The Plant Cell, Vol. 11, 31–42, January 1999, www.plantcell.org © 1999 American Society of Plant Physiologists

Polymorphisms and Genomic Organization of Repetitive DNA from Centromeric Regions of Arabidopsis Chromosomes

John S. Heslop-Harrison,[1,2] Minoru Murata, Yutaka Ogura, Trude Schwarzacher,[1] and Fusao Motoyoshi

Research Institute for Bioresources, Okayama University, Kurashiki 710-0046, Japan

A highly abundant repetitive DNA sequence family of Arabidopsis, AtCon, is composed of 178-bp tandemly repeated units and is located at the centromeres of all five chromosome pairs. Analysis of multiple copies of AtCon showed 95% conservation of nucleotides, with some alternative bases, and revealed two boxes, 30 and 24 bp long, that are 99% conserved. Sequences at the 3′ end of these boxes showed similarity to yeast CDEI and human CENP-B DNA–protein binding motifs. When oligonucleotides from less conserved regions of AtCon were hybridized in situ and visualized by using primer extension, they were detected on specific chromosomes. When used for polymerase chain reaction with genomic DNA, single primers or primer pairs oriented in the same direction showed negligible amplification, indicating a head-to-tail repeat unit organization. Most primer pairs facing in opposite directions gave several strong bands corresponding to their positions within AtCon. However, consistent with the primer extension results, some primer pairs showed no amplification, indicating that there are chromosome-specific variants of AtCon. The results are significant because they elucidate the organization, mode of amplification, dispersion, and evolution of one of the major repeated sequence families of Arabidopsis. The evidence presented here suggests that AtCon, like human α satellites, plays a role in Arabidopsis centromere organization and function.

INTRODUCTION

Understanding the organization, sequence, structure, and function of the DNA at the centromeres of chromosomes of fungi, plants, and animals is of both fundamental and applied importance because of the role of the centromere in chromosome segregation, in karyotypic stability, and in generating artificial chromosomes as cloning or expression vectors. In yeasts, the role of particular sequences and their protein interactions is becoming clear (Uzawa and Yanagida, 1992; Clarke et al., 1993; Hegemann and Fleig, 1993). However, despite these successes, the sequence requirements for the formation of a mammalian centromere remain unclear (Kipling and Warburton, 1997). In plants, various repetitive sequences have been isolated, and in situ hybridization with mitotic chromosome preparations has shown them to localize in broad centromeric regions (see below); however, little more is known.

Core centromeric DNA sequences and the flanking repetitive DNA motifs that are essential for function when reintroduced into the cell have been isolated from both budding yeast (*Saccharomyces cerevisiae*) and fission yeast (*Schizosaccharomyces pombe*), although the difference in average length of the centromere-associated DNA between the two species is enormous (Hegemann and Fleig, 1993). A functional centromere of *S. cerevisiae* is contained within a 125-bp sequence that carries three types of relatively simple protein binding DNA elements (see Clarke, 1990): CDEI (for centromere DNA element I), consisting of eight nucleotides (RTCACRTG, where R is a purine); CDEII, an AT-rich 78- to 86-nucleotide sequence; and CDEIII, a conserved sequence of 26 nucleotides (TGTYTYTGNT TTCCGAAANNNAAAAA, where Y is a pyrimidine, and N is A, T, C, or G). The palindromic core sequence of CDEI (CACGTG) has been shown to be important for in vivo function and in vitro binding of a key protein, Cpf1 (for centromere and promoter factor 1), and presumably allows the protein to bind to the DNA helix in either direction (Wilmen et al., 1994). In *S. pombe*, the structure of the centromere is much more complicated and may be a better model for multicellular eukaryotes; for example, the putative 100-kb functional centromeric region of chromosome 2 contains a specific, nonhomologous 7-kb core that is surrounded by a unique 1.5-kb inverted element and four repetitive DNA elements that occur on all centromeres (Clarke, 1990).

Many studies of mammals have focused on repetitive DNA sequences located at or near the centromeres to identify the sequences responsible for higher eukaryote centromere activity and their protein binding regions (Lee et al., 1997). There is no intellectual consensus regarding the importance of the abundant, highly repetitive (or satellite) sequences found at centromeres: many but by no means all authors (see Kipling and Warburton, 1997) regard them as

[1]Current address: John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK.

[2]To whom correspondence should be addressed. E-mail pat.heslop-harrison@bbsrc.ac.uk; fax 44-1603-456844.

playing key roles in centromere function and chromosome segregation (Tyler-Smith et al., 1998). Despite the functional similarity of the centromere in all mammals, the centromeric repetitive DNA sequences are variable between species (Sunkel and Coelho, 1995). Approximately 300 copies of the centromeric, alphoid, or α satellites in humans have been sequenced by various laboratories (see Choo et al., 1991). The centromere of each chromosome (except for Y and deleted or rearranged abnormal chromosomes) includes tandemly arrayed units of the 170-bp monomer repeat arranged in head-to-tail orientation, and many thousands of units occur in each single array that may be megabases long. Different chromosomes have both sequence variants and characteristic multimers of slightly variant units, giving a chromosome-specific subpattern (Willard, 1985).

Similar chromosome-specific variants have been identified in the centromeric minor satellite of the mouse (Kipling et al., 1994). These satellite sequences bind proteins, with the best characterized of these being CENP-B (for centromere protein B; reviewed in Brinkley et al., 1992), and monomers of many mammalian centromeric satellite sequences contain a degenerate 17-bp-long CENP-B protein binding motif ("box"; Muro et al., 1992). Although the DNA binding site is not highly conserved, the CENP-B protein itself is (Goldberg et al., 1996). The work of Willard and collaborators (Harrington et al., 1997), in which human microchromosomes were made starting from synthetic arrays of α-satellite DNA, provides evidence that the α-satellite DNA itself is the functional centromeric sequence and that an array of binding sites for CENP-B is sufficient to nucleate kinetochore assembly.

