Adaptive Evolution of Cid, a Centromere-Specific Histone in Drosophila

Harmit S. Malik* and Steven Henikoff*,†

†*Howard Hughes Medical Institute,* **Fred Hutchinson Cancer Research Center, Seattle, Washington 98109*

Manuscript received October 2, 2000
Accepted for publication November 29, 2000

ABSTRACT

Centromeric DNA is generally composed of large blocks of tandem satellite repeats that change rapidly due to loss of old arrays and expansion of new repeat classes. This extreme heterogeneity of centromeric DNA is difficult to reconcile with the conservation of the eukaryotic chromosome segregation machinery. Histone H3-like proteins, including Cid in *Drosophila melanogaster*, are a unique chromatin component of centromeres. In comparisons between closely related species of Drosophila, we find an excess of replacement changes that have been fixed since the separation of *D. melanogaster* and *D. simulans*, suggesting adaptive evolution. The last adaptive changes appear to have occurred recently, as evident from a reduction in polymorphism in the *melanogaster* lineage. Adaptive evolution has occurred both in the long N-terminal tail as well as in the histone fold of Cid. In the histone fold, the replacement changes have occurred in the region proposed to mediate binding to DNA. We propose that this rapid evolution of *Cid* is driven by a response to the changing satellite repeats at centromeres. Thus, centromeric H3-like proteins may act as adaptors between evolutionarily labile centromeric DNA and the conserved kinetochore machinery.

WHAT defines a centromere? In the budding yeast *Saccharomyces cerevisiae*, centromeric function can be assigned to a distinct 125-bp consensus sequence (Wiens and Sorger 1998); however, this simplicity turns out to be exceptional. In most complex eukaryotes, there is no such candidate sequence. Rather, centromeres are composed of large blocks of highly repetitive satellite sequences. These satellite sequences, such as α-satellites in primates (Willard 1991; Murphy and Karpen 1998) and 180-bp repeats in *Arabidopsis thaliana* (Round *et al.* 1997; Copenhaver *et al.* 1999), are not always unique to centromeres. This underlying complexity of sequence appears at odds with the extreme conservation of the chromosome segregation apparatus, with many components conserved throughout the history of Eukarya.

What all eukaryotic centromeres have in common is the presence of atypical H3-like proteins packaged into specialized nucleosomes (Palmer *et al.* 1991; Stoler *et al.* 1995). These H3-like proteins have been described as CENP-A in humans (Palmer *et al.* 1991; Shelby *et al.* 1997), Cse4p in *S. cerevisiae* (Stoler *et al.* 1995), Cid in *D. melanogaster* (Henikoff *et al.* 2000), HCP3 in *Caenorhabditis elegans* (Buchwitz *et al.* 1999), and spCENP-A in *Schizosaccharomyces pombe* (Takahashi *et al.* 2000). They share 35–50% identity with histone H3 in the 100-amino-acid core histone fold, compared to >75% identity among the histone H3s themselves (Figure 1B). However, they are distinguished from histone H3 by the presence of an atypical N-terminal tail (Figure 1A), which is conserved in neither sequence nor length, even among the centromeric H3-like molecules. It has been demonstrated that specific subdomains of CENP-A are responsible for centromeric localization (Shelby *et al.* 1997). Various models have been proposed to account for this localization: by recognition of satellite arrays by the CENP-A containing nucleosomes (Shelby *et al.* 1997), by recognition of pre-existing CENP-A (Palmer *et al.* 1990), or by the replication timing of centromeres (Shelby *et al.* 1997; Csink and Henikoff 1998).

Transfection studies have strongly suggested the competence of α-satellites for centromere formation in humans (Harrington *et al.* 1997; Masumoto *et al.* 1998). Three types of studies challenge this view. First is the absence of a universal "centromeric" sequence. In fact, centromeric satellite repeats can differ at different chromosomes, even within the same species (Lohe *et al.* 1993). Second, many neocentromeres appear to be devoid of α-satellites. In one case, no sequence differences could be found that could distinguish neocentromeric from parental, noncentromeric DNA (Barry *et al.* 2000). Third, *in vitro* nucleosome reconstitution experiments with CENP-A fail to show a specificity for α-satellites (Yoda *et al.* 2000). These observations have led to the prevailing view that centromere formation is not dependent on the particular DNA sequence (Karpen and Allshire 1997).

