

# Distribution of Retroelements in Centromeres and Neocentromeres of Maize

Rebecca J. Mroczek\* and R. Kelly Dawe\*<sup>†,1</sup>

\*Department of Plant Biology and <sup>†</sup>Department of Genetics, University of Georgia, Athens, Georgia 30602

Manuscript received March 23, 2003

Accepted for publication June 9, 2003

## ABSTRACT

Fluorescent *in situ* hybridization was used to examine the distribution of six abundant long terminal repeat (LTR) retroelements, *Opie*, *Huck*, *Cinful-1*, *Prem-2/Ji*, *Grande*, and *Tekay/Prem-1* on maize pachytene chromosomes. Retroelement staining in euchromatin was remarkably uniform, even when we included the structurally polymorphic abnormal chromosome 10 (Ab10) in our analysis. This uniformity made it possible to use euchromatin as a control for quantitative staining intensity measurements in other regions of the genome. The data show that knobs, known to function as facultative neocentromeres when Ab10 is present, tend to exclude retroelements. A notable exception is *Cinful-1*, which accumulates in TR-1 knob arrays. Staining for each of the six retroelements was also substantially reduced in centromeric satellite arrays to an average of 30% of the staining in euchromatin. This contrasted with two previously described centromere-specific retrotransposable (CR) elements that were readily detected in centromeres. We suggest that retroelements are relatively rare in centromeres because they interrupt the long satellite arrays thought to be required for efficient centromere function. CR elements may have evolved mutualistic relationships with their plant hosts: they are known to interact with the kinetochore protein CENH3 and appear to accumulate in clusters, leaving long satellite arrays intact.

**T**RANSPOSABLE elements are divided into two major groups: class I elements, which transpose via an RNA intermediate, and class II elements, which transpose through DNA replication. Class I transposable elements include LINES, SINES, and long terminal repeat (LTR) retroelements. The latter occupy a significant portion of large plant genomes like that of maize, where intergenic regions are composed primarily of nested LTR retrotransposons (SANMIGUEL *et al.* 1996; SANMIGUEL and BENNETZEN 1998). Roughly 50% of the maize, rye, barley, and wheat genomes are thought to be composed of LTR retroelements (SANMIGUEL *et al.* 1996; PEARCE *et al.* 1997; SANMIGUEL and BENNETZEN 1998; VINCENT *et al.* 1999; MEYERS *et al.* 2001; WICKER *et al.* 2001). The same variety of transposable elements appears to exist in smaller-genome species such as rice and Arabidopsis, but fewer representatives of each class are present. Only ~14% of the Arabidopsis genome is composed of transposable elements (ARABIDOPSIS GENOME INITIATIVE 2000). The fact that genome size varies greatly while gene number varies little (the *C*-value paradox; CAVALIER-SMITH 1978; PAGEL and JOHNSTONE 1992) can be largely attributed to extraordinary variation in the number of retroelements (KUMAR and BENNETZEN 1999; TIKHONOV *et al.* 1999; WICKER *et al.* 2001; BENNETZEN 2002).

LTR retroelements are separated into two major

groups on the basis of the organization of the domains within their *Pol* genes: in the Ty1/*copia*-like group the integrase (INT) domain is located upstream of the reverse transcriptase (RT) domain, while in the Ty3/*gypsy*-like group the INT domain is located downstream of the RT domain (XIONG and EICKBUSH 1990). The two groups can be further subdivided into families on the basis of the similarity of their LTR sequences, which evolve faster than the internal coding regions (SANMIGUEL and BENNETZEN 1998). The distribution of *gypsy*- and *copia*-like LTR retroelements has been analyzed in a number of large genome plants via fluorescent *in situ* hybridization (FISH; PEARCE *et al.* 1996a,b; BRANDES *et al.* 1997; KUMAR *et al.* 1997; FRIESEN *et al.* 2001). The patterns of localization suggest that retroelements have insertional preferences. Ty1/*copia* retroelements are found throughout the euchromatin of *Vicia faba* (PEARCE *et al.* 1996a) but are generally concentrated in the subtelomeric heterochromatin of *Allium cepa* (KUMAR *et al.* 1997). There is also evidence from *Allium* and gymnosperms that individual retroelement families have discernibly different patterns of chromosomal localization (PICH and SCHUBERT 1998; FRIESEN *et al.* 2001).

Although most retroelements are distributed nonrandomly throughout chromosomes, the most obvious discontinuities occur with respect to tandem repeat arrays. Brandes and co-workers (1997) examined a variety of organisms (*A. cepa*, *Beta vulgaris*, *Brassica campestris*, *Brassica oleracea*, *Pennisetum glaucum*, *Pinus elliotii*, *Selaginella apoda*, *V. faba*, and *V. narbonensis*) and demonstrated that the Ty1/*copia* group is dispersed throughout the euchromatic regions, but absent from regions where

<sup>1</sup>Corresponding author: Department of Plant Biology, Miller Plant Sciences Bldg., University of Georgia, Athens, GA 30602.  
E-mail: kelly@dogwood.botany.uga.edu

specialized tandem repeats are expected to lie, such as centromeres, telomeres, heterochromatin, and the nucleolus organizing region (NOR). FISH analyses in maize using portions of the *Opie* and *Prem-2/Ji* retroelements showed diffuse patterns of staining with reduced accumulation at the centromeres and NOR (EDWARDS *et al.* 1996; ANANIEV *et al.* 1998a; MILLER *et al.* 1998). In marked contrast are the *gypsy*-like centromeric retrotransposable (CR) elements of cereal grains, which accumulate specifically in centromeric satellites (JIANG *et al.* 1996; PRESTING *et al.* 1998; HUDAKOVA *et al.* 2001; ZHONG *et al.* 2002; NAGAKI *et al.* 2003a). Recent data indicate that the maize CR elements interact with the kinetochore protein centromeric histone H3 (CENH3; ZHONG *et al.* 2002), suggesting that they participate in centromere function.

Here we describe and quantify the accumulation patterns of a variety of maize retroelement families from both the Ty3/*Gypsy* and Ty1/*Copia* groups, including two different types of maize CR elements. We find that both the *Gypsy* and *Copia* groups are found throughout the euchromatic portions of the genome. In contrast, all but one of the retroelements analyzed are under-represented in knobs, which are known to function as facultative centromeres. Additionally, all of the retroelements outside of the CR clade are largely excluded from centromeric satellite arrays. The data suggest that centromeric satellite arrays are under selection for their function in chromosome movement, much like genic regions (SANMIGUEL *et al.* 1996).

## MATERIALS AND METHODS

**Maize stocks:** The standard maize inbred lines W23 and KYS were used for the bulk of the cytological analysis. The strain containing abnormal chromosome 10 (Ab10) was originally obtained from Marcus Rhoades and subsequently backcrossed into the W23 background seven times.

**Phylogenetic analysis:** Nucleotide sequences were retrieved from GenBank for the following maize retroelements: *Grande* 1-4 (GenBank X97604), *Huck* (AF391808-1), *Tekay* (AF050-455), *Fourf* (AF050436), *Mare 5* (AB033252.1), *Reina* (U69258), *Rle* (AF057037), *Cinful-1* (AF049110), *Cinful-2* (AF049111), *Prem-2* (U41000), *Opie-2* (U68408), *Cent-A* (AF078917), CRM (AY129008), CRR (AC022352), *Cereba* (AY040832), and two *Arabidopsis thaliana* retroelement sequences (AAD11616 and BAB40826). The reverse transcriptase regions were identified following the guidelines set forth by XIONG and EICKBUSH (1990). A progressive alignment of the RT regions was prepared using the Pileup option in GCG's SeqLab with a BLOSUM 30 transition matrix (GCG 1982–2000; FENG and DOOLITTLE 1987; HENIKOFF and HENIKOFF 1992). The alignments were adjusted manually using previously published RT alignments as visual templates (XIONG and EICKBUSH 1988; BOWEN and McDONALD 2001). Both parsimony heuristics and neighbor joining (NJ) were used to generate trees from our alignment using PAUP\* (SWOFFORD 1999). The NJ trees were derived using uncorrected pairwise distances (SWOFFORD 1999), and bootstrap values were determined using PAUP\*.

