past decade, several authors have argued that centromeres, the sites of spindle attachment at mitosis and meiosis, can also be maintained epigenetically (4–8). Here, we examine recent studies on the basis

for centromeric inheritance. These point to a novel chromatin-based mechanism for the maintenance of centromere location during multiple rounds of cell division. This mechanism may be responsible for the enigmatic organization of centromeric DNA and for the rapid onset of reproductive isolation as species emerge.

## DNA Sequences at Centromeres

Delimiting the precise boundaries of centromeres has proven to be a daunting task. In animals and plants, centromeres are contained within regions of highly repetitive satellite DNA, which confounds even the most powerful mapping methods. Chromosomes in

*Saccharomyces cerevisiae* are exceptional in that they lack satellite sequences and their centromeres have been precisely localized. Each of these “point” centromeres specifies spindle attachment with only ~125 base pairs (bp) of DNA. However, this simplicity is evolutionarily derived, as centromeres from other fungal lineages include arrays of repeats (9, 10), much like what is found in animals and plants.

The simplicity of *S. cerevisiae* centromeres on small chromosomes led to the idea that specific repeated sequence elements might specify centromere location for larger chromosomes when present in a sufficient number of copies (11). However, this hypothesis has fallen out of favor for reasons that have been extensively reviewed (5–7). Most compelling is the lack of any common repetitive elements in many “neocentromeres” that

Howard Hughes Medical Institute Research Laboratories, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA.

\*To whom correspondence should be addressed. E-mail: steveh@fhcrc.org

occasionally are found in humans. Recent studies have documented human neocentromeres that lack  $\alpha$ -satellite (12–15), which is the single common DNA sequence found in native human centromeres. Even fine-structure mapping of two human neocentromeres has failed to detect  $\alpha$ -satellite or other tandem repeats found near centromeres (15, 16).

Could there be more subtle sequence motifs that have gone undetected within complex centromeres? Unfortunately, there are several technical difficulties in characterizing centromeric satellites. Centromeric and flanking sequence appear to be indistinguishable in a well-studied *Drosophila* minichromosome (17), and it is unknown what subset of  $\alpha$ -satellite is found at human centromeres as opposed to satellites found in surrounding heterochromatin (18). In addition, there are no reports of a complete satellite-containing centromere that has been cloned, much less sequenced. In this “postgenomic” era, centromeric sequencing has only just begun.

A candidate centromeric motif would have to be reiterated over hundreds of kilobases, because this is the minimal size of fully functional centromeres in several organisms. In *Drosophila*, mapping of a minichromosome centromere has determined that 420 kb of primarily tandem repeats is required for full function and that removal of material from either end leads to a progressive reduction in transmission (19). In maize, a fully functional supernumerary B chromosome centromere contains ~500 kb of tandem repeats, and partial deletions reduce transmission (20). In humans, even the smallest minichromosomes retain at least 100 kb of  $\alpha$ -satellite (21), and human artificial chromosomes are found to contain  $\alpha$ -satellite arrays in the megabase range (22, 23). Therefore, “regional” rather than point centromeres are characteristic of both animals and plants.

One interesting feature of most centromeric satellite repeats is their unit length. Despite the lack of universal sequence motifs, repeat unit length can be remarkably similar between organisms. For example, the basic  $\alpha$ -satellite unit in primates is 171 bp, in the fish *Sparus aurata*, the centromeric repeat is 186 bp, in the insect *Chironomus pallidivittatus*, it is 155 bp, in both *Arabidopsis* and maize, it is 180 bp, and in rice, it is 168 bp (24–26). It is striking that the narrow range of repeat lengths found for these centromeric satellites corresponds closely to the range of nucleosomal unit lengths (11, 27). Larger repeat lengths, such as the 340-bp repeat found at pig centromeres, roughly encompass two nucleosomes. There may be exceptions, such as the pentameric satellite repeats in *Drosophila melanogaster*. These exceptions notwithstanding, selection for nucleosomal length might sometimes constrain evolution of centromeric satellites, consistent with their structural (noncoding) role in the genome.

### Centromeric Proteins

The failure to detect common motifs that distinguish centromeres from noncentromeric DNAs has led to speculations that non-DNA sequence determinants maintain centromeres (5, 7). Such determinants might account for examples of reversible centromere inactivation (28, 29). However, evidence is lacking for DNA modifications, noncoding RNAs, or enzymatic activities that might uniquely “mark” centromeres. Nevertheless, there are proteins found only at centromeres, and these are candidates for maintaining centromeric location. Many of these proteins are present at centromeres only at mitosis (30). Some of these form the kinetochore, a proteinaceous structure that assembles on centromeric chromatin and connects the centromere to spindle microtubules. The kinetochore appears at prophase and disappears at telophase (31). Other protein complexes, such as cohesins, are not part of the kinetochore but also disappear from centromeric regions at mitosis. Mutations in kinetochore or cohesion components compromise chromosome segregation; however, only constitutive proteins are candidates for directly maintaining centromeres.