Is centromeric sequence important at all? We have addressed this issue by investigating the evolution of centromeric H3-like proteins in closely related species of the *D. melanogaster* subgroup. We find strong evidence for the adaptive evolution of *Cid* in this subgroup. We propose that this adaptive evolution is driven in re-

Corresponding author: Steve Henikoff, 1100 Fairview Ave. N., A1-162, Seattle, WA 98109. E-mail: steveh@fhcrc.org

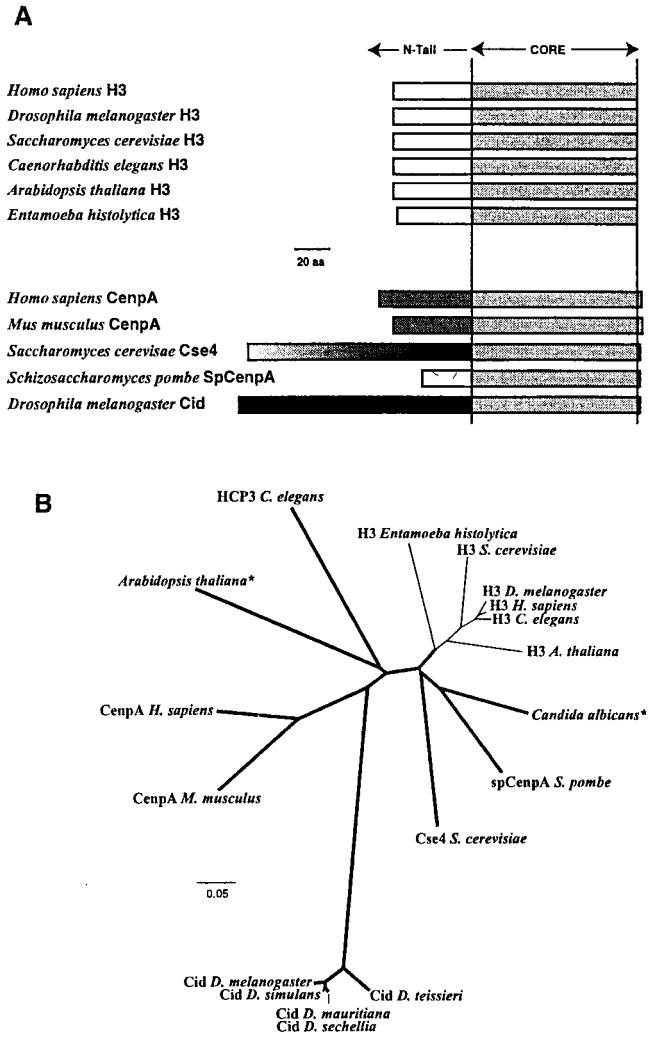

FIGURE 1.—The histone H3 family of proteins. This family consists of the typical, replication-dependent histone H3 proteins, as well as the atypical centromeric H3-like proteins. The latter are atypical in that their N-terminal tails bear little resemblance to those of the well-conserved histone H3 proteins, both in sequence similarity and in length, as shown in the schematic figure (A). Even in the nucleosomal core region, which can be well aligned across all members of the family, the centromeric H3-like proteins are more divergent as shown (B) in a neighbor-joining tree (SAITOU and NEI 1987). Asterisks mark those genes that have been putatively identified as being centromeric, on the basis of bioinformatic criteria, but not cytologically or biochemically. Figure modified and updated from HENIKOFF *et al.* (2000).

sponse to the rapidly changing satellite sequences at Drosophila centromeres. The evolution of *Cid* and Drosophila centromeric DNA may also have a direct impact on previously described meiotic drive processes involving chromosome orientation during meiosis (ZWICK *et al.* 1999).

MATERIALS AND METHODS

Fly strains: Various fly strains were obtained from the National Drosophila Species Resource Center (Bowling Green State University). Isofemale lines of *D. melanogaster* from Zimbabwe were kindly provided by Ying Chen (University of Rochester).

