

Molecular and Cytological Analyses of Large Tracks of Centromeric DNA Reveal the Structure and Evolutionary Dynamics of Maize Centromeres

Kiyotaka Nagaki,* Junqi Song,* Robert M. Stupar,* Alexander S. Parokony,* Qiaoping Yuan,[†] Shu Ouyang,[†] Jia Liu,[†] Joseph Hsiao,[†] Kristine M. Jones,[†] R. Kelly Dawe,[‡] C. Robin Buell[†] and Jiming Jiang*^{*,1}

*Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706, [†]The Institute for Genomic Research, Rockville, Maryland 20850 and [‡]Department of Plant Biology, University of Georgia, Athens, Georgia 30602

Manuscript received August 5, 2002

Accepted for publication November 22, 2002

ABSTRACT

We sequenced two maize bacterial artificial chromosome (BAC) clones anchored by the centromere-specific satellite repeat CentC. The two BACs, consisting of ~200 kb of cytologically defined centromeric DNA, are composed exclusively of satellite sequences and retrotransposons that can be classified as centromere specific or noncentromere specific on the basis of their distribution in the maize genome. Sequence analysis suggests that the original maize sequences were composed of CentC arrays that were expanded by retrotransposon invasions. Seven centromere-specific retrotransposons of maize (CRM) were found in BAC 16H10. The CRM elements inserted randomly into either CentC monomers or other retrotransposons. Sequence comparisons of the long terminal repeats (LTRs) of individual CRM elements indicated that these elements transposed within the last 1.22 million years. We observed that all of the previously reported centromere-specific retrotransposons in rice and barley, which belong to the same family as the CRM elements, also recently transposed with the oldest element having transposed ~3.8 million years ago. Highly conserved sequence motifs were found in the LTRs of the centromere-specific retrotransposons in the grass species, suggesting that the LTRs may be important for the centromere specificity of this retrotransposon family.

THE centromeres of eukaryotic chromosomes are responsible for sister chromatid cohesion and serve as the sites for kinetochore assembly and spindle fiber attachment during cell division. Thus, centromeres are critical for the segregation and transmission of genetic information. In the budding yeast *Saccharomyces cerevisiae*, the functional centromere is defined by a ~125-bp sequence (CLARKE 1998). However, in the majority of eukaryotic species, centromeres are embedded in long tracks of highly repetitive DNA sequences with satellite repeats often the major DNA component of centromeres in higher eukaryotic species (CSINK and HENIKOFF 1998). For example, a 171-bp tandem repeat, the α -satellite, is located in the centromeres of all human chromosomes. Human artificial chromosomes have been successfully assembled using either synthetic or cloned α -satellite DNA as the centromere component (HARRINGTON *et al.* 1997; IKENO *et al.* 1998; HENNING *et al.* 1999), suggesting that a long stretch of α -satellite DNA can act as a functional human centromere.

The centromeres of *Arabidopsis thaliana* chromosomes are among the most well-studied plant centromeres. *A. thaliana* centromeres were mapped genetically using tetrad-based genetic mapping (COPENHAVER *et al.* 1999).

DNA sequences within the genetically mapped centromeres were cloned and analyzed (COPENHAVER *et al.* 1999; ARABIDOPSIS GENOME INITIATIVE 2000; KUMEKAWA *et al.* 2000, 2001). The most abundant DNA element in *A. thaliana* centromeres is the pAL1 repeat, a 180-bp satellite repeat family (MARTINEZ-ZAPATER *et al.* 1986; MALUSZYNKA and HESLOP-HARRISON 1991; ROUND *et al.* 1997). The cytological locations of the pAL1 repeat coincide with the centromeric H3 histone (TALBERT *et al.* 2002). The pAL1 repeat is organized into long tandem arrays (JACKSON *et al.* 1998) that may be interrupted by the 106B repeat, a diverged copy of the long terminal repeat (LTR) of the *Athila* retrotransposon (FRANSZ *et al.* 2000). The *Athila* element, the most dominant retrotransposon family in *A. thaliana*, and a number of other repetitive DNA elements are highly enriched in pericentromeric regions of all five *A. thaliana* centromeres (FRANSZ *et al.* 2000; KUMEKAWA *et al.* 2000, 2001).

Two highly conserved repetitive DNA elements were reported in centromeres of grass species (ARAGON-ALCAIDE *et al.* 1996; JIANG *et al.* 1996). These two sequences are derived from a Ty3/*gypsy* class of retrotransposon (MILLER *et al.* 1998a; PRESTING *et al.* 1998; LANGDON *et al.* 2000). The centromere-specific retrotransposon sequences provide excellent probes to isolate DNA clones derived from grass centromeres. Such clones have been reported in a number of plant species, including rice (DONG *et al.* 1998; NONOMURA and KUR-

¹Corresponding author: Department of Horticulture, 1575 Linden Dr., University of Wisconsin, Madison, WI 53706.
E-mail: jjiang1@facstaff.wisc.edu

ATA 1999), barley (PRESTING *et al.* 1998), and maize (ANANIEV *et al.* 1998). DNA sequences associated with centromeric regions have also been reported in numerous other plant species (HARRISON and HESLOP-HARRISON 1995; MILLER *et al.* 1998b; NAGAKI *et al.* 1998; FRANCKI 2001; GINDULLIS *et al.* 2001; HUDAKOVA *et al.* 2001; KISHII *et al.* 2001; SAUNDERS and HOUBEN 2001).

Maize has become an important model for plant centromere research. ALFENITO and BIRCHLER (1993) isolated a repetitive DNA element that is specific to the centromeres of maize B chromosomes. This repeat is present in all significantly rearranged B centromeres (KASZAS and BIRCHLER 1996, 1998), suggesting that it is essential for B centromere function. A repetitive DNA element was recently isolated from the centromere of maize chromosome 4 on the basis of its partial sequence homology with the B centromeric repeat (PAGE *et al.* 2001). Cosmid clones derived from the centromeric region of maize chromosome 9 were identified in a library constructed from an oat-maize chromosome 9 addition line (ANANIEV *et al.* 1998). A 156-bp satellite repeat, CentC, was discovered from these cosmid clones. CentC is found only at maize centromeres, but the amount of CentC repeat is highly variable among the 10 maize centromeres (ANANIEV *et al.* 1998).

Although several DNA elements have been isolated from the maize centromeres, the large-scale organization of maize centromeric DNA, especially in the A chromosomes, is not known. In this study, we isolated and sequenced two maize bacterial artificial chromosome (BAC) clones derived from the centromeric regions. We found that the CentC satellite and retrotransposons, both centromere specific and noncentromere specific, are the primary DNA components of maize centromeres. Molecular and cytological analyses of the centromere-specific retrotransposons in maize and other cereal species revealed the structural diversity and evolutionary dynamics of this special retrotransposon family that may play an important role in grass centromere evolution.

MATERIALS AND METHODS

BAC library construction and screening: A BAC library was constructed from maize inbred line Mo17 according to SONG *et al.* (2000). The *Bam*HI cloning site of vector pBeloBAC11 (SHIZUYA *et al.* 1992) was used for library construction. The 9216 clones were placed on 24 384-well plates. Filter preparation and library screening were according to published protocols (NIZETIC *et al.* 1990). DNA sequences homologous to the maize centromeric repeats CentC and CentA (ANANIEV *et al.* 1998) were amplified from maize genomic DNA and cloned into plasmid vectors. Two plasmid clones, pCentA-int and pCentC-1, were used to screen the BAC library.

Fluorescent *in situ* hybridization: Maize inbred line Mo17 was used for cytological analysis. The fluorescence *in situ* hybridization (FISH) procedures on metaphase chromosomes and individual BAC molecules were essentially the same as previously published protocols (JIANG *et al.* 1995; JACKSON *et al.* 1999). All images were captured digitally using a SenSys

charge-coupled device (CCD) camera (Roper Scientific, Tucson, AZ) attached to an Olympus BX60 epifluorescence microscope. The camera control and image analysis were performed using IPLab Spectrum v3.1 software (Signal Analytics, Vienna, VA).