**Probe preparation:** Primers specific to LTRs were designed from the sequences noted above with the exceptions of *Huck*

(AF050438) and *Grande* (AF050437). The lengths of the amplified products and the primers used for the amplification are: *Cinful-1*, 562 bp (F-5'-CGCCGAAGTCTTCTAGGAA-3', R-5'-GGAGACTCGTTCTCAAGTGCTA-3'); *Grande*, 350 bp (F-5'-ATGCGAGGATAAGTCGGCGAAG-3', R-5'-GGTGTTTTATAGGAGTAGGACGGTG-3'); *Huck*, 673 bp (F-5'-TCCACTGACC GACCTGACAA-3', R-5'-GGTTTTGGCACCCTGTTTCAT-3'); *Opie-2*, 526 bp (F-5'-CAAACACAAGTGCTTAAAT-3', R-5'-GTC CCGTGCCCCGATTTGT-3'); *Prem-2/Ji*, 573 bp (F-5'-ACATTT GGTGTTGGGGCTA-3', R-5'-GGGTGAATAGGGCCGAAC TGAA-3'); *Tekay*, 537 bp (F-5'-ATTTGTGCGACCGCTCAA-3', R-5'-AGGAGTCCAGGCTGCTCTTA-3'); *Cent-A*, 1234 bp (F-5'-CATAACCCGCACAGATATGAC-3', R-5'-ATAAACCCAACG GGTAGAAGGG-3'); and CRM, 513 bp (F-5'-TCGTCAACT CAACCATCAGGTGAT-3', R-5'-GCAAGTAGCGAGAGCTAA ACTTGA-3'). The PCR fragments were each cloned into the TOPO cloning vector (Invitrogen, Carlsbad, CA) and verified by sequencing.

**In situ hybridization:** Anthers from maize inbred lines were fixed as previously described (HIATT *et al.* 2002). Probes were amplified from plasmids and labeled with FITC using a random primer labeling kit (Prime-It Fluor fluorescence labeling kit, Stratagene, La Jolla, CA). The CentC and TR-1 repeats were detected using rhodamine-labeled oligonucleotides specific to each sequence (HIATT *et al.* 2002; ZHONG *et al.* 2002). Chromosome straightening and *in situ* hybridization were carried out as previously described (DAWE *et al.* 1994; ZHONG *et al.* 2002). All data were collected and analyzed using a DeltaVision 3D light microscope workstation and associated software. Intensity data were collected by first selecting 9 × 9, two-dimensional "pixel boxes." From these boxes we recorded the total intensity values (amount of light detected by the CCD camera in that region) from the DNA 4',6-diamidino-2-phenylindole (DAPI) and retroelement (FITC) channels. Background, euchromatin, centromere, and knob readings were taken in each channel for each cell. Background measurements were averaged from 5 different pixel boxes and euchromatin measurements were averaged from 10 different pixel boxes. For knobs and centromeres the number of pixel boxes for which data were collected was limited by the number of centromeres or knobs that were unobstructed by chromosome arms; between 4 and 7 pixel boxes were taken for each structure in each cell. The pixel box data from each channel were averaged, and appropriate background levels were subtracted for each cell. The overall intensity values varied from cell to cell, as is typical for FISH experiments. We therefore calculated within-cell ratios of centromere/euchromatin or knob/euchromatin staining intensities for both channels and compared them. By taking advantage of an internal control (euchromatin staining), we effectively normalized the data, making it possible to average the information from different cells.

## RESULTS

**Phylogeny of the retroelement families analyzed:** We identified the RT regions of the retroelements examined in this report and compared their sequences to several other retroelements phylogenetically. Neighbor-joining and parsimony analysis of the retroelement RT regions produced identical trees. Figure 1 is a neighbor-joining tree, which shows that the retroelements can be divided into two major clades representing the *gypsy*- and *copia*-like groups of the LTR retroelements, a finding consistent with other phylogenetic analyses (XIONG and

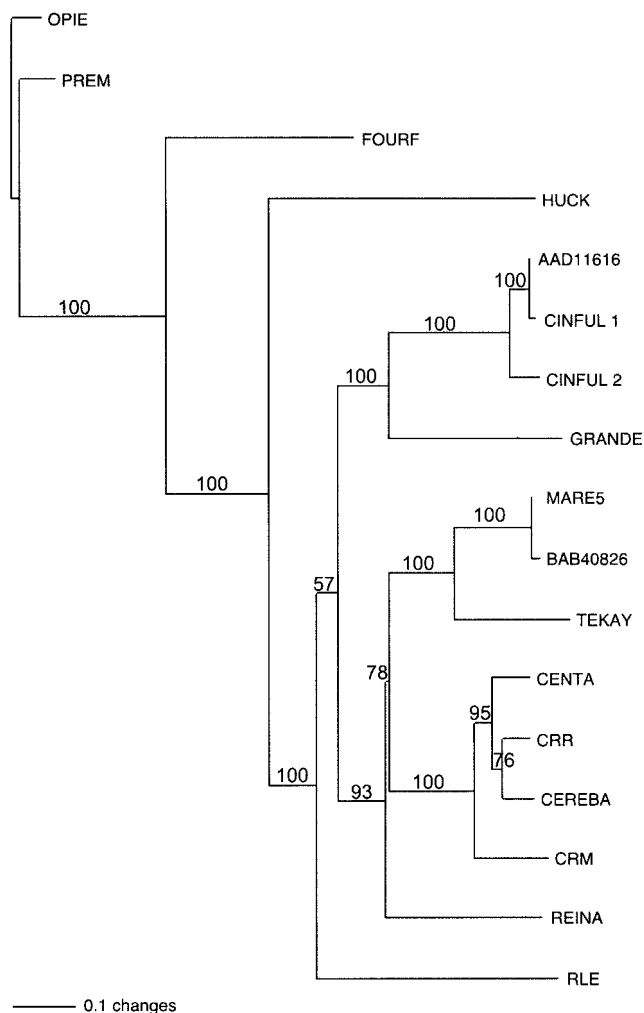


FIGURE 1.—Neighbor-joining tree of the RT amino acid sequences from a variety of plant retroelements. AAD11616 and BAB40826 represent RT sequences from *Arabidopsis*. CRM and *Cent-A* are CR elements from maize. CRR and *Cereba* are CR elements from rice and barley, respectively. All other retroelements are from maize. Numbers at the nodes represent the bootstrap values obtained using PAUP\*.

EICKBUSH 1990; MALIK and EICKBUSH 1999; BOWEN and McDONALD 2001). The known CR elements group together to form a monophyletic clade within the *gypsy*-like group (LANGDON *et al.* 2000; NAGAKI *et al.* 2003a). In maize there are two different CR elements, CRM and *Cent-A*, which have closely related internal regions (67% nucleotide similarity) but different LTRs (ANANIEV *et al.* 1998a; ZHONG *et al.* 2002). Among the known maize retroelements, *Tekay/Prem-1* and *Reina* are the most closely related to the CR elements. However, only *Tekay/Prem-1* is abundant enough (SANMIGUEL and BENNETZEN 1998; MEYERS *et al.* 2001) to be readily detected by FISH.