Cloning and sequencing: Flanking genes to *Cid* were used as queries in a BLASTP search and putative homologs identified in other metazoans. Primers were designed to conserved regions identified in the BLASTP search. PCR was done using primers to the upstream gene 5′-TCCGTGCCGAACAGCTC CGC-3′ (Cid-upstream) and to the *Cid* gene 5′-CTCGCG CACTAGACGCGAGAACGG-3′ (CidmidRev) for the 5′ half of the *Cid* gene. The 3′ half of the *Cid* gene was amplified using a *Cid*-specific primer, 5′-CCGGAGCCAGAAGACGGCACCG ACTACGG-3′ (CidmidFor) and a primer to the downstream gene, 5′-CTCGCTGCTGTTCCTCAACCAGTACTTCAA-3′ (Cid-downstream). Both strands of PCR products were sequenced in the *Cid* coding region, using ABI Big Dye sequencing. The sequences obtained have been deposited in GenBank under accession nos. AF321923–AF321926.

Tests for positive selection: The various sequences were aligned using CLUSTALX (THOMPSON *et al.* 1997) and analyzed for deviations from neutrality (McDonald-Kreitman test (MCDONALD and KREITMAN 1991) using DnaSP (ROZAS and ROZAS 1999). Other tests of deviations from the neutral hypothesis, *e.g.*, Hudson-Kreitman-Aguade test, Tajima's test, and Fu and Li's test) were also performed using DnaSP (the latter two tests do not reject neutrality).

RESULTS AND DISCUSSION

Adaptive evolution of Cid in the *melanogaster* subgroup: The single-copy gene *Cid* is 678 nucleotides long with no introns with the N-terminal tail encoded by bases 1–375, and the remainder (376–675) encoding the core-histone fold region (HENIKOFF *et al.* 2000). We determined the sequence of *Cid* from 15 strains of *D. melanogaster*, 8 strains of *D. simulans*, and 1 strain each from *D. mauritiana* and *D. sechellia* by PCR (see MATERIALS AND METHODS) and direct sequencing. We also sequenced *Cid* from *D. teissieri* and determined that it encoded a 231-amino-acid protein with no introns. Representative *Cid* sequences from the five species are shown aligned in Figure 2, along with the histone H3 sequences from *D. melanogaster* and *Entamoeba histolytica*, one of the earliest-branching eukaryotes. While the melanogaster subgroup is thought to be 2.3 million years old, the divergence between *D. melanogaster* and *E. histolytica* is likely to be greater than a billion years.

The polymorphic sites of the different strains of *D. melanogaster* and *D. simulans* (Figure 3) show an excess of replacement changes in the *melanogaster-simulans* comparison. To investigate whether these replacement changes have resulted from adaptive evolution, we separated all polymorphic sites into either replacement (R) or synonymous (S) sites and into either fixed differences between species or polymorphic within species. If mutations at synonymous and nonsynonymous sites were due to neutral mutations alone, the R:S ratio between species (divergence) would be expected to be the same as the R:S ratio within species (polymorphism). Significant deviations from this expectation are of two types: excess of replacement changes fixed between species is expected in the case of adaptive evolution (positive selec-