Polymerase chain reaction: To detect each subfamily of the centromere-specific retrotransposons in maize, primers specific to each subfamily were designed for the 5' LTR and 5' untranslated region (UTR). Primers include CRM1a-U (5'-ACACCAGCAGCACCTTCTCCAG-3'), CRM1a-L (5'-AGTTC TTATCCGTTCTTACCAA-3'), CRM2a-U (5'-GCTCGTCAAC TCAACCATCAGG-3'), and CRM2a-L (5'-GCCCATCTTTT CATTTCGTCAC-3'). Two primers were designed to amplify the 77-bp repeat discovered in BAC 15C5: ZMA77-U (5'-TTT TGCACGGATAGTCTTCG-3') and ZMA77-L (5'-TCCGTGCAA AAGTCGCCTAA-3'). The specific regions were amplified from the genomic DNA of Mo17 by 30 cycles of polymerase chain reaction (PCR) with the following conditions: 94° for 30 sec, 52° for 30 sec, and 72° for 2 min.

DNA sequencing: The sequences of the two maize BAC clones, 15C5 and 16H10, were determined essentially as described by YUAN *et al.* (2002). For 15C5, a 2- to 3-kb and a 10- to 15-kb shotgun library were constructed and these libraries were sequenced to provide a total of ~14× sequence coverage. For 16H10, a 2- to 3-kb and a 4- to 8-kb shotgun library were constructed and sequenced to provide >10× sequence coverage. Shotgun sequences for each BAC were assembled using TIGR assembler (SUTTON *et al.* 1995). Closure reactions were performed on the BACs using a combination of resequencing, alternative chemistries, transposon-based sequencing, and primer walking. Some of the assemblies could be ordered on the basis of clone mate pairs and the presence of the BAC vector. The sequences have been submitted to GenBank with accession nos. AC116034 (BAC 16H10) and AC116033 (BAC 15C5).

Sequence analysis: DNA sequences similar to the BAC assemblies were searched in the GenBank database using BLASTN. DNA elements in the sequences were analyzed by MegAlign software (DNASTAR, Madison, WI). The ages of the retrotransposons discovered in the two maize BACs were estimated by sequence comparison between the two LTRs of the elements. The LTRs were first aligned by CLUSTAL X v1.81 software (THOMPSON *et al.* 1997). Kimura's distance (KIMURA 1980) of the two LTRs of individual retrotransposons was estimated by the maximum-likelihood method using the baseml program with the K80 model in the PAML 3.11 PPC package (YANG 1997). The reported substitution rate per synonymous site per year in maize and Kimura's distances were then used to estimate the age of the elements (GAUT *et al.* 1996). The phylogeny of the retrotransposons in the BACs was analyzed by the neighbor-joining method with CLUSTAL X v1.81 software (SAITOU and NEI 1987; THOMPSON *et al.* 1997).

RESULTS

Isolation of centromeric BACs for sequencing: We constructed a BAC library of maize inbred line Mo17, which consists of 9216 clones with an average insert size of 120 kb. Two plasmid clones, pCentA-int and pCentC-1, were used as probes to identify centromeric clones from the BAC library. Probe pCentC-1 contains a 156-bp satellite DNA element CentC that is specific to the centromeres of maize chromosomes (ANANIEV *et al.* 1998). Probe pCentA-int is derived from a portion of the centromere-specific retrotransposon sequence

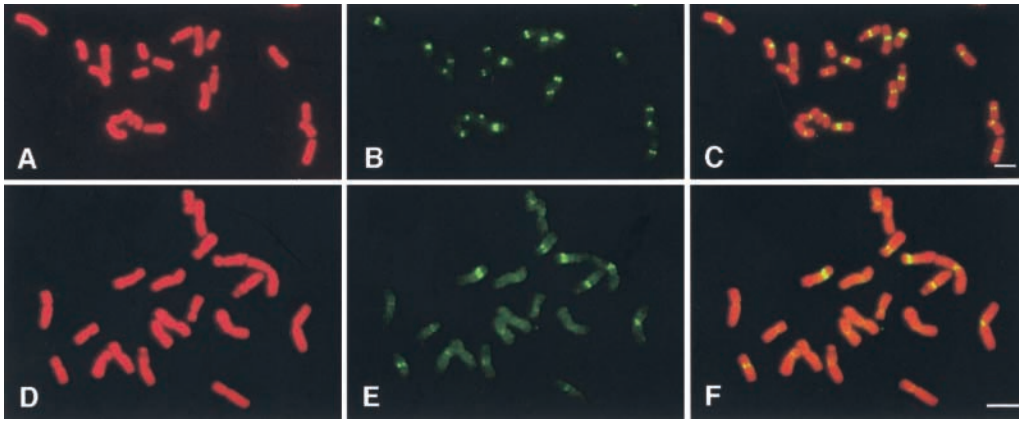


FIGURE 1.—FISH mapping of centromeric BACs 16H10 (A–C) and 15C5 (D–F) on somatic metaphase chromosomes of maize inbred Mo17. (A and D) Somatic metaphase chromosomes; (B and E) FISH signals; (C and F) merged images. Chromosomes are stained by 4',6-diamidino-2-phenylindole (DAPI) and presented by a pseudo-red color. Bars, 5 μ m.

CentA that is almost exclusively located in the centromeric regions of maize chromosomes (ANANIEV *et al.* 1998). BAC library screening using these two probes identified a total of 96 positive clones, including 18 specific to CentA, 64 specific to CentC, and 14 identified by both probes.

Two BAC clones, 16H10 and 15C5, were selected for further analysis. BACs 16H10 and 15C5 contain inserts of 95 and 100 kb, respectively, based on fingerprint analyses using both *NotI* and *BamHI* digestions (data not shown). FISH analysis on maize metaphase chromosomes showed that the signals derived from 16H10 were almost exclusively localized in the centromeres (Figure 1, A–C). Major FISH signals from 15C5 were also located in the centromeres. However, faint signals uniformly covered the entire length of all maize chromosomes (Figure 1, D–F). The amount and location of the CentC sequences in the two BAC clones were determined by FISH mapping on individual BAC molecules as described by JACKSON *et al.* (1999). The average sizes of the CentC tracts were calculated from 10 FISH images. BAC 16H10 contains three CentC tracts, and the sizes of the tracts are 18.0, 2.4, and 1.8% of the BAC molecule (including the vector), respectively (Figure 2, A and C). BAC 15C5 also contains three CentC tracts, and sizes of the tracts are 6.1, 1.5, and 1.7% of the BAC molecule, respectively (Figure 2, B and D).

Sequence analysis of BAC clone 16H10: BAC 16H10 was sequenced to $>10\times$ sequence coverage (see MATERIALS AND METHODS). The sequences generated from 16H10 were assembled into two large contigs (34,079 and 21,043 bp, respectively) and eight small contigs (9438, 4686, 3066, 2491, 2143, 1904, 1494, and 981 bp, respectively). The total length of these 10 contigs is 81,325 bp, slightly smaller than the 95 kb estimated by fingerprint analysis, suggesting that a portion of the highly conserved repetitive sequences within the BAC were collapsed within the contigs. However, a substantial portion of the 81-kb assembled sequence (74.8 kb) was correctly assembled as determined by inspection of clone mates and use of transposon-based sequencing of the large insert shotgun clones. The order of the contigs

in Figure 3 is determined on the basis of structure and locations of specific retroelements within the BAC insert and the presence of the BAC vector. Both large contigs (ASM 37376, 34,079 bp; and ASM 37375, 21,043 bp) and 4 of the 8 small contigs could be placed within the BAC insert using this approach (ASM 37379, 9438 bp; ASM 37381, 4686 bp; ASM 37378, 3066 bp; and ASM 37606, 981 bp; Figure 3).