The best candidate for maintaining mammalian centromeres is CENP-A (32), which is of special interest because it is the centromeric histone (33, 34). CENP-A is present at native centromeres and at neocentromeres (30, 35, 36) but is absent from centromeres that are mutated (13) or inactivated (37, 38). CENP-A, a member of the histone H3 protein family, copurifies with nucleosomes (33), replaces H3 in purified nucleosome particles (27), and can be assembled in vitro into nucleosomal core particles lacking H3 (39, 40). In vivo, each centromeric nucleosome consists of a histone octamer with CENP-A instead of H3 (27). Because CENP-A survives the protamine transition that removes all noncentromeric histones (33) and is present in sperm as distinct foci (4), it behaves as a heritable centromeric molecule.

CENP-A has counterparts in other eukaryotes (Fig. 1). Experimental evidence has confirmed that Cse4p in *S. cerevisiae* (41), HCP-3 in *Caenorhabditis elegans* (43) (Fig. 1B), Cid (for centromere identifier) in *D. melanogaster* (42) (Fig. 1D), and SpCENP-A in *Schizosaccharomyces pombe* (44) are exclusively centromeric. For example, in *S. pombe*, SpCENP-A is found only at the non-redundant central core region and not in the surrounding tandem repeats (44). In addition, molecular mapping of CENP-A in two human neocentromeres localizes it to the center of a region that includes the primary constriction (15, 16), which is the consistent cytological feature of regional centromeres in metaphase chromosomes. Indeed, wherever there is a centromere marker in metaphase chromosomes, the centromeric histone is found:

Even in *C. elegans*, where the holokinetochore extends from one end of the chromosome to the other, HCP-3 is found at the underlying holocentromere (Fig. 1B). At the other end of the spectrum, the point centromeres of *S. cerevisiae* have what may be a single Cse4p-containing centromeric nucleosome (41). Thus, a centromeric H3-like histone underlies the vastly different types of centromeres found across eukaryotes.

Centromeric histones contain sequence features that distinguish them from histone H3, including a noncanonical NH<sub>2</sub>-terminal tail, a more divergent core histone fold, and a slightly longer loop 1 region (27, 45) (Fig. 1A). Although histone H3 is evolutionarily constrained, centromeric histones are strikingly divergent. We attribute this difference to the necessity of H3 to interact with the entire genome, whereas each centromeric variant need only interact with the current centromeric DNA (45). This DNA consists of satellite repeats, which are the most rapidly evolving components of eukaryotic genomes (46). We propose that the interaction between the centromeric histone and centromeric DNA is responsible for the approximately nucleosomal repeat lengths found for satellites.

In addition to the centromeric histone, other constitutive components of centromeres have been identified. Only one of these, mammalian CENP-C, is evolutionarily conserved, sharing a single conserved motif of less than 20 amino acids in length with *S. cerevisiae* centromere protein, Mif2. Among sequenced eukaryotes, a homologous sequence has been found in many organisms, including HCP-4 in *C. elegans* (47, 48), but not in *Drosophila* or *Giardia*. The mitotic localization of CENP-C and HCP-4 depends on the presence of the centromeric histone (47–49) [but not vice-versa (47, 48, 50)]. Furthermore, HCP-4 does not localize to centromeres at interphase (47, 48). Other proteins have been found to be constitutive components of centromeres, but only in particular organisms (51). In yeasts, protein complexes that are constitutively present at centromeres appear to be required for localization of centromeric nucleosomes (44, 52). However, no counterparts to these complexes have been documented in other organisms. In *S. cerevisiae*, the presence of these complexes at centromeres could be a consequence of the constitutive nature of the kinetochore itself (53). Thus, the centromeric H3-like histone is the only current candidate protein for universally maintaining centromeres.