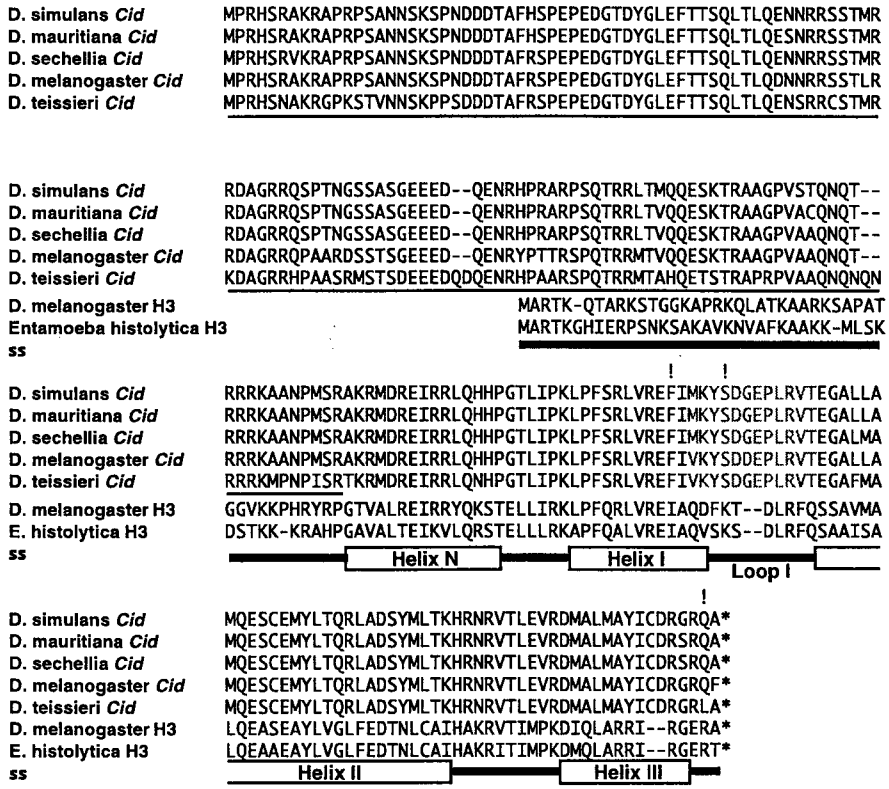

FIGURE 2.—Alignment of the Cid proteins from five species of the *melanogaster* subgroup with histone H3 proteins from *D. melanogaster* and *E. histolytica*. Various centromeric H3-like proteins are ~35–50% identical to histone H3, but even divergent histone H3 show >75% identity to each other. The thin vertical line separates the N-terminal tails of Cid from H3, as these are not similar to each other. A schematic secondary structure is also presented (modified from LUGER *et al.* 1997; SHELBY *et al.* 1997) with the core histone fold starting at Helix N. Loop I is between helices I and II. Fixed replacement changes in the core between *D. melanogaster* and *D. simulans* Cid are highlighted with a "!".

tion; MCDONALD and KREITMAN 1991), while an excess of replacement polymorphisms within species indicates the accumulation of deleterious mutations (NACHMAN *et al.* 1996).

We compared R:S ratios for interspecific fixed differences and intraspecific polymorphic sites in Table 1. Between *D. melanogaster* and *D. simulans*, the R:S ratio is 18:10, while the pooled polymorphic sites within the two species have an R:S ratio of 9:28. These ratios are different at highly significant levels ($P < 0.0025$) and demonstrate an excess of replacement changes fixed between species (18 instead of $9/28 \times 10 = \sim 3$ ex-