Four CentC tracts were found in 16H10 and were named as CentC tracts A1, A2, B, and C, respectively (Figure 3). The total length of CentC tracts A1 and tract A2, including the gap separating these two tracts, was determined to be ~ 25 kb by restriction digestions followed by Southern hybridization (data not shown), suggesting an ~ 12 -kb gap separating ASM 37375 and ASM 37379 (Figure 3). Nine retrotransposons were found in 16H10, including seven elements homologous to the

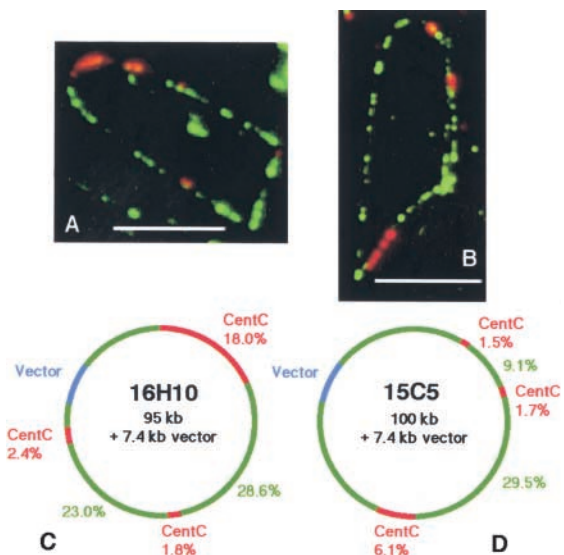
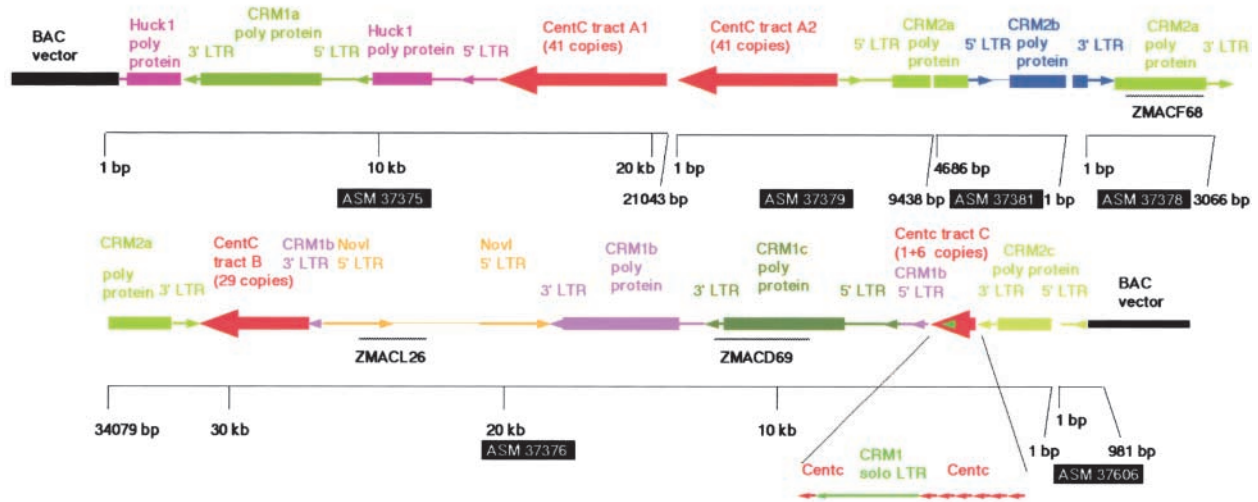


FIGURE 2.—Structure of maize BACs 16H10 and 15C5 revealed by fiber-FISH mapping. DNA from BACs 16H10 (A) and 15C5 (B) was labeled as green and pCentC-1 was labeled as red (bars, 5 μ m). The amount and locations of the CentC sequences within the BAC inserts were revealed by this method and are illustrated in C and D.

16H10



15C5

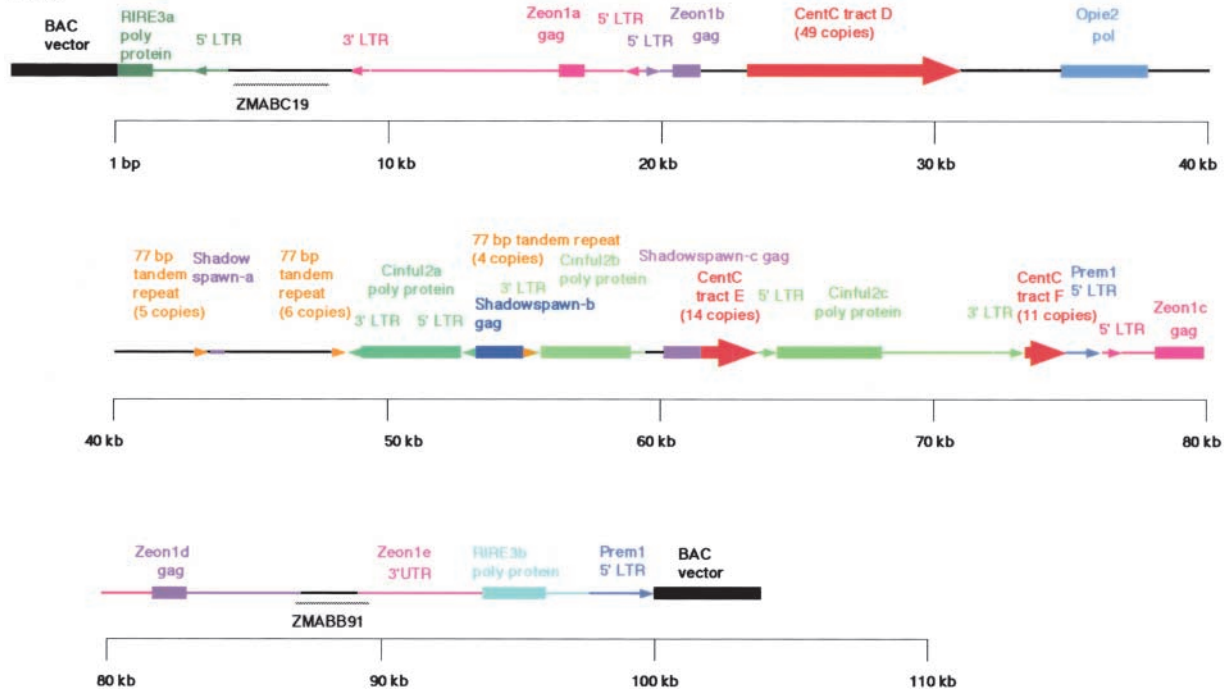


FIGURE 3.—Sequence organization of maize centromeric BACs 16H10 and 15C5. The order of the sequence contigs in 16H10 was determined on the basis of the sequence information of specific retroelements within the BAC insert and the presence of the BAC vector. Each retrotransposon is marked by a different color. The name, LTRs, and polyprotein of the same element are in the same color to facilitate the identification of interrupted retrotransposons.

centromeric retrotransposon of rice (CRR; CHENG *et al.* 2002) element (Table 1 and Figure 3). The CRR-like elements in maize were named *centromeric retrotransposon of maize* (CRM) thereafter. Six CRM elements, including CRM1a, CRM1b, CRM1c, CRM2a, CRM2b, and CRM2c, are complete or near-complete elements. The seventh CRM element is a solo LTR inserted in the middle of CentC tract C (Figure 3).