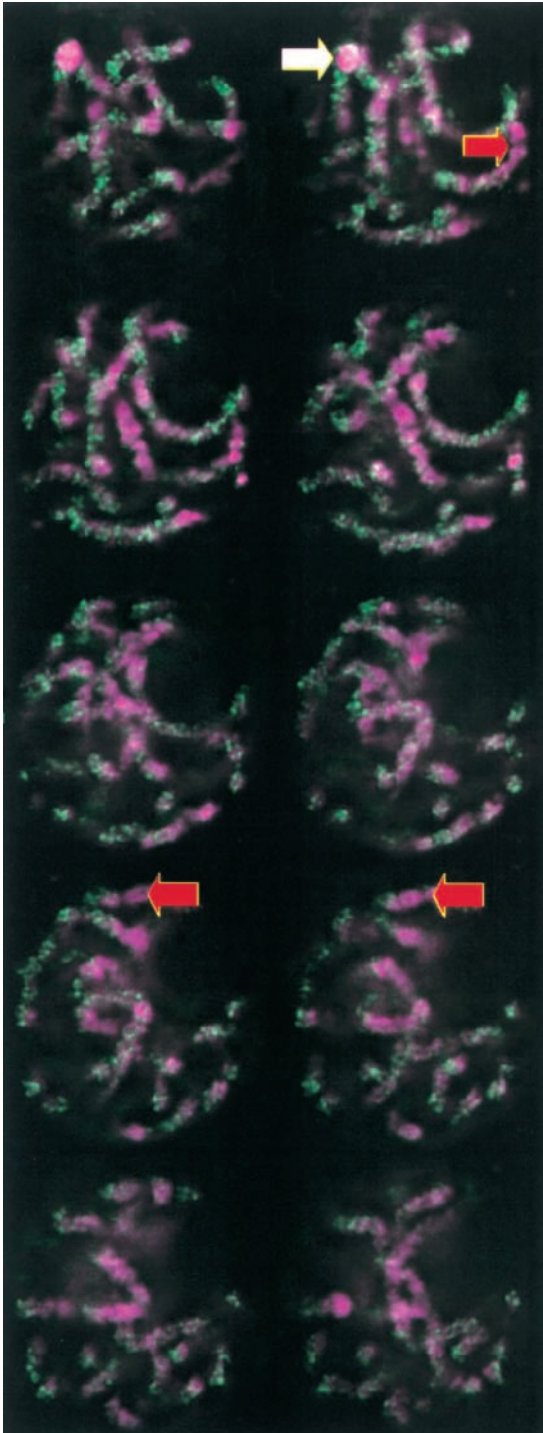
**The abundant maize retroelement families show genome-wide distributions:** We chose six of the most abundant maize retrotransposon families for *in situ* hybridization analysis: *Huck*, *Opie*, *Grande*, *Prem-2/Ji*, *Cinful-1*, and

*Tekay/Prem-1* (SANMIGUEL *et al.* 1996; SANMIGUEL and BENNETZEN 1998; MEYERS *et al.* 2001). Probes were generated to portions of the LTRs that are unique to individual families. These LTR fragments were then fluorescently labeled and hybridized to maize pachytene chromosome preparations, and the results were analyzed using deconvolution 3D light microscopy. Each of the retroelement families outside of the CR clade showed roughly uniform euchromatic staining patterns throughout the genome (Figures 2 and 3), with subtle differences among families. *Opie* is evenly distributed in a pattern of neat dots along the chromosomes (Figure 3), *Huck* has a patchy distribution (Figure 3), and *Prem-2/Ji* has a genome-wide distribution but on some chromosomes is underrepresented in pericentromeric heterochromatin (Figure 2, arrows; Figure 3).

**Retroelement abundance in euchromatic regions is relatively uniform but varies in knobs:** Visual inspection of the images in Figures 2 and 3 indicated a remarkable uniformity in the distribution of retroelements along chromosome arms, suggesting that euchromatin staining might serve as a suitable internal control for the staining in centromeric regions. To test this idea, we compared the euchromatin on Ab10 to other chromosome arms. Ab10 is an alternative version of the normal chromosome 10 (N10) that is present in ~10% of teosinte (the ancestor of maize) and ~2% of known maize strains (KATO 1976). The terminal portion of the long arm of Ab10 is responsible for the phenomena of neocentromere activity and meiotic drive. This region contains few essential genes (HIATT and DAWE 2003) and an ~14 map unit inversion (RHOADES and DEMPSEY 1985). In addition, there are three small knobs containing arrays of a 350-bp TR-1 knob repeat and a large knob composed primarily of a 180-bp repeat (PEACOCK *et al.* 1981; RHOADES and DEMPSEY 1985; ANANIEV *et al.* 1998b; HIATT *et al.* 2002). The striking structural polymorphism between Ab10 and N10, most notably the inverted region, is thought to be responsible for the fact that recombination between the distal portion of Ab10 and N10 rarely occurs (KIKUDOME 1959; RHOADES and DEMPSEY 1985). The unusual structural features and evolutionary history of Ab10 suggested that it might have an unusual distribution or abundance of retroelements.

We assayed the localization patterns of a sample of retroelements (*Huck*, *Prem-2/Ji*, and *Cinful-1*) on the distal portion of the Ab10 chromosome. As can be seen in Figure 4, retroelement staining in the euchromatic portion of the distal region of Ab10 was nearly identical to the intensity and pattern of staining observed for the rest of the genome. To quantify this observation we took advantage of the fact that chromatin (DAPI) and retroelement (FITC) staining are measured and stored as separate images during data collection and that deconvolution microscopy is quantitative. Staining intensity readings were taken for both chromatin and retro-

elements and compared to the staining intensities found in other euchromatic regions. We found no significant difference between Ab10 and the rest of the genome (*t*-test,  $P < 0.01$ ). Although the DAPI intensity in Ab10 euchromatin was relatively low compared to the genome as a whole (mean = 0.73, SD  $\pm 0.53$ ), retro-element intensities were similarly reduced (mean = 0.77, SD  $\pm 0.21$ ). These data support the view that retroelements are spread evenly and uniformly throughout maize euchromatin.



Our analysis of Ab10 also revealed that euchromatin and knobs stain differently for retroelements. For instance, *Huck* is nearly absent from the large knob of Ab10 (Figure 4C), and *Cinful-1* is highly abundant within the TR-1-containing chromomeres (indicated by staining that appears yellow in Figure 4D). As described below, these observations were pursued in more detail by analyzing a variety of other knobs in the genome.

**Retroelement families are variably interspersed in maize knob satellite DNA:** The TR-1 and 180-bp repeats present on Ab10 also occur at 22 other knob loci in differing proportions (KATO 1976; ANANIEV *et al.* 1998c; BUCKLER *et al.* 1999; HIATT *et al.* 2002). Many other classes of repeats may also be present in maize knobs. To obtain a more general perspective on knobs without a bias toward particular repeats, we scored retroelement staining in knobs as identified by their characteristic ball-shaped heterochromatic structure (Figure 5; see also Figures 2 and 3). We found that all families except *Huck* showed some staining in nearly every knob observed (Figure 5). The retroelement families *Cinful-1*, *Grande*, *Tekay/Prem-1*, *Opie*, and *Prem-2/Ji* all showed staining in the majority of knobs examined (84.8, 86.7, 87.1, 94.9, and 96.3%, respectively). In contrast, the *Huck* retroelement family was nearly absent in most knobs with some staining in only 20% of the knobs examined (Table 1). Retroelements from the *Huck* family apparently avoid or are selectively removed from knobs even though they occupy  $\sim 10\%$  of the maize genome (SANMIGUEL and BENNETZEN 1998; MEYERS *et al.* 2001). Our observations are consistent with the findings of ANANIEV and co-workers (1998b), who found no copies of the *Huck* retroelement family in any of 23 cloned knob segments.

Probes for retroelements outside of the *Huck* and CR families appeared to stain euchromatin and knobs at about the same intensity (Figure 5). Since knobs stain with DAPI very brightly, this observation implied a relatively low abundance of retroelements within the knobs. Intensity measurements confirmed the interpretation: we found that knobs were 2.6 times brighter in the DAPI channel than an average segment of euchromatin, but that retroelement staining within knobs was only 1.1 times brighter (Figure 6). In the small percentage of knobs that showed *Huck* staining, retroelement staining intensities were only 15% of the levels in euchromatin. These data indicate that although most retroelement families are present in the satellite repeats of knobs,

FIGURE 2.—Genome-wide distribution of the *Prem-2/Ji* retroelement family in maize. Images are single optical sections taken from a maize pachytene meiotic section, separated by 4  $\mu\text{m}$ . DNA is represented in magenta and the *Prem-2/Ji* retroelement family is represented in green. Absence of *Prem-2/Ji* staining at centromeres is indicated with red arrowheads. *Prem-2/Ji* staining at a knob is shown with a white arrowhead.