### Maintenance of Centromeric Chromatin

Nucleosomes are distributed between daughter strands at replication (54), and current evidence favors a similar distributive segregation of centromeric nucleosomes (55). For

centromeres to be maintained, new centromeric nucleosomes must be assembled, and H3-containing nucleosomes must be excluded

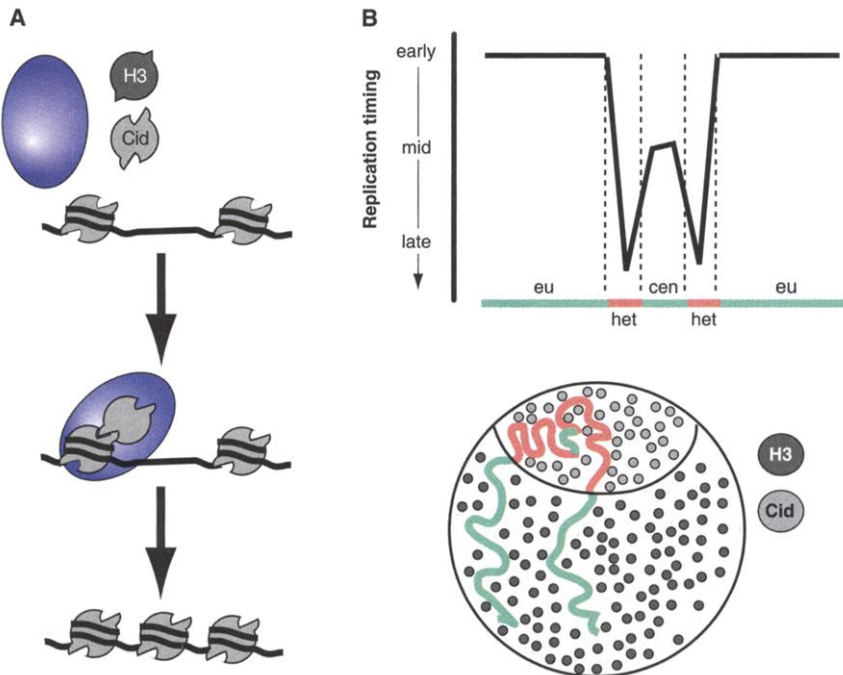
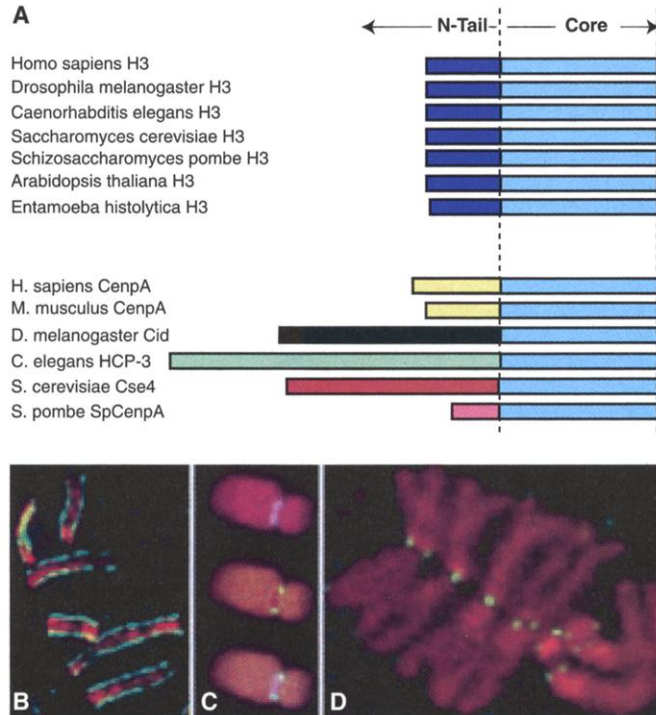
to prevent a gradual degradation of centromeric identity. Two general models have been proposed for this selective assembly. In

one, old centromeric nucleosomes specifically direct the deposition of new centromeric nucleosomes (6, 8) (Fig. 2A). This process could be mediated by an interaction between centromeric nucleosomes or facilitated by a nucleosome loading factor (44). Adenosine triphosphate-dependent nucleosome remodeling complexes are attractive mediators of this process, given that centromeric nucleosomes can be assembled without DNA replication, both in humans (55) and in flies (56). A second model supposes that centromeric histones are available only at a restricted time or nuclear location during the cell cycle (27, 42, 46) (Fig. 2B). This restriction would prevent the poisoning of centromeres by the vast excess of histone H3 and promote the assembly of centromeric nucleosomes. These two models are not mutually exclusive, as it seems likely that centromere maintenance involves both restricted availability of histones during centromere replication (56) and nucleosome remodeling after replication (8).