The two non-CRM elements include a Huck1 element

and a nonautonomous retroelement that is novel and different from any published maize retrotransposon families. We named this a Nov1 element. A shotgun clone containing sequences derived from the Nov1 element was used as a probe for FISH analysis (Figure 4, A and B). Dispersed signals were observed from the probe, indicating that the Nov1 element is not specific to the centromeres. The last CRM element, CRM2c, is located between CentC tract C and the BAC vector. A

TABLE 1
Retrotransposons in the two sequenced centromeric BACs of maize

Name	Homologous sequence ^a	Type	Clone	Note
CRM1a	RIRE7 (AB033235)	<i>gypsy</i>	16H10	
CRM1b	RIRE7 (AB033235)	<i>gypsy</i>	16H10	
CRM1c	RIRE7 (AB033235)	<i>gypsy</i>	16H10	
CRM1 solo LTR	RIRE7 (AB033235)	<i>gypsy</i>	16H10	Solo LTR
CRM2a	RIRE7 (AB033235)	<i>gypsy</i>	16H10	
CRM2b	RIRE7 (AB033235)	<i>gypsy</i>	16H10	Internal region truncated
CRM2c	RIRE7 (AB033235)	<i>gypsy</i>	16H10	5' and internal region truncated
Huck1	Huck1b (AF391808)	<i>gypsy</i>	16H10	5' truncated
Nov1		Unknown	16H10	
RIRE3a	RIRE3 (AB014738)	<i>gypsy</i>	15C5	Polyprotein
RIRE3b	RIRE3 (AB014738)	<i>gypsy</i>	15C5	Polyprotein
Cinful1	cinful1 (AF049110)	<i>gypsy</i>	15C5	LTR
Zeon1a	Zeon1 (ZMU11059)	Non- <i>gypsy</i> /copia	15C5	
Zeon1b	Zeon1 (ZMU11059)	Non- <i>gypsy</i> /copia	15C5	5' LTR and gag
Zeon1c	Zeon1 (ZMU11059)	Non- <i>gypsy</i> /copia	15C5	5' LTR, gag, and 3' UTR
Zeon1d	Zeon1 (ZMU11059)	Non- <i>gypsy</i> /copia	15C5	5' LTR, gag, and 3' UTR
Zeon1e	Zeon1 (ZMU11059)	Non- <i>gypsy</i> /copia	15C5	3' UTR
Shadowspawn1a	Shadowspawn1 (AF326577)	Non- <i>gypsy</i> /copia	15C5	A part of noncoding region
Shadowspawn1b	Shadowspawn1 (AF326577)	Non- <i>gypsy</i> /copia	15C5	gag
Shadowspawn1c	Shadowspawn1 (AF326577)	Non- <i>gypsy</i> /copia	15C5	gag
Cinful2a	Cinful2 (AF123535)	<i>gypsy</i>	15C5	
Cinful2b	Cinful2 (AF123535)	<i>gypsy</i>	15C5	
Cinful2c	Cinful2 (AF123535)	<i>gypsy</i>	15C5	
Opie2	Opie2 (ZMU68408)	<i>copia</i>	15C5	pol

^a GenBank accession numbers are in parentheses.

solo LTR, which is most likely derived from a different CRM element, is found in the middle of CentC track C (Figure 3).

Sequence analysis of BAC clone 15C5: The sequences generated from 15C5 were assembled into a single contig with a length of 99,979 bp, which is consistent with the estimated size of 100 kb based on fingerprint analyses.

Three CentC tracts, named D, E, and F, were found in 15C5. A total of 15 retrotransposons were discovered in 15C5 (Table 1 and Figure 3), including two complete Cinful2-like elements and one complete Zeon1 element. The remaining 12 retrotransposons have significantly decayed and their structures were difficult to determine. A novel 77-bp tandem repeat was found in BAC 15C5 (Figure 3). Two primers, ZMA77-U and ZMA77-L (see MATERIALS AND METHODS), were designed to amplify this repeat from maize genomic DNA and the PCR product was labeled as a probe for FISH analysis. Dispersed FISH signals were observed on maize metaphase chromosomes, indicating that this repeat is not specific to maize centromeres (Figure 4, C and D).

Several regions within BAC 15C5 did not show any homology with known repeats or transposons within GenBank. Shotgun clones derived from these regions were used as FISH probes, and they all generated dispersed signals that are enriched in the pericentromeric regions (Figures 4, E–H), suggesting that much of the

novel sequence is composed of degenerated retrotransposons.

Phylogenetic analysis of the centromere-specific retrotransposons: Ty3/*gypsy*-type retrotransposons similar to those in the CRM family have been found in the centromeric regions of all grass chromosomes (MILLER *et al.* 1998a; PRESTING *et al.* 1998; LANGDON *et al.* 2000). These centromeric retrotransposons in grass species (referred to as CR elements) can be divided into “autonomous” and “nonautonomous” subfamilies (LANGDON *et al.* 2000). The autonomous CR elements are full-size elements. The nonautonomous CR elements have an internal deletion leading to the loss of all enzymatic functions, resulting in the retrotransposons having only LTRs, a 5' UTR, and a gag structural gene fragment, truncated before the canonical RNA-binding motif (LANGDON *et al.* 2000).

A number of CR elements from rice, maize, and barley were used in phylogenetic analysis. These CR elements were described in previous reports or were directly deposited in GenBank (Table 2). The polyprotein regions from autonomous CR elements and two typical Ty3/*gypsy* retrotransposons of rice (RIRE3) and maize (Huck2) were analyzed by the neighbor-joining method (Figure 5A). Consistent with previous data (LANGDON *et al.* 2000) we found that the CR elements formed a cluster distinct from other rice and maize Ty3/*gypsy*

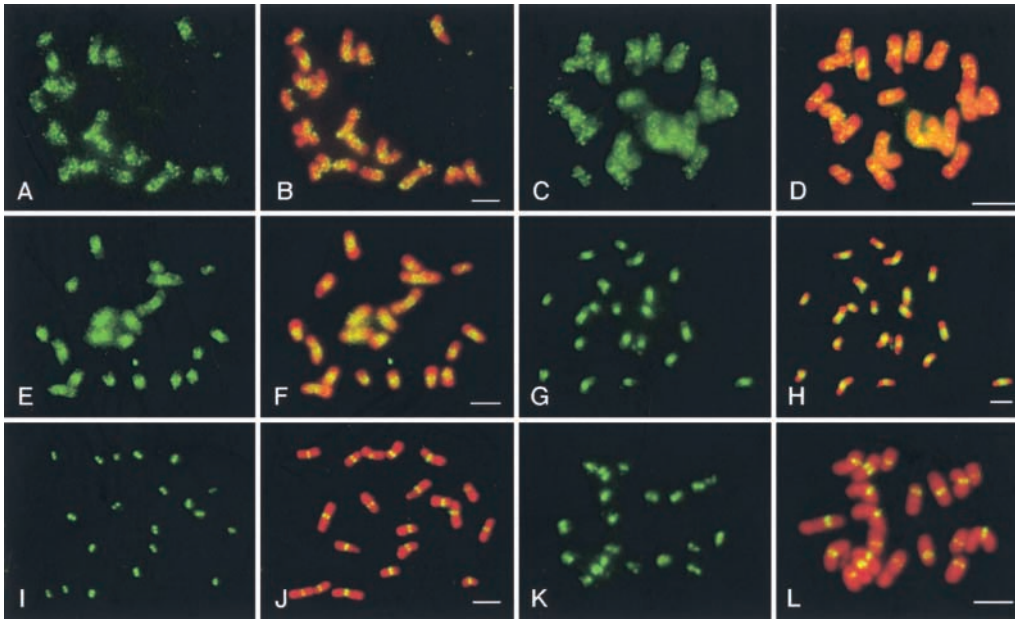


FIGURE 4.—FISH analysis using shotgun plasmid clones from maize BACs 16H10 and 15C5. The locations of plasmids within the BAC inserts are marked in Figure 3. (A and B) FISH pattern of plasmid ZMACL26 derived from retrotransposon Nov1. (C and D) Chromosomal locations of the ZMA77bp tandem repeat. This repeat was amplified from maize genomic DNA using primers ZMA77-U and ZMA77-L (see MATERIALS AND METHODS). The PCR product was labeled as a FISH probe that hybridized almost uniformly to the chromosomes although enhanced pericentromeric signals were observed in some chromosomes. (E and F) FISH-pattern of plasmid ZMABC19

derived from possibly decayed retrotransposon sequences in BAC 15C5. (G and H) FISH pattern of plasmid ZMABC91 derived from possibly decayed retrotransposon sequences in BAC 15C5. (I and J) FISH pattern of plasmid ZMACD69 derived from CRM1c. (K and L) FISH pattern of plasmid ZMACD68 derived from CRM2a. Chromosomes are stained by DAPI and presented by a pseudo-red color. Bars, 5 μ m.

elements (Figure 5A). This CR cluster can be divided into five species-specific subclusters. The maize sequences fall into two of these subclusters. CRM1a, -1b, and -1c fall into one subcluster, while the second maize subcluster, including CRM2a, -2b, and -2c, is more closely related to one of the two rice subclusters (Figure 5A). Our FISH data showed that the elements in both subclusters are centromere specific (Figure 4, I–L).