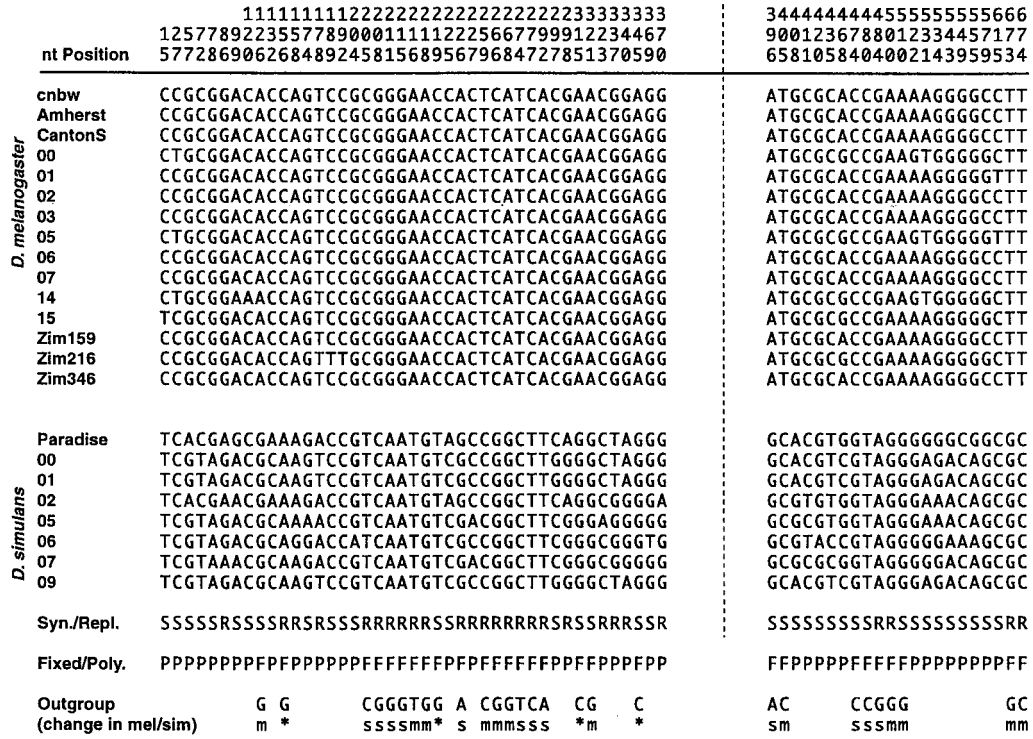

FIGURE 3.—Polymorphic positions among *D. melanogaster* (15 strains) and *D. simulans* (8 strains). A particular polymorphism is classified as being fixed (F) between species or polymorphic (P) if it is variable within either or both species. Sites are further separated into replacement (R) and synonymous (S) positions (see Table 1). Of the 28 fixed changes between *D. melanogaster* and *D. simulans*, 24 can be assigned to either lineage (m/s) by comparison with an outgroup, *D. teissieri*.

TABLE 1

McDonald-Kreitman test

	Fixed (interspecies)	Polymorphic (intraspecies)	Ratios	*G*-value*	*P*-value
		D. melanogaster and D. simulans (pooled)			
Replacement	14 (T) + 4 (C)	9 (T) + 0 (C)	14:5::9:15 (T)	5.58 (T)	0.0185
Synonymous	5 (T) + 5 (C)	15 (T) + 13 (C)	4:5::0:13 (C)	7.48 (C)	0.006
			18:10::9:28 (W)	10.43 (W)	0.0012
		D. melanogaster lineage alone			
Replacement	8	0	8:4::0:9 (W)		0.0006
Synonymous	4	9			
		D. simulans lineage alone			
Replacement	9	9	9:3::9:20 (W)		0.011
Synonymous	3	20			

The test is applied to the N-terminal tail (T), C-terminal core (C), and the whole *cid* gene (W) for *D. melanogaster* and *D. simulans*. *G-test of independence with Williams' correction for continuity was carried out (Sokal and Rohlf 1995). Tajima's and the Fu and Li's tests were also carried out but do not reject the neutral model ($P > 0.1$).

pected). Thus, we can reject the neutral mutation hypothesis for the entire *Cid* gene. Further, when the analysis is partitioned to just the N-terminal tail, or just the C-terminal core histone fold, we can still reject neutrality, implying that *both* the N-terminal tail and the C-terminal core have undergone positive selection. This is the first instance where positive selection has been documented in a structural protein that is both ubiquitous and essential in all eukaryotes.

All 9 replacement polymorphisms (Table 1) have occurred within strains of *D. simulans*, none within *D. melanogaster* (Figure 3). We investigated the possibility of disparate evolutionary pressures on *Cid* along the two lineages, one leading to *D. melanogaster* and the other to *D. simulans*, by comparisons using *D. teissieri* Cid as an outgroup (Figure 3). There are 28 fixed changes between *D. simulans* and *D. melanogaster*, of which 4 cannot be assigned to either lineage using *D. teissieri* (asterisks in Figure 3). Starting with the most recent common ancestor of *D. simulans* and *D. melanogaster*, 8 replacement and 4 synonymous changes are assigned to have taken place in the lineage leading to *D. melanogaster* (0 replacement and 9 synonymous polymorphisms within species) while 9 replacement and 3 synonymous changes occurred in the lineage leading to *D. simulans* (9 replacement and 20 synonymous polymorphisms within species). This indicates that adaptive evolution has occurred on *both* the *D. melanogaster* ($P < 10^{-3}$) and the *D. simulans* ($P = 0.01$) lineages since their split from a common ancestor.

With each adaptive substitution, polymorphisms are lost from the surrounding region. If such a selective sweep had occurred recently, then we should detect fewer polymorphisms than expected in a region around the adaptive site. We find that in *D. melanogaster*, the region from nucleotides 121–467 is devoid of any polymorphisms, but has at least five fixed replacement changes, which is consistent with a selective sweep. Using the Hudson-Aguade-Kreitman test, we find θ in region 1–120 and 468–675 = 0.0237 to be significantly different from θ in region 121–467 = 0.01 ($\chi^2 = 4.227$, $P = 0.04$; Hudson *et al.* 1987) consistent with a selective sweep recent enough that no synonymous polymorphisms have appeared in this window (121–467) since the sweep.