Centromeres replicate in mid S phase (56–58); however, they are embedded within a heterochromatin compartment in which all other DNA replicates late (56). Should this nonreplicating compartment exclude replication-dependent histone H3 but sequester centromeric histones, the compartment would ensure that only the latter are available for assembly. This model is supported by four lines of evidence. First, both the *Drosophila cid* and *S. pombe Crp1* genes are expressed during a brief period of early S phase (42, 44) [although human Cenp-A does not follow this pattern (55)], suggesting early cell cycle availability of centromeric histones. Second, centromeres replicate before heterochromatin in mammals (57), barley (58), and *Drosophila* (56) (Fig. 2B). Third, histone H3 is not deposited at replicating *Drosophila* centromeres (56). Fourth, centromeric histones from diverse species appear to become sequestered within both *Drosophila* and human heterochromatin (42).

Consistent with the sequestration model, most native centromeres are inextricably linked to late-replicating heterochromatin and both consist of highly repetitive DNA. This has made it difficult to determine whether there is a causal relationship between centromeres and pericentric heterochromatin. However, human neocentromeres lacking repeats have been compared with their noncentromeric parents, and in this way, essential centromeric components can be identified. Several nonrepetitive neocentromeres accumulate Heterochromatin-associated Protein 1 (HP1), suggesting a causal link between heterochromatin and centromere function (30). Consistent with this view, late replication timing, a hallmark of heterochromatin, is an acquired feature of the regions on both sides flanking the ~300 kb CENP-A–

**Fig. 1. (A)** Schematic alignment of centromeric histones. Centromeric histones are similar to H3 in their core domains (light blue) but differ in their NH<sub>2</sub>-terminal tails even from each other, indicated by the various colors. **(B)** Antibody to HCP-3 labeling (green) of *C. elegans* holocentric chromosomes (red) at prophase (94). **(C)** Antibody to CENP-A (green) and antibody to CENP-B labeling (purple) of human metaphase chromosomes (red) [from (38) (with permission)]. CENP-A is located precisely at the centromere, whereas CENP-B binds to surrounding  $\alpha$ -satellite. **(D)** Antibody to Cid labeling (green) of *D. melanogaster* prometaphase chromosomes (red) (95). The sister centromeres and holocentromeres have split.



**Fig. 2. Proposed mechanisms for centromere maintenance in dividing cells. (A)** Recognition model (8). After replication of centromeric DNA, gaps in nucleosome arrays are filled by a replication-independent mechanism. Old centromeric nucleosomes discriminate centromeric octamers (light gray) from H3-containing octamers (black), either by a direct interaction or indirectly through a chromosome remodeling factor (blue), leading to specific assembly of new centromeric nucleosomes. **(B)** Sequestration model, showing replication timing along the chromosome (15, 56) (top). In the early S phase nucleus, histone H3 is excluded from the heterochromatic compartment, but centromeric histones are enriched (bottom). This leads to exclusive deposition of centromeric histones as centromeres replicate. Centromeres (cen), like euchromatin (eu) (both green) replicate earlier than their flanking heterochromatin (het) (orange).

associated region of the mardel(10) neocentromere (15). Thus, the best characterized human centromere appears to be remarkably similar to *S. pombe* centromeres, where a central core containing the SpCENP-A centromeric histone is surrounded by repeats with heterochromatic features (44, 59).

These recent cytogenetic and molecular studies complement a growing body of genetic evidence that has revealed that heterochromatin is required for centromere function. Reduced chromosome transmission results from removal of flanking heterochromatin in *Drosophila* (60) and *S. pombe* (61). In both organisms, modifiers that disrupt heterochromatic silencing impair chromosome transmission (59, 60, 62). In *S. pombe*, this disruption of silencing and impaired transmission is accompanied by the hyperacetylation of histones flanking the central core (63, 64). Similarly, inhibition of histone deacetylases in mammalian cells causes delocalization of HP1 from heterochromatin and reduced chromosome transmission (65). Thus, flanking heterochromatin is a prerequisite for maintaining regional centromeres. Perhaps loss of heterochromatin was responsible for the utilization of DNA-binding proteins for centromere localization in budding yeast and the emergence of holocentricity in nematodes.

### Centromere Evolution

Centromeric repeats comprise the most rapidly evolving DNA sequences in eukaryotic genomes, differing even between closely related species (46, 66, 67). This is not to say that satellites are hypermutable but rather that sequence variants are fixed by expansion and contraction (68) and can arise de novo at new sites (69). These satellite changes are brought about by a variety of mutational processes, including replication slippage, unequal exchange, transposition, and excision (68, 70). Such rapid change is paradoxical: Why has not a single optimal sequence been fixed at centromeres? A clue comes from examination of centromeric histones. These are expected to maintain favorable interactions with centromeric satellite (27). Comparison of Cid from closely related *Drosophila* species reveals that both the NH<sub>2</sub>-terminal tail and the histone core domain contain regions that have undergone frequent episodes of adaptive evolution (45). This is unexpected for a histone molecule, as histones are among the most evolutionarily constrained eukaryotic proteins. Within the histone core domain, most adaptive changes lie in loop 1, a region that makes direct H3-DNA contacts (71), suggesting that centromeric histone binding is sequence dependent (27). Consistent with this finding, all confirmed centromeric histones have a longer loop 1 than H3 (27, 72). Thus, the adaptive signal and its location provide compelling evidence that Cid has evolved in concert with centromeric DNA, reminiscent of the co-