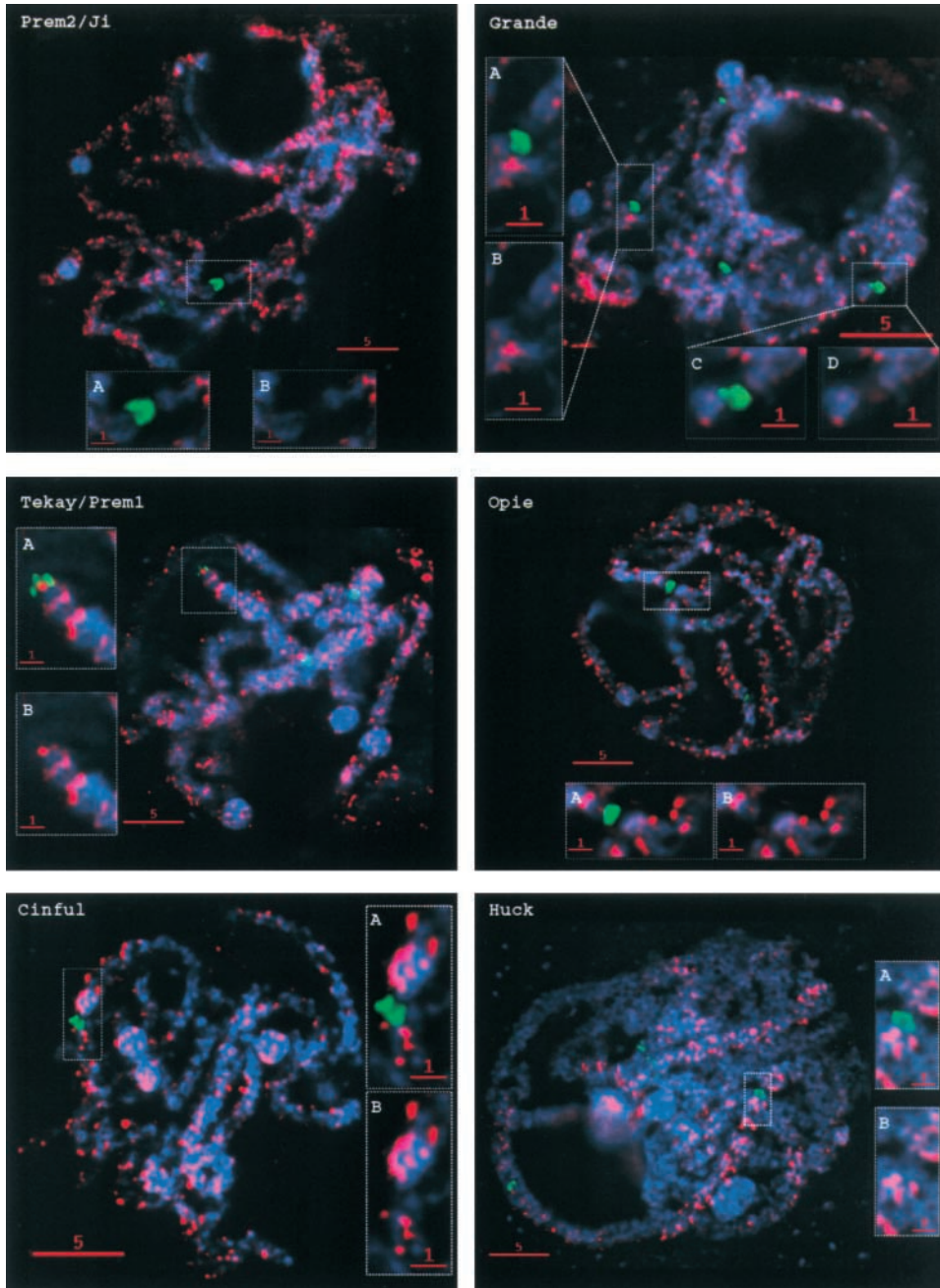


FIGURE 3.—Staining patterns of maize retroelement families throughout the genome with respect to the centromeres. Each image shows a single optical section of a cell at pachytene. DNA is shown in blue, CentC in green, and retroelement LTR staining in red. (A and C) Enlargements (2 $\times$ ) of the boxed areas showing retroelement staining around the centromere. (B and D) Same images as A and C, except with the CentC staining removed.

they are present at a reduced frequency when compared to euchromatin.

Data from the Ab10 chromosome indicated that *Cinful-1* retroelements accumulate in TR-1 knobs (Figure 4D). To determine if *Cinful-1* accumulation was limited to Ab10 knobs, we took intensity readings from the TR-1 knobs of Ab10 as well as three other TR-1-containing knobs. Overall, *Cinful-1* staining was 2.2 times higher in TR-1 knobs than in euchromatin (Figure 6), and there was no significant difference between the TR-1 knobs on Ab10 and those elsewhere in the genome (*t*-test,  $P < 0.01$ ). In contrast, the staining intensities for *Huck* and *Prem-2/ji* in TR-1 knobs were similar to those in non-TR-1-containing knobs. The fact that *Cinful-1* is substan-

tially overrepresented in TR-1 arrays provides evidence that the reduced staining we detect for all the other retroelements is not a consequence of the heterochromatic nature of knobs.

**Abundant retroelement families are largely absent from centromeres:** Maize centromeres contain tandem arrays of the 156-bp Cent-C satellite repeat, interspersed with the CR elements *Cent-A* and CRM (ZHONG *et al.* 2002; NAGAKI *et al.* 2003a). Noncentromere-specific retroelements are also present to some extent, although the overall frequency of these elements in centromeres is not known (ANANIEV *et al.* 1998a; NAGAKI *et al.* 2003a) and will be difficult to determine by sequence analysis, given the inherent limitations associated with cloning

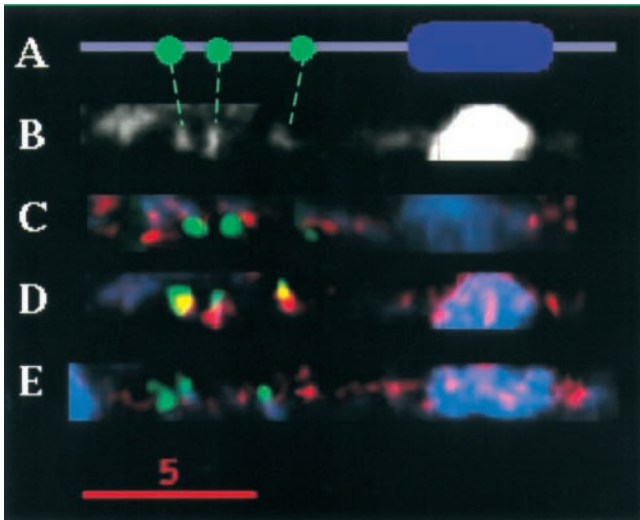


FIGURE 4.—Staining patterns of retroelements in the polymorphic portion of Ab10. (A) Graphical representation of the Ab10 chromosome showing the small TR-1 knobs in green, the large 180-bp repeat knob in blue, and the euchromatic regions in purple. (B) Black and white image of a computationally straightened Ab10 chromosome showing only DNA (DAPI) staining; chromomeres are indicated with connecting lines to the graphical representation in A. (C) *Huck* staining on a computationally straightened Ab10 pachytene chromosome. TR-1 staining is in green and retroelement staining is in red. (D) Straightened chromosome from a cell stained with *Cinfu1-1* in red and TR-1 in green. Intense *Cinfu1* staining at the TR-1 knobs makes them appear yellow, due to the overlap of the red and green colors. (E) Straightened chromosome showing *Prem-2/Ji* in red and TR-1 in green.

and contigging long repeat arrays (SONG *et al.* 2001; HENIKOFF 2002). In an effort to further examine and quantify the accumulation patterns of retroelements in centromeres, we labeled the major retroelement LTRs and Cent-C with different fluorescent dyes and analyzed the results on DAPI-stained maize pachytene chromosomes.