Similar phylogenetic results were obtained from the 5'

UTR (data not shown) and LTR regions (Figure 5B). Three nonautonomous CR elements were included in the LTR-based phylogenetic tree, including the CentA element (ANANIEV *et al.* 1998), a CRR element in RCB11 (RCB11-1; NONOMURA and KURATA 1999; LANGDON *et al.* 2000), and the CRR4.4kb element in rice BAC 17p22 (CHENG *et al.* 2002). These nonautonomous elements made independent clusters from the full-size elements in both rice and maize (Figure 5B).

TABLE 2
CR elements used in phylogenetic analysis

Element	Type	GenBank accession	Species	Reference
CRR AP003054	Autonomous	AP003054	Rice	
CRR AP003562	Autonomous	AP003562	Rice	
CRR AC022352	Autonomous	AC022352	Rice	
CRR AC079634	Autonomous	AC079634	Rice	
CRR AC090441	Autonomous	AC090441	Rice	
CRR AC092749	Autonomous	AC092749	Rice	
CRR9.2kb	Autonomous	AY101510	Rice	CHENG <i>et al.</i> (2002)
RIRE7	Autonomous	AB033235	Rice	
CentA-target ^a	Autonomous	AF078917	Maize	ANANIEV <i>et al.</i> 1998; LANGDON <i>et al.</i> (2000)
<i>Cereba</i>	Autonomous	AF078801	Barley	PRESTING <i>et al.</i> (1998)
<i>cereba-1</i>	Autonomous	AY040832	Barley	HUDAKOVA <i>et al.</i> (2001)
<i>cereba-2</i>	Autonomous	AY040832	Barley	HUDAKOVA <i>et al.</i> (2001)
CentA	Nonautonomous	AF078917	Maize	ANANIEV <i>et al.</i> (1998)
RCB11-1	Nonautonomous	AB013613	Rice	NONOMURA and KURATA (1999)
CRR4.4kb	Nonautonomous	AY101510	Rice	CHENG <i>et al.</i> (2002)

^a The CentA repeat is a nonautonomous CR element (LANGDON *et al.* 2000). The CentA element reported by ANANIEV *et al.* (1998) inserted into the polyprotein region of another CR element, which is named "CentA-target element" (LANGDON *et al.* 2000).

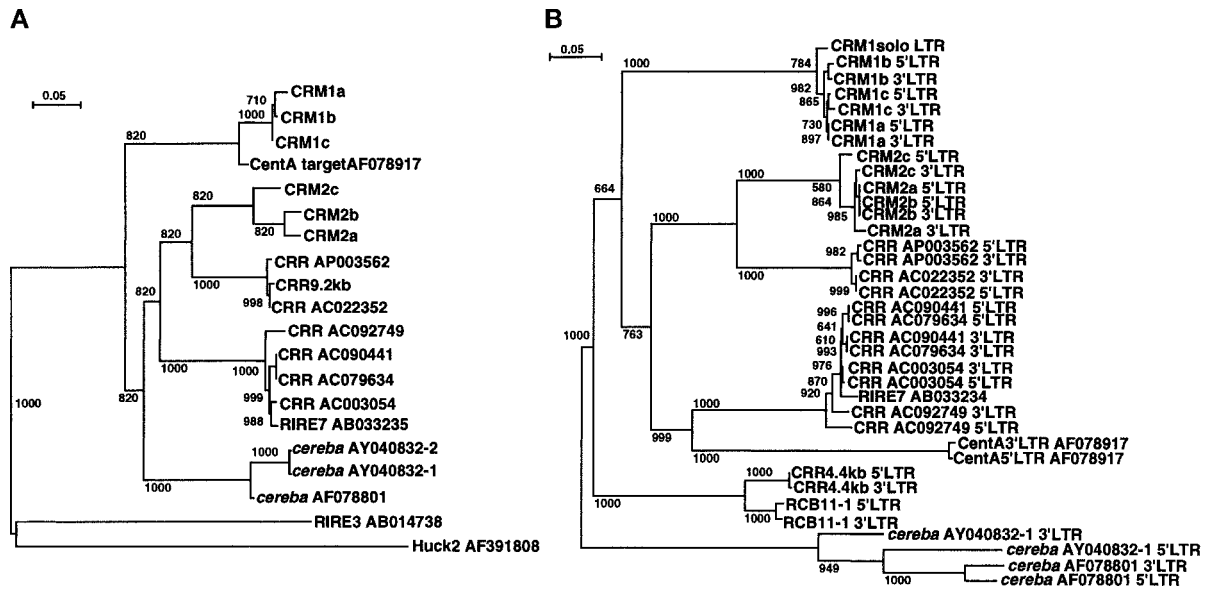


FIGURE 5.—Phylogenetic analysis of the CR elements from barley, rice, and maize. Bootstrap values in 1000 tests are indicated on the branches. (A) Phylogenetic tree constructed from the gag-pol polyprotein genes. For CRM2b, the polyprotein region in ASM 37378 (Figure 3) was used in the phylogenetic analysis. (B) Phylogenetic tree constructed from the LTRs. For CRM2b, the 3' LTR in ASM 37376 (Figure 3) was used in the phylogenetic analysis.

Four conserved domains were observed in the LTRs of the CR elements from different species (Figure 6). These highly conserved DNA motifs were found in both autonomous and nonautonomous CR elements despite the fact that these elements fall in different clusters in the phylogenetic tree (Figure 5B), suggesting that these motifs may be important for the targeting of the CR elements in centromeric regions.

Phylogenetic studies revealed that the full-size CR elements in rice and maize can be grouped into two distinct subfamilies (Figure 5, A and B). We analyzed the sequence similarity between the two subfamilies in maize and rice using the MegAlign program in DNASTAR and found that the LTRs and 5' UTRs are significantly more diverged than the pol and gag regions (data not shown). To reveal potential differences in the distribution of