evolution of rDNA and its transcription factor (70). Similarly, the NH<sub>2</sub>-terminal tail may be adapting under constraints imposed by centromeric DNA. Understanding the basis of these adaptive changes could resolve the paradox of rapidly evolving centromeres.

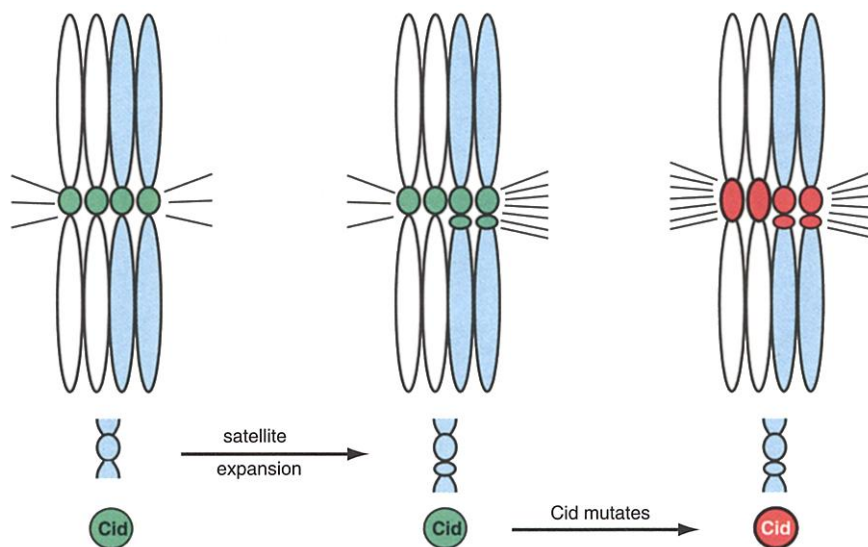
Asymmetry at female meiosis may be the key. Of the four products of meiosis, three are lost and only one becomes the oocyte nucleus. There is evidence that the asymmetry of the meiotic tetrad provides an opportunity for chromosomes to compete for inclusion into the oocyte nucleus by attaining a preferable orientation at meiosis (73–75). Centromeres that can exploit this opportunity at meiosis I will “win,” and even a slight advantage at each female meiosis will be enough to rapidly drive a centromere to fixation. Additional recruitment of centromeric nucleosomes, for example, by the expansion of a centromeric satellite, would confer this advantage (Fig. 3). Genetic evidence that some animal and plant centromeres are “stronger” at meiosis dates back nearly half a century (73, 76). In maize, centromere strength is characteristic of heterochromatic “knobs,” which display poleward movement and meiotic drive during female meiosis (25, 73), and a similar drive process might contribute to the success of selfish B chromosomes (77). In humans, a variety of Robertsonian translocations, with two adjacent centromeres, consistently display a higher than expected transmission ratio (78).

In females, the winning centromeres simply exploit the inherently destructive process of forming the egg and thus might not reduce fecundity. However, in *Drosophila* males, heterochromatic differences between paired

chromosomes at meiosis I can cause nondisjunction (79) manifested as skewed sex ratios or infertility. We propose that these chromosome pairs have centromeric imbalances. Cid is the best candidate to relieve deleterious effects associated with centromere meiotic drive (45). For example, if Cid were to mutate such that it preferentially bound the weaker centromere, centromeric balance would be restored (Fig. 3). Such a beneficial *cid* allele will drive to fixation itself. This process suffices to explain both the evolutionary dynamics of satellite DNA and the adaptive evolution of Cid (80). Episodes of drive and deleterious mutation by transposons (17, 18) would lead to the accumulation of satellites representing centromeric relics surrounding functional centromeres (46, 81). This would also provide a mechanism for the well-documented fixation of chromosome-specific satellites (67) in successive episodes of drive.

Consider this process occurring in two isolated populations of the same species. Satellite-Cid configurations will diverge rapidly. In each population, Cid will evolve to suppress the deleterious effects of satellites that have driven through that population. By so doing, Cid becomes incompatible with the independently evolving centromeric satellites in the other population. Crosses between the populations will result in hybrid defects as centromeric drive is released again. Thus, the satellite-Cid drive process results in the onset of reproductive isolation between the two populations. In other words, speciation is an inevitable consequence of centromere evolution (82).