We found that each of the retroelement families, *Huck*, *Opie*, *Grande*, *Prem-2/Ji*, *Cinfu1-1*, and *Tekay/Prem-1*, are poorly represented at centromeres (Figures 3 and 6 and Table 1). The percentage of centromeres with detectable staining for these retroelements ranged from 3.4 to 37%, depending on the family (Table 1). Similarly, while DAPI-staining intensities were roughly equivalent in centromeres and chromosome arms, non-CR retroelement staining in centromeres averaged only 30% of the levels found in euchromatin (Figure 6).

In marked contrast to other maize retroelement families, *in situ* hybridization with the CR elements *Cent-A* (ANANIEV *et al.* 1998a) and CRM (ZHONG *et al.* 2002; NAGAKI *et al.* 2003a) revealed strict centromeric localization on all 10 maize chromosomes (Figure 7, Table 1). We often observed overlap of the CentC and CR signals (Figure 7, CRM), and in many cases the two signals were clearly separate (Figure 7, *Cent-A*). These data are consistent with fiber-FISH data from rice, which suggest

that CR elements tend to insert into large clusters distinct from the regions composed mainly of satellite repeats (CHENG *et al.* 2002).

## DISCUSSION

In this study we provide a perspective on the distribution of the most abundant retroelements in maize with particular emphasis on the centromeres and neocentromeres. Previous reports have used general reverse transcriptase probes to examine overall distribution patterns of retroelements (PEARCE *et al.* 1996a,b, 1997; BRANDES *et al.* 1997; KUMAR *et al.* 1997) or have localized specific families without a special emphasis on centromeres (PICH AND SCHUBERT 1998; FRIESEN *et al.* 2001). Retroelement distribution is also being analyzed in the sequenced plant genomes of Arabidopsis and rice (ARABIDOPSIS GENOME INITIATIVE 2000; FENG *et al.* 2002). However, sequence data are generally unreliable in regions containing long satellite arrays (HENIKOFF 2002). Even the size of most eukaryotic centromeres is still under debate (HAUPT *et al.* 2001; HOSOUCHI *et al.* 2002). The FISH strategy employed here has a lower resolution than DNA sequencing, but can provide a general and quantitative perspective on the frequency of retroelements in large centromeres. By focusing on a single organism and using retroelement family-specific probes and quantitative light microscopy, we have been able to draw new conclusions and hypotheses about the forces driving retrotransposon accumulation in large-genome species.

**Chromosomal localization with respect to evolutionary history of the elements:** The families examined represent both the *gypsy*- and *copial*-like groups of LTR retroelements. *Huck*, *Cinfu1-1*, *Tekay/Prem-1*, *Grande*, *Cent-A*, and CRM belong to the *gypsy*-like group and *Prem-2/Ji* and *Opie* fall into the *copial*-like group. In our analysis of eight retroelement families from maize we saw no obvious correlation between the type of retroelement (*gypsy*- or *copial*-like) and chromosomal localization patterns. The rapid evolution of localization patterns is particularly evident with regard to the CR elements and the closely related *Tekay/Prem-1* retroelement family. Although *Tekay/Prem-1* shares a more recent common ancestor with the CR elements than do the other families examined here, it is no more likely to be found in or near the centromere than are more distantly related families. As can be seen in Table 1 and Figure 6, the *Cinfu1-1* and *Opie* families are just as likely to show centromeric staining as the *Tekay/Prem-1* family is. The available data suggest that the strict centromeric localization pattern and apparent function of CR elements in recruiting centromeric histone are recently evolved features primarily limited to the cereal grains.

**Accumulation of retroelements in Ab10 and the satellite repeats of maize knobs:** The abnormal 10 chromosome of maize provides all knobs in the genome with the

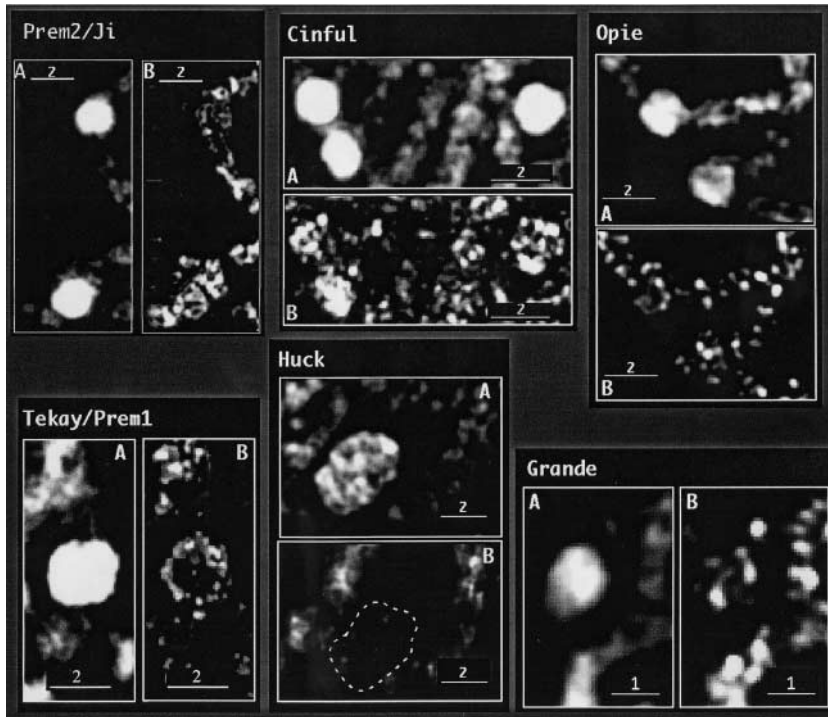


FIGURE 5.—Patterns of retroelement staining at knobs. Enlargements of knobs are from the cells shown in Figure 3, although some are from different optical sections. (A) Image of DNA showing intensely staining heterochromatic knobs. (B) Same image showing only the FITC (retroelement) staining. Note that *Huck* staining at the knob is markedly reduced.

capacity to move as neocentromeres and preferentially segregate to progeny (RHOADES 1942). Within the terminal region of the long arm of Ab10 is an inversion of chromatin from normal 10, three small knobs composed primarily of TR-1 repeats, a large knob composed primarily of 180-bp repeats, and regions of apparently novel chromatin (HIATT *et al.* 2002). Genetic data suggest that there are few essential genes in the latter half of this large structural polymorphism (HIATT and DAWE 2003). Meiotic drive systems may spread to some extent without regard to organismal fitness and are known to accumulate deleterious mutations (ARDLIE 1998). Similarly, retroelements have been shown to accumulate in *Drosophila* inversions, presumably because recombination events that would eliminate them are reduced there

(SNEGOWSKI and CHARLESWORTH 1994). Because of its association with meiotic drive, reduced recombination, and relative lack of genes, we considered whether the distal portion of Ab10 would be a favored spot for the accumulation of retroelements. However, outside of the *Cinful-1* family, which accumulates in all TR-1 arrays, the retroelement families we examined appear to be no more abundant on this chromosome than they are throughout the rest of the genome (Figure 4). These data suggest that the reduced recombination and meiotic drive typical of the long arm of Ab10 have had little impact on retroelement distribution over the time span in which the drive system has existed (see BUCKLER *et al.* 1999).

The remarkable uniformity of retroelement distribu-

TABLE 1  
Presence of retroelements at centromeres and knobs

Retroelement	Total centromeres counted	% centromeres with detectable staining	Total knobs counted	% knobs with detectable staining <sup>a</sup>
Opie	19	26.3	39	94.9
Prem-2/Ji	58	3.4	80	96.3
Cinful-1	27	37.0	46	84.8
Grande	13	15.4	15	86.7
Tekay/Prem-1	31	22.6	31	87.1
Huck	30	6.7	40	20.0
Cent-A	32	100.0	<sup>b</sup>	<sup>b</sup>
CRM	60	100.0	<sup>b</sup>	<sup>b</sup>

<sup>a</sup> Number of knobs observed for which retroelement staining was at least as intense as that observed throughout the euchromatin of the chromosomes.