Where are the targets of positive selection? Nucleotide differences tabulated in Figure 3 are represented graphically in a sliding window analysis in Figure 4 as either polymorphisms within species for *D. simulans* (represented by π) or divergent sites between species (K). Two peaks of interspecific divergence are observed, consistent with positive selection in both the N-terminal tail and the core. *Cid*'s histone core sequence is similar to that of histone H3, and previous studies have suggested an identical topology of the centromeric H3-like proteins within the nucleosomal particle (Shelby *et al.* 1997; Yoda *et al.* 2000). We find it compelling that two out of the three amino acid replacements that are fixed between *D. melanogaster and D. simulans* in the core map at or adjoining the Loop 1 region (Figure 1A). Loop 1 from histone H3 has been shown in previous studies to make direct contacts with the DNA that is wrapped around the nucleosome (Arents *et al* 1991; Luger *et al.* 1997). The Loop 1 region in Cid and other centromeric H3-like molecules is typically two to three amino acids longer and may play a more significant role in the specificity of DNA binding (Shelby *et al.* 1997). Studies that have swapped the Loop 1 region between the centromeric histone H3-like proteins and histone H3 have confirmed the importance of this region for centromeric localization (Shelby *et al.* 1997) and chromosome segregation (Keith *et al.* 1999). Consistent with these previous results, we find that the signal of positive selection we have observed in the Cid core is in the region

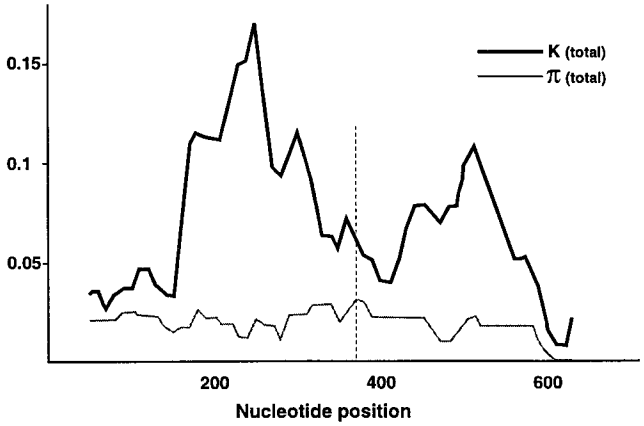

FIGURE 4.—A sliding window analysis [50-nucleotide (nt) window, steps of 10 nt] of the intraspecific polymorphism (within *D. simulans*) represented by π and the interspecific divergence (K) for *Cid* performed using all sites (synonymous and replacement) with the *x*-axis indicating nucleotide position and the *y*-axis indicating π or K. A majority of the divergent sites in the first peak are nonsynonymous, while a majority of those in the second peak (but not all) are synonymous changes. The dashed line separates the N-terminal tail region from the C-terminal core.

that mediates contacts with centromeric DNA. Since the N-terminal tail of Cid is not similar to those in other lineages of centromeric histones, it is not possible to map the observed amino acid replacement changes onto a predetermined structure. However, the hydrophilicity and location of the N-terminal tail as it exits between the two DNA gyres of the nucleosome are consistent with the hypothesis that the positively selected sites in the N-terminal tail may also mediate DNA binding.

What drives the adaptive evolution of *Cid*? Satellite changes can occur rapidly in species. This is because insertion of transposable elements can interrupt preexisting satellite repeat elements [such as L1 elements in human α-satellites (LAURENT *et al.* 1999) and various transposons in a Drosophila mini-chromosome centromere (LE *et al.* 1995)] and sequence-variant satellite repeats can arise and expand to fixation stochastically. For example, the 2.3-million-year divergence of the *melanogaster* subgroup has been sufficient to eliminate some satellite sequences or to significantly alter their abundance (LOHE and BRUTLAG 1987). Similarly, α-satellite repeats at centromeres of human and chimpanzee homologous chromosomes are not orthologous (JORGENSEN *et al.* 1992; HAAF and WILLARD 1999). To maintain centromeric function, Cid (or CENP-A etc.)-containing nucleosomes could be expected to recognize these centromeres in transition. Under this model, Cid is predicted to be constantly driven to keep pace with the rapidly changing satellite sequences at centromeres. In contrast, histone H3, among the most highly conserved proteins in eukaryotes, would be evolutionarily constrained by the necessity to interact with the entire genome.