Hybrids between closely related species



**Fig. 3.** Centromere drive model. Expansion of a satellite that binds Cid provides more microtubule attachment sites. This stronger centromere drives in female meiosis but also leads to increased nondisjunction. A mutation in Cid that alters sequence specificity leads to more extensive binding of the weaker centromere, providing more microtubule attachment sites. This restores meiotic balance and alleviates nondisjunction.

display a common syndrome of defects, most prominently, the sterility and inviability of the heterogametic sex, referred to as Haldane's rule (83). In most species, the XY male is the heterogametic sex; however, in many lineages, such as birds and butterflies, the ZW female is heterogametic. We think that our model is sufficient to explain the bias for hybrid sterility. In hybrids, there would be two sets of centromeric satellites and two alleles encoding the centromeric histone. The imbalance in centromere strength would be highest between the heterogametic pair, regardless of whether that pair is XY or ZW, because the centromeres of the heterogametic pair of centromeres are always the most dissimilar. Thus, infertility and distortion of the sex ratio (84) would be symptoms displayed by hybrids bringing together incompatible satellite-Cid combinations (85).

Because centromeric histones are found in all eukaryotes, and asymmetric female meiosis has been documented in both animals and plants, our model for the origin of species is widely applicable. The model is testable, in that adaptive evolution should be found for centromeric histones in all organisms with asymmetric meiosis, but not necessarily in organisms with symmetric meiosis. Moreover, we predict that species in early stages of postzygotic reproductive isolation will show critical changes in residues that mediate contacts between the centromeric histone and centromeric DNA. If this model survives these tests, then we will have had a glimpse into how the basic machinery of chromosome segregation can result in the astonishing diversifications of life.