<sup>b</sup> Cent-A and CRM are not detected at knobs or euchromatin.

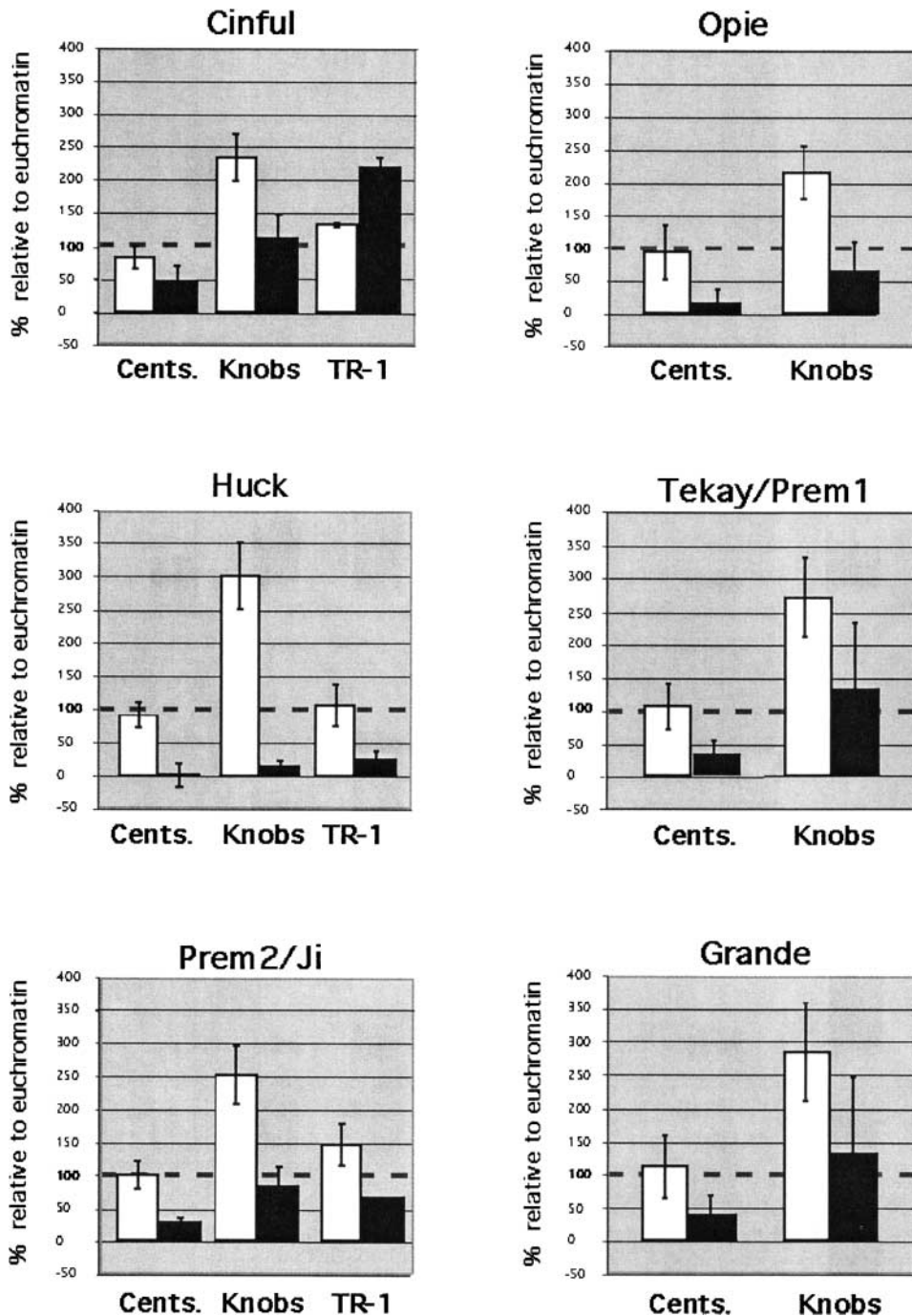


FIGURE 6.—Staining intensities of (□) DNA and (■) retroelements in centromeres, cytologically defined knobs, and TR-1 arrays. Staining intensities are expressed relative to euchromatin. Values below the dashed line indicate relatively low staining intensities; values above the line indicate high staining intensities. Vertical lines indicate the standard deviation.

tion in maize euchromatin gave us a useful internal control for the staining of retroelements in knobs and centromeres. The analysis indicates that retrotransposons are substantially underrepresented in knobs relative to euchromatin (Figures 4 and 6). One explanation for the underrepresentation of retroelements in knobs is that the structural organization of knob heterochromatin is such that FISH probes cannot gain access. However, the fact that we detected an abundance of *Cinful-1* elements in TR-1 arrays strongly suggests that knob structure did not serve as a substantial barrier to FISH probes. In addition, we have corroborated a previous

conclusion, on the basis of the analysis of cosmid clones, that *Huck* is underrepresented in knobs (ANANIEV *et al.* 1998b). We believe that our data provide an accurate view of the prevalence of retroelements in knob repeat arrays.

The fact that retroelements are underrepresented in knobs indicates that long tracts of tandem repeats may be required for neocentromere function, as argued below for centromeres. Although there are clearly more retroelements in knobs than in centromeres (Figure 6), this may reflect the fact that neocentromeres are activated only when Ab10 is present, and as a result



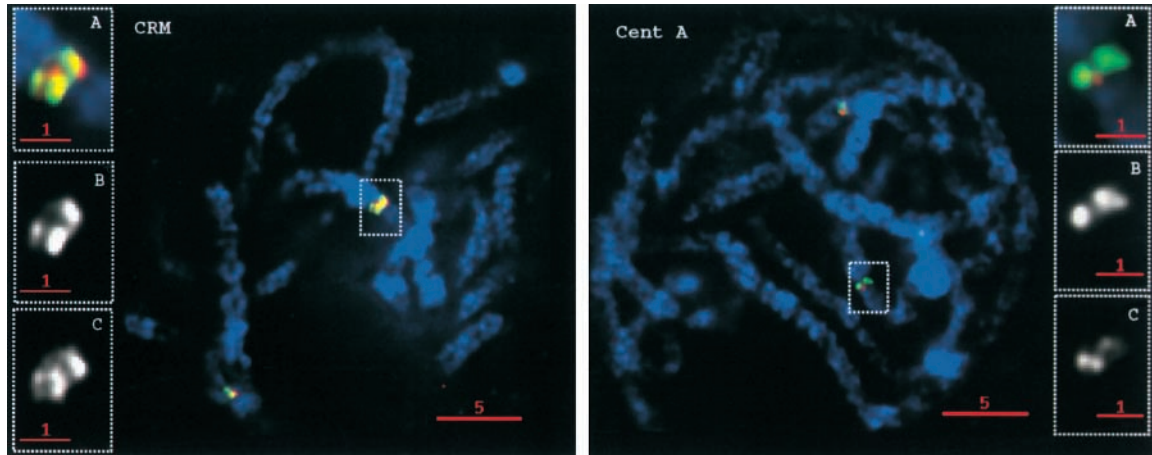


FIGURE 7.—Staining patterns of the maize CR elements CRM and CentA. The DNA is shown in blue, the retroelement is shown in red, and the centromeric repeat CentC is shown in green. Large images are single optical sections from maize pachytene cells. (A) Enlargement (3 $\times$ ) of the boxed regions. (B) Same images as A except only CentC staining is shown. (C) Same images as A showing only the retroelement staining.

there are less stringent evolutionary restraints upon the content of knobs. Interestingly, TR-1 satellite arrays are considerably more neocentric (active on the spindle) than 180-bp arrays and are controlled by different transacting factors (HIATT *et al.* 2002). Our data suggest that *Cinful-1* does not actively interfere with TR-1-mediated neocentromere activity and leave open the possibility that *Cinful-1* may contribute to knob motility, as has been suggested for the CR elements (ZHONG *et al.* 2002).