Cid is the first centromere-associated protein whose pattern of variation parallels that of the satellite repeats found at centromeres. The 2.3-million-year evolutionary divergence of Cid is amplified when we compare the heterogeneity of the N-terminal tail of centromeric H3-like proteins in different lineages (Figure 1; HENIKOFF *et al.* 2000). Thus, N-terminal tails from Cse4, CENP-A, HCP3, and Cid show no recognizable similarity to each other, despite apparently carrying out analogous functions in their respective genomes.

Are the satellites themselves driven? In female meiosis, an opportunity for competition between centromeres can arise, since only one of the four meiotic products is included as a pronucleus in the oocyte. Any chromosomal element that improves the likelihood of inclusion into the oocyte will thus have a significant advantage (ZWICK *et al.* 1999). Among laboratory stocks of Drosophila, centromere strength can vary, influenced by the amount of flanking heterochromatin (NOVITSKI 1955; AULT and LYTTLE 1988). Since a satellite block can have innate centromeric competence (PLATERO *et al.* 1999), chromosomes with different satellite expansions may have differential centromeric strength, leading to meiotic competition in their orientation and propagation. Particular satellite variants may be preferred because of preferential binding by a host factor. Cid could be that host factor.

Any drive process is inherently deleterious because it can lead to takeover by unfit alleles as well as to lowered fertility. A satellite-centromere drive can be especially deleterious. Consider a situation in which a satellite sequence that can be bound by Cid expands near a centromere. By a process of stochastic expansion, this satellite could supplement the pre-existing centromeric satellite sequence, leading to an increase in centromere strength, and consequently a preferential retention of the chromosome in female meiosis. Left unchecked, this meiotic advantage could lead to runaway centromere expansions and chromosomal imbalances in mitosis. The host genome would prefer to restore balance among different chromosomes. Cid is the best candidate for alleviating this centromere-drive process by modulating binding to different satellite variants.

The observation that Cid is subject to adaptive evolution provides us the opportunity to test specific models about causes of satellite evolution (ZWICK *et al.* 1999) and Cid's potential role in mediating isolation between species, as species arrive at different Cid-satellite configurations. The conservation of the kinetochore apparatus and the generality of rapidly evolving centromeric satellites (CSINK and HENIKOFF 1998) predict that similar adaptive evolution will be found for centromeric H3-like proteins in other complex eukaryotes, including humans. Conversely, strong adaptive evolution is not expected for the centromeric H3-like Cse4 protein in

S. cerevisiae, which has simple centromeres. These possibilities can now be tested using methods that we have applied to the analysis of Cid evolution.

We thank C. H. Langley, W. Gilliland, members of our lab, and anonymous referees for helpful comments on the manuscript. We thank the Drosophila Species stock center and Ying Chen for the various Drosophila strains used in this study and Judith O'Brien for technical support. This work was supported by the Howard Hughes Medical Institute and a postdoctoral fellowship from the Helen Hay Whitney Foundation.

LITERATURE CITED

Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love and E. N. Moudrianakis, 1991 The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc. Natl. Acad. Sci. USA **88:** 10148–10152.

Ault, J. G., and T. W. Lyttle, 1988 A transmissible dicentric chromosome in Drosophila melanogaster. Chromosoma **97:** 71–79.

Barry, A. E., M. Bateman, E. V. Howman, M. R. Cancilla, K. M. Tainton *et al.*, 2000 The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification. Genome Res. **10:** 832–838.

Buchwitz, B. J., K. Ahmad, L. L. Moore, M. B. Roth and S. Henikoff, 1999 A histone-H3-like protein in *C. elegans.* Nature **401:** 547–548.

Copenhaver, G. P., K. Nickel, T. Kuromori, M. I. Benito, S. Kaul *et al.*, 1999 Genetic definition and sequence analysis of Arabidopsis centromeres. Science **286:** 2468–2474.

Csink, A. K., and S. Henikoff, 1998 Something from nothing: the evolution and utility of satellite repeats. Trends Genet. **14:** 200–204.