#### References and Notes

- Martienssen, *Science* **293**, 1070 (2001).
- R. B. Wickner, D. C. Masion, H. K. Edskes, *Yeast* **11**, 1671 (1995).
- F. Lyko, R. Paro, *Bioessays* **21**, 824 (1999).
- D. K. Palmer, K. O'Day, R. L. Margolis, *Chromosoma* **100**, 32 (1990).
- G. H. Karpen, R. C. Allshire, *Trends Genet.* **13**, 489 (1997).
- H. F. Willard, *Curr. Opin. Genet. Dev.* **8**, 219 (1998).
- K. H. Choo, *Trends Cell Biol.* **10**, 182 (2000).
- K. F. Sullivan, *Curr. Opin. Genet. Dev.* **11**, 182 (2001).
- B. Fishel, H. Amstutz, M. Baum, J. Carbon, L. Clarke, *Mol. Cell. Biol.* **8**, 754 (1988).
- E. B. Cambareli, R. Aisner, J. Carbon, *Mol. Cell. Biol.* **18**, 5465 (1998).
- H. F. Willard, *Trends Genet.* **6**, 410 (1990).
- T. W. Depinet et al., *Hum. Mol. Genet.* **6**, 1195 (1997).
- C. Tyler-Smith et al., *Am. J. Hum. Genet.* **64**, 1440 (1999).
- P. E. Warburton et al., *Am. J. Hum. Genet.* **66**, 1794 (2000).
- A. W. Lo et al., *EMBO J.* **20**, 2087 (2001).
- A. W. Lo et al., *Genome Res.* **11**, 448 (2001).
- X. Sun, J. Wahlstrom, G. Karpen, *Cell* **91**, 1007 (1997).
- J. Puechberty et al., *Genomics* **56**, 274 (1999).
- T. D. Murphy, G. H. Karpen, *Cell* **82**, 599 (1995).
- E. Kaszas, J. A. Birchler, *Genetics* **150**, 1683 (1998).
- J. W. Yang et al., *Hum. Mol. Genet.* **9**, 1891 (2000).
- J. J. Harrington, G. Van Bokkelen, R. W. Mays, K. Gustashaw, H. F. Willard, *Nature Genet.* **15**, 345 (1997).
- M. Ikeno et al., *Nature Biotechnol.* **16**, 431 (1998).
- K. H. A. Choo, *The Centromere* (Oxford Univ. Press, Oxford, 1997).
- R. K. Dawe, W. Z. Cande, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8512 (1996).
- F. Dong et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8135 (1998).
- R. D. Shelby, O. Vafa, K. F. Sullivan, *J. Cell Biol.* **136**, 501 (1997).
- A. W. Higgins, M. G. Schueler, H. F. Willard, *Chromosoma* **108**, 256 (1999).
- M. Agudo et al., *Chromosoma* **109**, 190 (2000).
- R. Saffery et al., *Hum. Mol. Genet.* **9**, 175 (2000).
- J. M. Craig, W. C. Earnshaw, P. Vagnarelli, *Exp. Cell Res.* **246**, 24 (1999).
- W. C. Earnshaw, N. Rothfield, *Chromosoma* **91**, 313 (1985).
- D. K. Palmer, K. O'Day, H. L. Trong, H. Charbonneau, R. L. Margolis, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3734 (1991).
- K. F. Sullivan, M. Hechenberger, K. Masri, *J. Cell Biol.* **127**, 581 (1994).
- O. Vafa, K. F. Sullivan, *Curr. Biol.* **7**, 897 (1997).
- O. Vafa, R. D. Shelby, K. F. Sullivan, *Chromosoma* **108**, 367 (1999).
- B. A. Sullivan, H. F. Willard, *Nature Genet.* **20**, 227 (1998).
- P. E. Warburton et al., *Curr. Biol.* **7**, 901 (1997).
- K. Yoda et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7266 (2000).
- The presence of centromeric histones may generate a chromatin structure unique to active centromeres (44, 86).
- P. B. Meluh, P. Yang, L. Glowczewski, D. Koshland, M. M. Smith, *Cell* **94**, 607 (1998).
- S. Henikoff, K. Ahmad, J. S. Platero, B. van Steensel, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 716 (2000).
- B. J. Buchwitz, K. Ahmad, L. L. Moore, M. B. Roth, S. Henikoff, *Nature* **401**, 547 (1999).
- K. Takahashi, E. S. Chen, M. Yanagida, *Science* **288**, 2215 (2000).
- H. S. Malik, S. Henikoff, *Genetics* **157**, 1293 (2001).
- A. K. Csirik, S. Henikoff, *Trends Genet.* **14**, 200 (1998).
- L. L. Moore, M. B. Roth, *J. Cell Biol.* **153**, 1199 (2001).
- K. Oegema, A. Desai, S. Rybina, M. Kirkham, A. A. Hyman, *J. Cell Biol.* **153**, 1209 (2001).
- E. V. Howman et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1148 (2000).
- P. Kalitsis, K. J. Fowler, E. Earle, J. Hill, K. H. Choo, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1136 (1998).
- One example is the kinesinlike protein Cenp-meta in *Drosophila* (87). Several others localize to regions that surround centromeres, including Heterochromatin-associated Protein 1, mammalian CenpB, and Topoisomerase II (30).
- J. Ortiz, O. Stemmann, S. Rank, J. Lechner, *Genes Dev.* **13**, 1140 (1999).
- M. Winey, E. T. O'Toole, *Nature Cell Biol.* **3**, E23 (2001).
- V. Jackson, *Biochemistry* **27**, 2109 (1988).
- R. D. Shelby, K. Monier, K. F. Sullivan, *J. Cell Biol.* **151**, 1113 (2000).
- K. Ahmad, S. Henikoff, *J. Cell Biol.* **153**, 101 (2001).
- R. T. O'Keefe, S. C. Henderson, D. L. Spector, *J. Cell Biol.* **116**, 1095 (1992).
- Z. Jasencakova, A. Meister, I. Schubert, *Chromosoma* **110**, 83 (2001).
- J. F. Partridge, B. Borgstrom, R. C. Allshire, *Genes Dev.* **14**, 783 (2000).
- D. R. Wines, S. Henikoff, *Genetics* **131**, 683 (1992).
- N. C. Steiner, L. Clarke, *Cell* **79**, 865 (1994).
- K. Ekwall, G. Cranston, R. C. Allshire, *Genetics* **153**, 1153 (1999).