#### Abundance of maize retroelements in centromeres:

An unusual feature of cereal centromeres is that they specifically accumulate a group of Ty3-*gypsy*-like retroelements known as CR elements. CR elements show remarkable sequence conservation and are found in the centromeres of rice, wheat, sorghum, barley, rye, and oats (ARAGON-ALCAIDE *et al.* 1996; JIANG *et al.* 1996; MILLER *et al.* 1998; PRESTING *et al.* 1998; LANGDON *et al.* 2000). Zhong and co-workers have recently reported that maize CR elements interact with CENH3, indicating that they may have been coopted to perform a function in chromosome segregation (ZHONG *et al.* 2002). The strict centromeric localization of CR elements is even more striking in light of data shown here that several other maize retroelement families are poorly represented within centromeres, at least as defined by colocalization with the CentC satellite repeat. CentC strongly interacts with the kinetochore protein CENH3, suggesting that it is a legitimate marker for the functional centromere (ZHONG *et al.* 2002). *Cent-A* and CRM are abundant in centromeres (Figure 7), whereas retroelements that occur throughout the chromosome arms (*Grande*, *Opie*, *Prem-2/Ji*, *Huck*, *Cinful-1*, and *Tekay/Prem-1*) are significantly underrepresented in centromeres (Figures 2, 3, and 6, and Table 1). There is some family-to-family variation, but overall the non-CR retroelements are found threefold less frequently in centromeres than in euchromatin. The tandem repeats in maize centro-

meres apparently represent a niche of the genome that is favorable to CR elements and seemingly hostile to families of retroelements that proliferate throughout the intergenic, euchromatic niches of the genome.

Extended tandem repeat arrays are thought to be required for centromere function in animals (GRIMES *et al.* 2002) and are characteristic of the centromeres in *Arabidopsis* and rice (KUMEKAWA *et al.* 2001; CHENG *et al.* 2002). Rampant insertion of retroelements in the centromere would interrupt the continuity of these long arrays. On the basis of our data, we hypothesize that such insertions adversely impact the formation of centromeric chromatin. Other functional domains of the chromosome, most notably those containing genes, also rarely contain the large insertions that retrotransposons generate (SANMIGUEL *et al.* 1996). Non-CR retrotransposons may be excluded from the centromere by the chromatin structure at the centromere or, alternatively, could be eliminated during the process of unequal crossing over that is thought to homogenize tandem repeat arrays (EICKBUSH 2002). The CR elements may circumvent elimination via recombination by continuously transposing back into the centromeric satellite arrays, as has been proposed for the maintenance of R1 and R2 elements in *Drosophila* rRNA gene arrays (EICKBUSH 2002; PEREZ-GONZALEZ and EICKBUSH 2002). This interpretation is supported by sequencing data showing that most CR elements are recent insertions (NAGAKI *et al.* 2003a). Further, fiber-FISH analyses (CHENG *et al.* 2002) and our own localization data (Figure 7) suggest that CRs transpose into specific domains of the centromere, leaving the tandem repeat arrays largely intact.

At least a subset of known retroelements seems to have evolved mutualistic relationships with their hosts. Two particularly well-studied examples in this category are the HeT-A and TART elements of *Drosophila*, which substitute for telomerase by continually transposing into

chromosome ends (PARDUE and DEBARYSHE 2000). Although the host normally responds to retroelements by actively repressing them, *Drosophila* may tolerate HeT-A and TART because it benefits from their presence. Similarly, CR elements accumulate in specific domains of the centromere where they interact with the key kinetochore protein CENH3 (ZHONG *et al.* 2002). These data suggest that CR elements contribute to the specification and/or maintenance of the centromere/kinetochore complex. In the same manner that HeT-A and TART elements are limited to *Drosophila* species (CASACUBERTA and PARDUE 2003), the CR elements and their association with centromeres may be limited to cereal grains. No *Arabidopsis* retroelements transpose specifically into the centromere core (HAUPT *et al.* 2001; KUMEKAWA *et al.* 2001) and chromatin immunoprecipitation experiments have failed to reveal any association between retroelements and the *Arabidopsis* homolog of CENH3 (NAGAKI *et al.* 2003b). The functional centromeric sequences in *Arabidopsis*, as in humans (SCHUELER *et al.* 2001; GRIMES *et al.* 2002), appear to be long uninterrupted arrays of satellite repeats (NAGAKI *et al.* 2003b).

We thank Carolyn Lawrence for advice and help throughout the study, and Susan Wessler for suggesting that we analyze retroelement distribution on the Ab10 chromosome, as well as Evelyn Hiatt, Cedric Feschotte, and an anonymous reviewer for detailed comments on the manuscript. This work was supported by a grant from the National Science Foundation (9975827) to R.K.D. Additional support was provided to R.J.M. by a National Science Foundation interdisciplinary research training grant (BIR9220329).

#### LITERATURE CITED

- ANANIEV, E. V., R. L. PHILLIPS and H. W. RINES, 1998a Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc. Natl. Acad. Sci. USA* **95**: 13073–13078.
- ANANIEV, E. V., R. L. PHILLIPS and H. W. RINES, 1998b Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. *Genetics* **149**: 2025–2037.
- ANANIEV, E. V., R. L. PHILLIPS and H. W. RINES, 1998c A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: Are chromosome knobs megatransposons? *Proc. Natl. Acad. Sci. USA* **95**: 10785–10790.
- ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- ARAGON-ALCAIDE, L., T. MILLER, T. SCHWARZACHER, S. READER and G. MOORE, 1996 A cereal centromeric sequence. *Chromosoma* **105**: 261–268.
- ARDLIE, K. G., 1998 Putting the brake on drive: meiotic drive of t-haplotypes in natural populations of mice. *Trends Genet.* **14**: 189–193.
- BENNETZEN, J. L., 2002 Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29–36.
- BOWEN, N., and J. F. McDONALD, 2001 *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* **11**: 1527–1540.
- BRANDES, A., J. S. HESLOP-HARRISON, A. KAMM, S. KUBIS, R. L. DOUDRICK *et al.*, 1997 Comparative analysis of the chromosomal and genomic organization of *Ty1-copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.* **33**: 11–21.
- BUCKLER, E. S. I., T. L. PHELPS-DURR, C. S. KEITH BUCKLER, R. K. DAWE, J. F. DOEBLEY *et al.*, 1999 Meiotic drive of knobs reshaped the maize genome. *Genetics* **153**: 415–426.
- CASACUBERTA, E., and M. L. PARDUE, 2003 Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the *virilis* group. *Proc. Natl. Acad. Sci. USA* **100**: 3363–3368.
- CAVALIER-SMITH, T., 1978 Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J. Cell Sci.* **34**: 247–278.
- CHENG, Z., F. DONG, T. LANGDON, S. OUYANG, C. R. BUELL *et al.*, 2002 Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14**: 1691–1704.
- DAWE, R. K., J. W. SEDAT, D. A. AGARD and W. Z. CANDE, 1994 Meiotic chromosome pairing in maize is associated with novel chromatin organization. *Cell* **76**: 901–912.
- EDWARDS, K. J., J. VEUSKENS, H. RAWLES, A. DALY and J. L. BENNETZEN, 1996 Characterization of four dispersed repetitive DNA sequences from *Zea mays* and their use in constructing contiguous DNA fragments using YAC clones. *Genome* **39**: 811–817.
- EICKBUSH, T. H., 2002 R2 and related site-specific non-long terminal repeat retrotransposons, pp. 813–835 in *Mobile DNA II*, edited by N. L. CRAIG, R. CRAIGIE, M. GELLERT and A. LAMBOWITZ. ASM Press, Washington, DC.
- FENG, D. F., and R. F. DOOLITTLE, 1987 Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25**: 351–360.
- FENG, Q., Y. ZHANG, P. HAO, S. WANG, G. FU *et al.*, 2002 Sequence and analysis of rice chromosome 4. *Nature* **420**: 316–320.
- FRIESE, N., A. BRANDES and J. S. HESLOP-HARRISON, 2001 Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. *Mol. Biol. Evol.* **18**: 1176–1188.
- GCG, 1982–2000 Wisconsin package. Genetics Computer Group, Madison, WI.
- GRIMES, B. R., A. A. RHOADES and H. F. WILLARD, 2002 Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. *Mol. Ther.* **5**: 798–805.
- HAUPT, W., T. C. FISCHER, S. WINDERL, P. FRANSZ and R. A. TORRES-RUIZ, 2001 The centromere1 (CEN1) region of *Arabidopsis thaliana*: architecture and functional impact of chromatin. *Plant J.* **27**: 285–296.
- HENIKOFF, S., 2002 Near the edge of a chromosome's "black hole." *Trends Genet.* **18**: 165–167.
- HENIKOFF, S., and J. G. HENIKOFF, 1992 Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**: 10915–10919.
- HIATT, E. N., and R. K. DAWE, 2003 The meiotic drive system on maize abnormal chromosome 10 contains few essential genes. *Genetica* **117**: 67–76.
- HIATT, E. N., E. K. KENTNER and R. K. DAWE, 2002 Independently regulated neocentromere activity of two classes of tandem repeat arrays. *Plant Cell* **14**: 407–420.
- HOSOUCI, T., N. KUMEKAWA, H. TSURUOKA and H. KOTANI, 2002 Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. *DNA Res.* **9**: 117–121.
- HUDAKOVA, S., W. MICHALEK, G. G. PRESTING, R. TEN HOOPEN, K. DOS SANTOS *et al.*, 2001 Sequence organization of barley centromeres. *Nucleic Acids Res.* **29**: 5029–5035.
- JIANG, J., A. NASUDA, F. DONG, C. SCHERRER, S. S. WOO *et al.*, 1996 A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proc. Natl. Acad. Sci. USA* **93**: 14210–14213.
- KATO, Y. T. A., 1976 Cytological studies of maize (*Zea mays* L.) and teosinte (*Zea mexicana*) in relation to their origin and evolution. *Mass. Agric. Exp. Stat. Bull.* **635**: 1–185.
- KIKUDOME, G., 1959 Studies on the phenomenon of preferential segregation in maize. *Genetics* **44**: 815–831.
- KUMAR, A., and J. L. BENNETZEN, 1999 Plant retrotransposons. *Annu. Rev. Genet.* **33**: 479–532.
- KUMAR, A., S. R. PEARCE, K. McLEAN, G. HARRISON, J. S. HESLOP-HARRISON *et al.*, 1997 The *Ty1-copia* group of retrotransposons in plants: genomic organization, evolution, and use as molecular markers. *Genetica* **100**: 205–217.
- KUMEKAWA, N., T. HOSOUCI, H. TSURUOKA and H. KOTANI, 2001 The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. *DNA Res.* **8**: 285–290.
- LANGDON, T., C. SEAGO, M. MENDE, M. LEGGETT, H. THOMAS *et al.*,