Haaf, T., and H. F. Willard, 1999 Chromosome-specific alpha-satellite DNA from the centromere of chimpanzee chromosome 4. Chromosoma **106:** 226–232.

Harrington, J. J., G. V. Bokkelen, R. W. Mays, K. Gustashaw and H. Willard, 1997 Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. Nat. Genet. **15:** 345–355.

Henikoff, S., K. Ahmad, J. S. Platero and B. vanSteensel, 2000 Heterochromatic deposition of centromeric histone H3-like proteins. Proc. Natl. Acad. Sci. USA **97:** 716–721.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Jorgensen, A. L., H. B. Laursen, C. Jones and A. L. Bak, 1992 Evolutionarily different alphoid repeat DNA on homologous chromosomes in human and chimpanzee. Proc. Natl. Acad. Sci. USA **89:** 3310–3314.

Karpen, G. H., and R. C. Allshire, 1997 The case for epigenetic effects on centromere identity and function. Trends Genet. **13:** 489–496.

Keith, K. C., R. E. Baker, Y. Chen, K. Harris, S. Stoler *et al.*, 1999 Analysis of primary structural determinants that distinguish the centromere-specific function of histone variant Cse4p from histone H3. Mol. Cell. Biol **19:** 6130–6139.

Laurent, A. M., J. Puechberty and G. Roizes, 1999 Hypothesis: for the worst and for the best, L1Hs retrotransposons actively participate in the evolution of the human centromeric alphoid sequences. Chromosome Res. **7:** 305–317.

Le, M. H., D. Duricka and G. H. Karpen, 1995 Islands of complex DNA are widespread in Drosophila centric heterochromatin. Genetics **141:** 283–303.

Lohe, A. R., and D. L. Brutlag, 1987 Identical satellite DNA sequences in sibling species of Drosophila. J. Mol. Biol **194:** 161–170.

Lohe, A. R., A. J. Hilliker and P. A. Roberts, 1993 Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics **134:** 1149–1174.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, 1997 Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389:** 251–260.

Masumoto, H., M. Ikeno, M. Nakano, T. Okazaki, B. Grimes *et al.*, 1998 Assay of centromere function using a human artificial chromosome. Chromosoma **107:** 406–416.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351:** 652–654.

Murphy, T. D., and G. H. Karpen, 1998 Centromeres take flight: alpha satellite and the quest for the human centromere. Cell **93:** 317–320.

Nachman, M. E., W. M. Brown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.

Novitski, E., 1955 Genetic measures of centromere activity in Drosophila melanogaster. J. Cell. Comp. Physiol. **45:** 151–169.

Palmer, D. K., K. O'Day and R. L. Margolis, 1990 The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. Chromosoma **100:** 32–36.

Palmer, D. K., K. O'Day, H. L. Trong, H. Charbonneau and R. L. Margolis, 1991 Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. USA **88:** 3734–3738.

Platero, J. S., K. Ahmad and S. Henikoff, 1999 A distal heterochromatic block displays centromeric activity when detached from a natural centromere. Mol. Cell **4:** 995–1004.

Round, E. K., S. K. Flowers and E. J. Richards, 1997 Arabidopsis thaliana centromere regions: genetic map positions and repetitive DNA structure. Genome Res. **7:** 1045–1053.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Shelby, R. D., O. Vafa and K. F. Sullivan, 1997 Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. J. Cell Biol. **136:** 501–513.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry,* Ed. 3. W. H. Freeman, New York.

Stoler, S., K. C. Keith, K. E. Curnick and M. Fitzgerald-Hayes, 1995 A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. **9:** 573–586.

Takahashi, K., E. S. Chen and M. Yanagida, 2000 Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science **288:** 2215–2219.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

Wiens, G. R., and P. K. Sorger, 1998 Centromeric chromatin and epigenetic effects in kinetochore assembly. Cell **93:** 313–316.

Willard, H. F., 1991 Evolution of alpha satellite. Curr. Opin. Genet. Dev. **1:** 509–514.

Yoda, K., S. Ando, S. Morishita, K. Houmura, K. Hashimoto *et al.*, 2000 Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc. Natl. Acad. Sci. USA **97:** 7266–7271.

Zwick, M. E., J. L. Salstrom and C. H. Langley, 1999 Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in Drosophila melanogaster. Genetics **152:** 1605–1614.

Communicating editor: J. A. Birchler