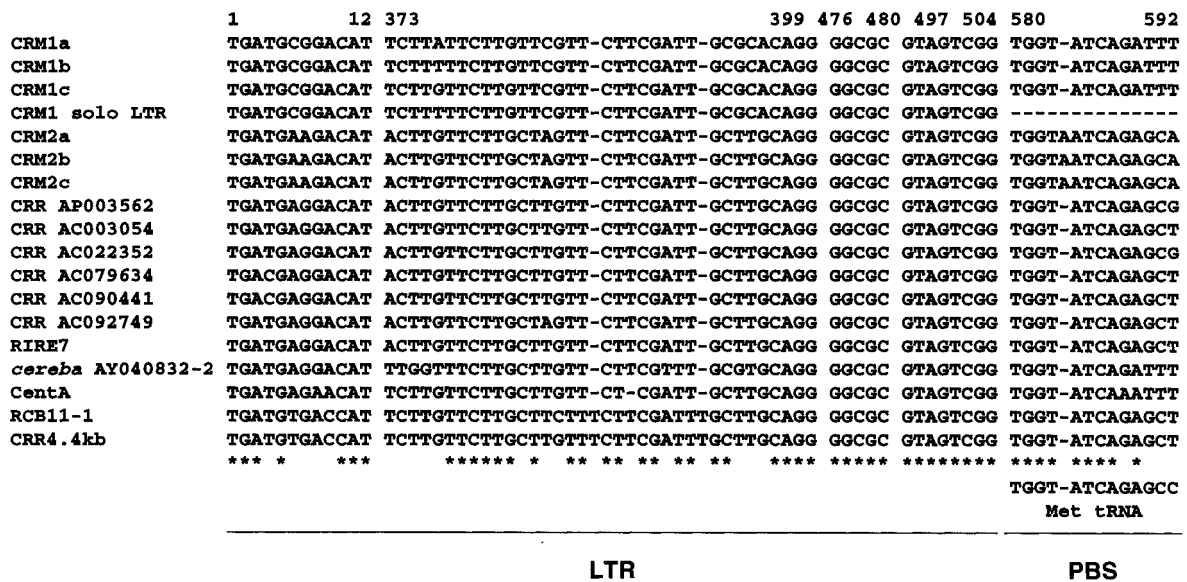


FIGURE 6.—Conserved motifs in the LTR and PBS of the CR elements in barley, rice, and maize. The 5' LTR and PBS of the retrotransposons were aligned. In the CRM2b and the cereba element, the 3' LTR was used instead of its 5' LTR, because the 5' end of the 5' LTR was truncated. Nucleotide positions of the conserved regions in CRM1a are indicated above the sequences. Stars at the bottom of the sequence indicate conserved base in the sequences. A complement sequence of methionine tRNA is indicated at the bottom of the PBS region.

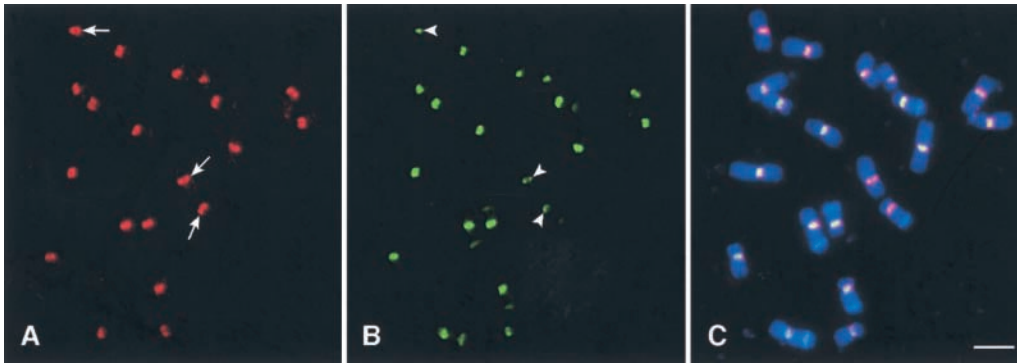


FIGURE 7.—Chromosomal localization of the PCR products amplified from the CRM1a and CRM2a subfamilies. (A) Signals derived from the PCR products amplified from CRM2a. (B) Signals derived from the PCR products amplified from CRM1a. (C) The FISH signals were merged with the metaphase chromosomes. Chromosomes are stained by DAPI. Note that the PCR products amplified

from CRM1a generated minor signals in the knob regions that are stained more intensively by DAPI than were the rest of the chromosomes. Some centromeres (arrows in A and arrowheads in B) show significant differences in the size and intensity of the FISH signals from the two subfamilies. Bar, 5 μ m.

these two subfamilies we double labeled DNA probes amplified from the LTR/5' UTR regions. Signals from both subfamilies were mainly located in the centromeric regions of maize metaphase chromosomes. However, the size and intensity of the signals were significantly different in some maize centromeres (Figure 7), suggesting that the elements from the two subfamilies are not uniformly dispersed in these centromeres.

Estimation of the age of the retrotransposons in the centromeric BACs: The two LTRs of a retrotransposon are identical at the time of its insertion into the host genome. If the mutation rate is constant after the transposition, the age of the retrotransposon since transposition can be estimated by the number of substitutions per nucleotide site within the LTRs (SANMIGUEL *et al.* 1998). An average substitution rate at the *adh* locus among grasses was estimated at 6.5×10^{-9} substitutions per synonymous site per year (GAUT *et al.* 1996). This rate was used to estimate the insertion time of the retrotransposons in this study (Table 3).

The insertion timing or the ages of the retrotransposons in BAC clone 16H10 are summarized in Figure 8 and Table 3. Sequence analysis suggests that the insert of BAC 16H10 was an intact CentC DNA fragment. This CentC fragment was separated into three CentC tracts due to retrotransposon invasions. All retroelements within 16H10 are younger than 1.3 million years. Four CRM elements inserted directly into the CentC fragment (Figure 8), but the locations within the CentC 156-bp repeat unit of the four insertions are different, indicating that targeting sites of the CRM elements are not sequence specific.

The insertion timing of the majority of the retrotransposons in BAC 15C5 was difficult to determine due to the significant sequence degeneracy. Only three retrotransposons retained a pair of complete LTRs. One of these three elements, Cinfu2a, is highly rearranged and its structure is difficult to define. The ages of the other two retrotransposons, Cinfu2c and Zeon1a, were estimated to be 2.63 and 42.22 million years, respectively (Table 3).

Organization and divergence of the CentC repeat:

Several large shotgun clones covering the CentC tract regions were sequenced using transposon-based sequencing methods to confirm the sequence and the order of the highly similar CentC monomers. The CentC repeats in the two BAC clones were aligned and grouped by the neighbor-joining method. The CentC repeat sequences can be divided into 18 groups (groups A–R; Figure 9). All the CentC repeats from 15C5 are different from those of 16H10, suggesting that the CentC sequences in these two BACs have significantly diverged. Some of the CentC groups periodically appeared in multiple CentC tracts (Figure 9). For example, a JCFFI motif is observed in both A1 and A2 tracts (Figure 9). The physical gap between tract A1 and A2 may contain CentC repeats with identical sequence and organization patterns to those within tracts A1 and A2. Such CentC sequences may be assembled into the “duplicated regions” in tracts A1 and A2. Similarly, HE, QMRPO, or KRLRR motifs are observed periodically in tracts B and C, D, and E and F, respectively (Figure 9). These results indicate that the CentC sequences have been amplified and maintained by higher-order structures of specific CentC monomers.

The 3' end of CentC tract A2 and the 5' end of CentC tract B are located in the same position in a CentC monomer, suggesting that these two CentC tracts were separated by the insertion of the CRM2a that transposed ~ 1.22 million years ago. Interestingly, CentC tracts A2 and B showed completely different patterns (Figure 9), suggesting that retrotransposon invasion may significantly impact the divergence of the centromeric satellite repeats.

DISCUSSION

DNA sequences located within centromeric regions have been isolated in numerous plant species. However, large-scale sequencing and organization studies of centromeric DNA have been documented in only a few

TABLE 3
Estimated age of CR elements and retrotransposons in BACs 16H10 and 15C5

Name	Clone ^a	Reference	Sites	<i>k</i> ^b	Time ^c
Cinful2c	15C5	This article	607	0.0342	2.63
Zeon1a	15C5	This article	548	0.5488	42.22
Nov1	16H10	This article	2471	0.0024	0.19
CRM1a	16H10	This article	577	0.0000	0.00
CRM1b	16H10	This article	569	0.0090	0.69
CRM1c	16H10	This article	574	0.0071	0.55
CRM2a	16H10	This article	918	0.0158	1.22
CRM2b	16H10	This article	931	0.0011	0.08
CRM2c	16H10	This article	608	0.0000	0.00
CRR	B1108H10 (AP003562)		901	0.0057	0.44
CRR	OSJNBb0023M11 (AC092749)		815	0.0496	3.82
CRR	OSJNBb0052C09 (AC090441)		849	0.0109	0.84
CRR	OSJNBa0045C13 (AC079634)		849	0.0109	0.84
CRR	OSJNBa0034E23 (AC022352)		908	0.0000	0.00
CRR	P0436D06 (AP003054)		787	0.0013	0.10
<i>cereba-1</i>	(AY040832)	HUDAKOVA <i>et al.</i> (2001)	251	0.0040	0.31
<i>cereba-2</i>	(AY040832)	HUDAKOVA <i>et al.</i> (2001)	865	0.0155	1.19
CentA	(AF078917)	ANANIEV <i>et al.</i> (1998)	1233	0.0100	0.77
RCB11-1	(AB013613)	NONOMURA and KURATA (1999)	778	0.0132	1.02
CRR4.4kb	(AY101510)	CHENG <i>et al.</i> (2002)	779	0.0000	0.00

^a GenBank accession numbers are in parentheses.