- 2000 Retrotransposon evolution in diverse plant genomes. *Genetics* **156**: 313–325.
- MALIK, H. S., and T. H. EICKBUSH, 1999 Modular evolution of the integrase domain in the Ty3/*gypsy* class of LTR retrotransposons. *J. Virol.* **73**: 5186–5190.
- MEYERS, B. C., S. V. TINGEY and M. MORGANTE, 2001 Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* **11**: 1660–1676.
- MILLER, J. T., F. DONG, S. A. JACKSON, J. SONG and J. JIANG, 1998 Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**: 1615–1623.
- NAGAKI, K., J. SONG, R. M. STUPAR, A. S. PAROKONNY, Q. YUAN *et al.*, 2003a Molecular and cytological analyses of large tracks of centromeric DNA reveals the structure and evolutionary dynamics of maize centromeres. *Genetics* **163**: 759–770.
- NAGAKI, K., P. B. TALBERT, C. X. ZHONG, R. K. DAWE, S. HENIKOFF *et al.*, 2003b Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* **163**: 1221–1225.
- PAGEL, M., and R. A. JOHNSTONE, 1992 Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. *Proc. R. Soc. Lond. B Biol. Sci.* **249**: 119–124.
- PARDUE, M. L., and P. G. DEBARYSHE, 2000 *Drosophila* telomere transposons: genetically active elements in heterochromatin. *Genetica* **109**: 45–52.
- PEACOCK, W. J., E. S. DENNIS, M. M. RHOADES and A. J. PRYOR, 1981 Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proc. Natl. Acad. Sci. USA* **78**: 4490–4494.
- PEARCE, S. R., G. HARRISON, D. LI, J. S. HESLOP-HARRISON, A. KUMAR *et al.*, 1996a The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. *Mol. Gen. Genet.* **250**: 305–315.
- PEARCE, S. R., U. PICH, G. HARRISON, A. J. FLAVELL and J. S. HESLOP-HARRISON, 1996b The Ty1-copia group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chromosome Res.* **4**: 357–364.
- PEARCE, S. R., G. HARRISON, J. S. HESLOP-HARRISON, A. J. FLAVELL and A. KUMAR, 1997 Characterization and genomic organization of Ty1-copia group retrotransposons in rye (*Secale cereale*). *Genome* **40**: 617–625.
- PEREZ-GONZALEZ, C., and T. H. EICKBUSH, 2002 Rates of R1 and R2 retrotransposition and elimination from the rDNA locus of *Drosophila melanogaster*. *Genetics* **162**: 799–811.
- PICH, U., and I. SCHUBERT, 1998 Terminal heterochromatin and alternative telomeric sequences in *Allium cepa*. *Chromosome Res.* **6**: 315–321.
- PRESTING, G. G., L. MALYSHEVA, J. FUCHS and I. SCHUBERT, 1998 A Ty3/*gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J.* **16**: 721–728.
- RHOADES, M. M., 1942 Preferential segregation in maize. *Genetics* **27**: 395–407.
- RHOADES, M. M., and E. DEMPSEY, 1985 Structural heterogeneity of chromosome 10 in races of maize and teosinte, pp. 1–18 in *Plant Genetics*, edited by M. FREELING. Alan R. Liss, New York.
- SANMIGUEL, P., and J. L. BENNETZEN, 1998 Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **82**: 37–44.
- SANMIGUEL, P., A. TIKHONOV, Y. JIN, N. MOTCHOULSKAIA, D. ZAKHAROV *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- SCHUELER, M. G., A. W. HIGGINS, M. K. RUDD, K. GUSTASHAW and H. F. WILLARD, 2001 Genomic and genetic definition of a functional human centromere. *Science* **294**: 109–115.
- SNIEGOWSKI, P. D., and B. CHARLESWORTH, 1994 Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics* **137**: 815–827.
- SONG, J., F. DONG, J. W. LILLY, R. M. STUPAR and J. JIANG, 2001 Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA sequences. *Genome* **44**: 463–469.
- SWOFFORD, D., 1999 *PAUP\**. Sinauer Associates, Sunderland, MA.
- TIKHONOV, A. P., P. J. SANMIGUEL, Y. NAKAJIMA, N. M. GORENSTEIN, J. L. BENNETZEN *et al.*, 1999 Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* **96**: 7409–7414.
- VINCENT, C. M., A. SUONIEMI, K. ANAMTHAWAT-JONSSON, J. TANSKANEN, A. BEHARAV *et al.*, 1999 Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**: 1769–1784.
- WICKER, T., N. STEIN, L. ALBAR, C. FEUILLET, E. SCHLAGENHAUF *et al.*, 2001 Analysis of a contiguous 211kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J.* **26**: 307–316.
- XIONG, Y., and H. EICKBUSH, 1988 Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* **5**: 675–690.
- XIONG, Y., and H. EICKBUSH, 1990 Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**: 3353–3362.
- ZHONG, C., J. B. MARSHALL, C. TOPP, R. MROCZEK, A. KATO *et al.*, 2002 Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* **14**: 2825–2836.

Communicating editor: S. HENIKOFF