- K. Ekwall, T. Olsson, B. M. Turner, G. Cranston, R. C. Allshire, *Cell* **91**, 1021 (1997).
- S. I. Grewal, M. J. Bonaduce, A. J. Klar, *Genetics* **150**, 563 (1998).
- A. Taddei, C. Maisson, D. Roche, G. Almouzni, *Nature Cell Biol.* **3**, 114 (2001).
- T. D. Murphy, G. H. Karpen, *Cell* **93**, 317 (1998).
- T. Haaf, H. F. Willard, *Chromosoma* **106**, 226 (1997).
- B. Charlesworth, P. Sniegowski, W. Stephan, *Nature* **371**, 215 (1994).
- For example, different satellites have arisen on homologous chromosomes since the divergence of humans and chimpanzees (67). In *Drosophila*, the enigma is even more severe, as completely unrelated satellites are found on homologous chromosomes in closely related species (46). Even the best characterized *Drosophila* centromere consists primarily of two unrelated pentameric satellite blocks (17).
- G. A. Dover, *Curr. Opin. Genet. Dev.* **3**, 902 (1993).
- K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **389**, 251 (1997).
- K. C. Keith et al., *Mol. Cell. Biol.* **19**, 6130 (1999).
- M. M. Rhoades, in *Heterosis*, J. W. Gowen, Ed. (Iowa St. College Press, Ames, IA, 1952), pp. 66–80.
- M. E. Zwick, J. L. Salstrom, C. H. Langley, *Genetics* **152**, 1605 (1999).
- F. Pardo-Manuel de Villena, C. Sapienza, *Mam. Genome* **12**, 331 (2001).
- E. Novitski, *J. Cell. Comp. Physiol.* **45**, 151 (1955).
- M. M. Rhoades, *Genetics* **27**, 395 (1942).
- F. Pardo-Manuel de Villena, C. Sapienza, *Hum. Genet.* **108**, 31 (2001).
- B. D. McKee, K. Wilhelm, C. Merrill, X. Ren, *Genetics* **149**, 143 (1998).
- Cid might not be the only protein whose adaptive evolution can alleviate the deleterious effects of satellite drive. Heterochromatic DNA-binding proteins, such as HOAP, might rapidly evolve in order to occlude satellite DNAs from centromeric status (88, 89).
- E. S. Buckler 4th et al., *Genetics* **153**, 415 (1999).
- The satellite-Cid model for reproductive isolation can be compared with the classical two-locus hypothesis for speciation (90). In that model, an ancestral population ( $A_1A_1B_1B_1$ ) evolves in one population to  $A_1A_2B_1B_1$  and in the other population to  $A_2A_2B_1B_1$ . The alleles  $A_2$  and  $B_2$  are neutral in the background of  $B_1$  and  $A_1$ , respectively; however, in hybrids ( $A_1A_2B_1B_2$ ), an epistatic interaction between  $A_2$  and  $B_2$  is deleterious. One problem with this model has been the lack of a plausible mechanism for the deleterious interaction between  $A_2$  and  $B_2$ . In addition, neither  $A_1A_2B_2B_2$  nor  $A_2A_2B_1B_1$  would be incompatible with their ancestor  $A_1A_1B_1B_1$ . Our model for satellite-Cid evolution implies an inherently rapid drive process (91, 92) together with a molecular mechanism for hybrid incompatibilities. This drive is irreversible in that  $A_1A_1B_1B_1$  would give rise to  $A_2A_2B_2B_2$  and  $A_2A_2B_1B_1$ , and so both derived populations become incompatible with their ancestral population.
- C. C. Laurie, *Genetics* **147**, 937 (1997).
- In birds, sex ratio distortion can result from a competition between the Z and W sex chromosomes for inclusion in the oocyte. Remarkably, female birds appear to have the ability to choose between the Z and W at meiosis on the basis of external cues (93).
- The rapid evolution of centromeric histones might also have mitotic consequences. For example, dicentric chromosomes might appear in hybrids because of the different sequence preferences of their centromeric histones. The resulting breakage and aneuploidy will reduce viability. The effects of X or Z dicentrics will be more severe in XY or ZW than XX or ZZ, consistent with the "dominance" theory of Haldane's rule for hybrid inviability (90).
- G. Florida, A. Zatterale, O. Zuffardi, C. Tyler-Smith, *EMBO Rep.* **1**, 489 (2000).
- J. K. Yucel et al., *J. Cell Biol.* **150**, 1 (2000).
- M. M. Shareef et al., *Mol. Biol. Cell* **12**, 1671 (2001).
- K. J. Schmid, L. Nigro, C. F. Aquadro, D. Tautz, *Genetics* **153**, 1717 (1999).
- J. A. Coyne, H. A. Orr, *Philos. Trans. R. Soc. London Ser. B* **353**, 287 (1998).
- S. A. Frank, *Evolution* **45**, 252 (1991).
- L. D. Hurst, A. Pomiankowski, *Genetics* **128**, 841 (1991).
- B. C. Sheldon, *Curr. Biol.* **9**, R487 (1999).
- B. Buchwitz, L. Moore, M. Roth, unpublished data.
- J. S. Platero, K. Ahmed, unpublished data.
- We thank S. Biggins, B. Byers, D. Gottschling, C. Langley, M. Roth, P. Talbert, and D. Vermaak for helpful discussions and P. Warburton for Fig. 1C. We are supported by the Howard Hughes Medical Institute (S.H.), the American Cancer Society (K.A.), and the Helen Hay Whitney Foundation (H.S.M.).