^b The estimated number of substitutions per nucleotide site.

^c Million years.

plant species. In rice, the central domains of rice centromeres are occupied by a 155-bp satellite repeat CentO (CHENG *et al.* 2002). Surprising sequence similarity between CentO and the CentC satellite in maize was discovered (CHENG *et al.* 2002). The CentO satellite arrays are interrupted irregularly by the CRR elements (CHENG *et al.* 2002) and other retrotransposons (NONOMURA and KURATA 2001). In general, the organization of centromeric DNA in rice, as well as in several other species, including *Beta* species (GINDULLIS *et al.* 2001), barley (HUDAKOVA *et al.* 2001), and *Zingieria biebersteiniana* (SAUNDERS and HOUBEN 2001), are all similar to that of *A. thaliana* and contain mainly satellite repeats and retrotransposons.

Previous work by ANANIEV *et al.* (1998) suggested that maize centromeres also contain a centromere-specific satellite repeat (CentC) and the centromere-specific retrotransposon, which we have named CRM. Molecular and cytological data suggest that some maize centromeres contain very limited amounts of CentC and CRM-related sequences (ANANIEV *et al.* 1998). These results imply that these centromeres may contain additional centromere-specific DNA sequence families. In this study, we sequenced two maize BACs containing the CentC satellite repeat. Sequence analysis revealed that these two BACs exclusively contain satellite repeats and retrotransposons. BAC 16H10 contains retrotransposons both specific and nonspecific to centromeres, while BAC 15C5 contains only retrotransposons that are not specific to centromeres. The results indicate that these

two centromeric DNA fragments were derived from the insertion of retroelements into intact CentC arrays (Figure 3). These findings add additional evidence that satellite repeats and retrotransposons are the main DNA components of plant centromeres.

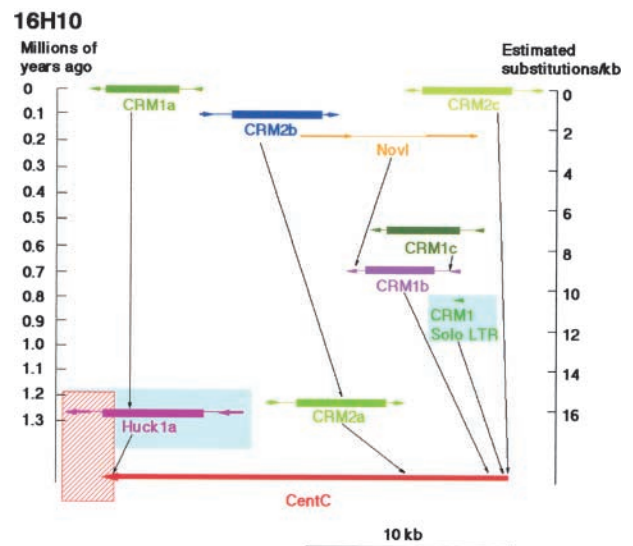


FIGURE 8.—Timing of transposition of the retrotransposons in BAC 16H10. Each retroelement is marked by a different color. The age of the retrotransposons is estimated on the basis of sequence divergence of the two LTRs. The DNA in the red-shadowed box is not cloned in BAC 16H10. The ages of the retroelements within blue boxes are not known.

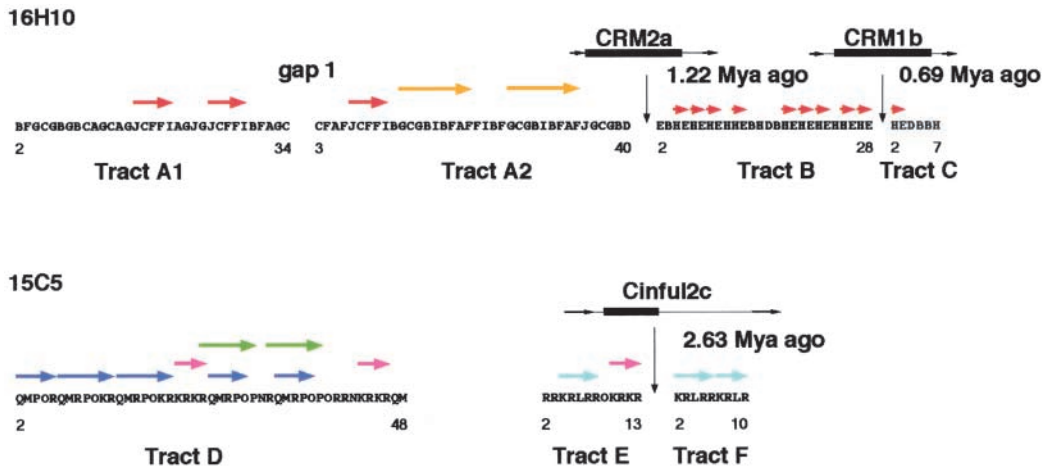


FIGURE 9.—Higher-order structure of the CentC repeats in BACs 16H10 and 15C5. Each subgroup of the CentC monomer is indicated by a different letter and then the subgroups are aligned sequentially. The arrows above the sequence indicate the higher-order repeat.

LANGDON *et al.* (2000) demonstrated that all CR elements reported in grass species were derived from a single ancient family. The CR family has a conventional organization and its protein components are highly conserved even in Arabidopsis homologs (LANGDON *et al.* 2000). Our sequencing results, coupled with the sequenced CRR elements recently deposited in GenBank, provide new data for evolutionary studies of this special retrotransposon family. Phylogenetic analysis demonstrated that the nonautonomous CR elements in both maize and rice are significantly diverged from the full-size CR elements (Figure 5B). The full-size CR elements in maize can be divided into two groups on the basis of sequence similarity analysis (Figure 5, A and B). The most diverged sequences between the two groups are located within the LTRs and 5' UTR. Cytological analyses suggest that the full-size elements from the two groups are not uniformly intermingled at least in some maize centromeres (Figure 7).

The most striking characteristic of this retrotransposon family is its centromere specificity. All the subfamilies in different species have maintained their exclusive centromere locations. The mechanism of this centromere-specific insertion is unknown. In rice, many of the CRR elements inserted either in the CentO satellite repeat or in other CRR elements (CHENG *et al.* 2002), suggesting that the satellite repeat or the CRR element itself may create the conditions such as chromatin confirmation (LANGDON *et al.* 2000) for direct targeting. We found strikingly conserved motifs within the LTRs of the CR elements. Although the grass species were diverged >55 million years ago (KELLOGG 2001), these motifs were found in all the subfamilies (Figure 6). These results suggest that the LTRs may be critical for the centromere-specific transposition.

LANGDON *et al.* (2000) cloned and sequenced PCR products of the CR elements from a number of grass species. The sequence information was expected to provide a basis for estimating the age of individual insertion events, although this would be a substantial underesti-

mate as retrotransposition itself is an error-prone process. A total of 45 reverse transcriptase-encoding clones were obtained from five species and 31 integrase-encoding clones were obtained from eight species. All clones conformed closely to the relevant species consensus, and total variation was in the range of a few percent. The ages of the elements within most species were <1 million years of divergence. LANGDON *et al.* (2000) suggested that the CR family is likely to still be active in most if not all species, while the failure to detect "old" elements implies that either the family is rapidly increasing in abundance at an equivalent rate in each of the divergent species sampled or ancestral sequences are relatively rapidly removed in their entirety before significant levels of degradation occur.

We estimated the age of centromere-specific retrotransposons by comparing the sequences of the two LTRs in individual retrotransposons, an approach more accurate than the method employed by LANGDON *et al.* (2000). All CRM elements discovered in BAC 16H10 transposed within the last 1.22 million years (Table 3 and Figure 8). We also analyzed six CRR elements recently deposited in GenBank. The oldest CRR element transposed 3.82 million years ago, and the other five were transposed within 1 million years. The *cereba* elements in barley transposed between 0.31 and 1.19 million years ago. These data suggest that a majority of the CR elements discovered in all these three species transposed recently, consistent with previous conclusions (LANGDON *et al.* 2000). In contrast, the non-centromere-specific retrotransposons discovered in BAC 15C5 are significantly rearranged, suggesting that these elements transposed much earlier. A *Zeon1a* element present in 15C5 was estimated to be 42 million years old (Table 3). The young age of the CR elements in different grass species suggests that certain parts of the centromeres, possibly the functional domains, are highly dynamic and evolve rapidly at the DNA sequence level. The recent discovery of the interaction between CRM sequences and centromeric histone H3 in maize (ZHONG

et al. 2002) provided the first evidence that the CR elements participate in centromere function and may be a driving force in grass centromere evolution.

We thank Evelyn Hiatt for technical assistance. This research was supported by National Science Foundation grant 9975827 to R.K.D. and J.J.

LITERATURE CITED

- ALFENITO, M. R., and J. A. BIRCHLER, 1993 Molecular characterization of a maize B chromosome centric sequence. *Genetics* **135**: 589–597.
- ANANIEV, E. V., R. L. PHILLIPS and H. W. RINES, 1998 Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc. Natl. Acad. Sci. USA* **95**: 13073–13078.
- 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.
- 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.
- CLARKE, L., 1998 Centromeres: proteins, protein complexes, and repeated domains at centromeres of simple eukaryotes. *Curr. Opin. Genet. Dev.* **8**: 212–218.
- 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.
- DONG, F., J. T. MILLER, S. A. JACKSON, G. L. WANG, P. C. RONALD *et al.*, 1998 Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc. Natl. Acad. Sci. USA* **95**: 8135–8140.
- FRANCKI, M. G., 2001 Identification of Bilby, a diverged centromeric *Ty1-copia* retrotransposon family from cereal rye (*Secale cereale* L.). *Genome* **44**: 266–274.
- FRANZ, P. F., A. ARMSTRONG, J. H. DE JONG, L. D. PARNELL, C. VAN DRUNEN *et al.*, 2000 Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* **100**: 367–376.
- GAUT, B. S., B. R. MORTON, B. C. McCAIG and M. T. CLEGG, 1996 Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *Proc. Natl. Acad. Sci. USA* **93**: 10274–10279.
- GINDULLIS, F., C. DESEL, I. GALASSO and T. SCHMIDT, 2001 The large-scale organization of the centromeric region in *Beta* species. *Genome Res.* **11**: 253–265.
- HARRINGTON, J. J., G. V. BOKKELEN, R. W. MAYS, K. GUSTASHAW and H. F. WILLARD, 1997 Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nat. Genet.* **15**: 345–355.
- HARRISON, G. E., and J. S. HESLOP-HARRISON, 1995 Centromeric repetitive DNA in the genus *Brassica*. *Theor. Appl. Genet.* **90**: 157–165.
- HENNING, K. A., E. A. NOVOTNY, S. T. COMPTON, X.-Y. GUAN, P. P. LIU *et al.*, 1999 Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. *Proc. Natl. Acad. Sci. USA* **96**: 592–597.
- 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.
- IKENO, M., B. GRIMES, T. OKAZAKI, M. NAKANO, K. SAITOH *et al.*, 1998 Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotech.* **16**: 431–439.
- JACKSON, S. A., M. L. WANG, H. M. GOODMAN and J. JIANG, 1998 Application of fiber-FISH in genome analysis of *Arabidopsis thaliana*. *Genome* **41**: 566–572.
- JACKSON, S. A., F. DONG and J. JIANG, 1999 Digital mapping of bacterial artificial chromosomes by fluorescence *in situ* hybridization. *Plant J.* **17**: 581–587.
- JIANG, J., B. S. GILL, G. L. WANG, P. C. RONALD and D. C. WARD, 1995 Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **92**: 4487–4491.
- JIANG, J., S. NASUDA, F. DONG, C. W. SCHERRER, 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.
- KASZAS, E., and J. A. BIRCHLER, 1996 Misdivision analysis of centromere structure in maize. *EMBO J.* **15**: 5246–5255.
- KASZAS, E., and J. A. BIRCHLER, 1998 Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. *Genetics* **150**: 1683–1692.
- KELLOGG, E. A., 2001 Evolutionary history of the grasses. *Plant Physiol.* **125**: 1198–1205.
- KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KISHII, M., K. NAGAKI and H. TSUJIMOTO, 2001 A tandem repetitive sequence located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes. *Chromosome Res.* **9**: 417–428.
- KUMEKAWA, N., T. HOSOUCI, H. TSURUOKA and H. KOTANI, 2000 The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. *DNA Res.* **7**: 315–321.
- 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.
- MALUSZYNSKA, J., and J. S. HESLOP-HARRISON, 1991 Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *Plant J.* **1**: 159–166.
- MARTINEZ-ZAPATER, J. M., M. A. ESTELLE and C. R. SOMERVILLE, 1986 A high repeated DNA sequence in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **204**: 417–423.
- MILLER, J. T., F. DONG, S. A. JACKSON, J. SONG and J. JIANG, 1998a Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**: 1615–1623.
- MILLER, J. T., S. A. JACKSON, S. NASUDA, B. S. GILL, R. A. WING *et al.*, 1998b Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor*. *Theor. Appl. Genet.* **96**: 832–839.
- NAGAKI, K., H. TSUJIMOTO and T. SASAKUMA, 1998 A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. *Chromosome Res.* **6**: 295–302.
- NIZETIC, D., R. DRMANAC and H. LEHRACH, 1990 An improved bacterial colony lysis procedure enables direct DNA hybridization using short (10, 11 bases) oligonucleotides to cosmids. *Nucleic Acids Res.* **19**: 182.
- NONOMURA, K. I., and N. KURATA, 1999 Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. *Mol. Gen. Genet.* **261**: 1–10.
- NONOMURA, K. I., and N. KURATA, 2001 The centromere composition of multiple repetitive sequences on rice chromosome 5. *Chromosoma* **110**: 284–291.
- PAGE, B. T., M. K. WANOUS and J. A. BIRCHLER, 2001 Characterization of a maize chromosome 4 centromeric sequence: evidence for an evolutionary relationship with the B chromosome centromere. *Genetics* **159**: 291–302.
- 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.
- 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.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SANMIGUEL, P., B. S. GAUT, A. TIKHONOV, Y. NAKAJIMA and J. L.

- BENNETZEN, 1998 The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**: 43–45.
- SAUNDERS, V. A., and A. HOUBEN, 2001 The pericentromeric heterochromatin of the grass *Zingera biebersteiniana* (2n=4) is composed of Zbcen1-type tandem repeats that are intermingled with accumulated dispersedly organized sequences. *Genome* **44**: 955–961.
- SHIZUYA, H., B. BIRREN, U. J. KIM, V. MANCINO, T. SLEPAK *et al.*, 1992 Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**: 8794–8797.
- SONG, J., F. DONG and J. JIANG, 2000 Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research. *Genome* **43**: 199–204.
- SUTTON, G. G., O. WHITE, M. D. ADAMS and A. R. KERLAVAGE, 1995 TIGR assembler: a new tool for assembling large shotgun sequencing projects. *Genome Res.* **1**: 9–19.
- TALBERT, P. B., R. MASUELLI, A. P. TYAGI, L. COMAI and S. HENIKOFF, 2002 Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *Plant Cell* **14**: 1053–1066.
- 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.
- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- YUAN, Q., J. HILL, K. MOFFAT, J. HSIAO, Z. CHENG *et al.*, 2002 Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. *Mol. Genet. Genomics* **267**: 713–720.
- ZHONG, C. X., 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: V. SUNDARESAN