Compared with mammals and yeast, there has been relatively little molecular and structural analysis of centromeric DNA in higher plants, although many highly repetitive satellite and nonsatellite sequences have been localized to the centromeric regions of chromosomes by using in situ hybridization (crucifers: Harrison and Heslop-Harrison, 1995; Brandes et al., 1997b; grasses: Kamm et al., 1994; Leach et al., 1995; Aragon-Alcaide et al., 1996; Jiang et al., 1996; Nagaki et al., 1998). Centromeric regions of many other plant species stain strongly with dyes such as 4′,6-diamidino-2-phenylindole (DAPI) or chromomycin A3, showing enhanced AT or GC base pair–dependent fluorescence or altered C banding, and they probably consist of highly repetitive DNA sequence motifs. Although any function of these sequences remains unknown, it has been suggested that, like the human α satellites, centromeric repetitive sequences may be related to centromere function because of their location (Maluszynska and Heslop-Harrison, 1991), their sometimes widespread species distribution, and their high repetition. The inclusion of DNA sequence boxes defined for centromere function in mammals and yeasts is also considered significant.

Sequencing of the genome of Arabidopsis is under way, but major blocks of repetitive DNA are initially excluded until a detailed strategy for sequencing the more complex regions, such as centromeres, is devised (Bevan et al., 1997). Although the centromeres of all five chromosomes have now been mapped (Round et al., 1997; Copenhaver et al., 1998), these regions may not be well represented in the existing libraries. Because of the sequence repetition, clones consisting mainly of repetitive DNA cannot be placed in overlapping units by using straightforward current technologies, and thus, they remain as gaps (Schmidt et al., 1995; Zachgo et al., 1996).

Repetitive sequence families have been identified in Arabidopsis, and one such family, with members that include monomers AS1, AS2, and the dimer AL1, with a monomer ~180 bp long, has been isolated by several groups of researchers (e.g., Martinez-Zapater et al., 1986; Simoens et al., 1988; Murata et al., 1994; see Figure 1). Following Martinez-Zapater et al. (1986), the family is referred to here as AtCon. It constitutes between 2 and 5% of the genome (Murata et al., 1994) and is surpassed only by the 18S–25S rDNA units (constituting 8% of the genome). The centromeric regions of all five chromosome pairs have been shown to contain long tracts of up to 4 Mb of this tandemly repeated sequence, as was determined by in situ hybridization and pulse-field mapping (Maluszynska and Heslop-Harrison, 1991; Murata et al., 1994). Other repetitive DNA elements, including retrotransposons (Pelissier et al., 1996), are clustered near the centromeres and have been shown by DNA fiber in situ hybridization (Brandes et al., 1997a; Heslop-Harrison et al., 1997) and two-dimensional pulse-field gel electrophoresis to intersperse AtCon. Similarly, uninterrupted tandem arrays of AtCon up to 1 Mb long were shown to be flanked by complex interspersed domains (Round et al., 1997). Sequences with high homology to AtCon, the AaKB27 family, have been isolated from the closely related species *Arabidopsis arenosa* (Kamm et al., 1995).

In this study, our goal was to elucidate the structure, variability, and physical organization of AtCon associated with the centromere of Arabidopsis. It is important to know the variation and chromosome specificity of such sequences to (1) understand modes of genome evolution; (2) compare their organization with that of mammalian and yeast sequences associated with centromeres; (3) learn something about functionality from possible conserved regions; and (4) examine their use as chromosomal markers. We also wanted to learn more about the structure and organization of the centromeric regions of Arabidopsis chromosomes that will complement information from the sequencing of other regions of the genome.

RESULTS

Sequence Analysis

Figure 1A shows the sequences of a trimer and a tetramer of AtCon (pAtHR220/2 and pAtHR220/3) and their alignment

with 12 other *A. thaliana* sequences and five *A. arenosa* AaKB sequences from the GenBank, EMBL, and DDBJ databases. The *A. thaliana* AtCon family had a median length of 178 bp and was 64% AT rich, with an equal content of A and C bases on both strands. Alignment of AtCon sequences showed an overall homology level of 94.5% (monomers within the trimer and tetramer were only slightly more homologous), but this value disguised features of special interest. Twelve positions had alternative nucleotides, for which 50 to 73% of the nucleotides were one base, and most of the remaining bases had a second base: no consensus nucleotide could be given at these positions. Furthermore, in AtCon, there was a region of 30 bp, designated box A, with 99.3% conservation (only four variants outside of the three alternative nucleotide positions). Within box A, 13 nucleotides showed 100% conservation. In the five AaKB sequences from *A. arenosa*, box A showed 100% conservation over its 29 bp (including one deletion and five substitutions with respect to AtCon). A second 9-bp conserved region was present within a 24-bp region, box B, with 98.9% conservation (five variant nucleotides). We found no evidence for nonrandom nucleotide changes in several AtCon sequences: within both the alternative and other variant sites, and between the *A. thaliana* and *A. arenosa* sequences, transitional (A↔G or T↔C) changes were not significantly more frequent than transversions (A↔C, T↔G, C↔G, or A↔T; chi square probability $>5\%$). However, it is notable that six of the seven changes in boxes A and B were transitional (expectation, 2.3 of 7; chi square probability of 3%).

Polymerase Chain Reaction Amplification of Genomic DNA

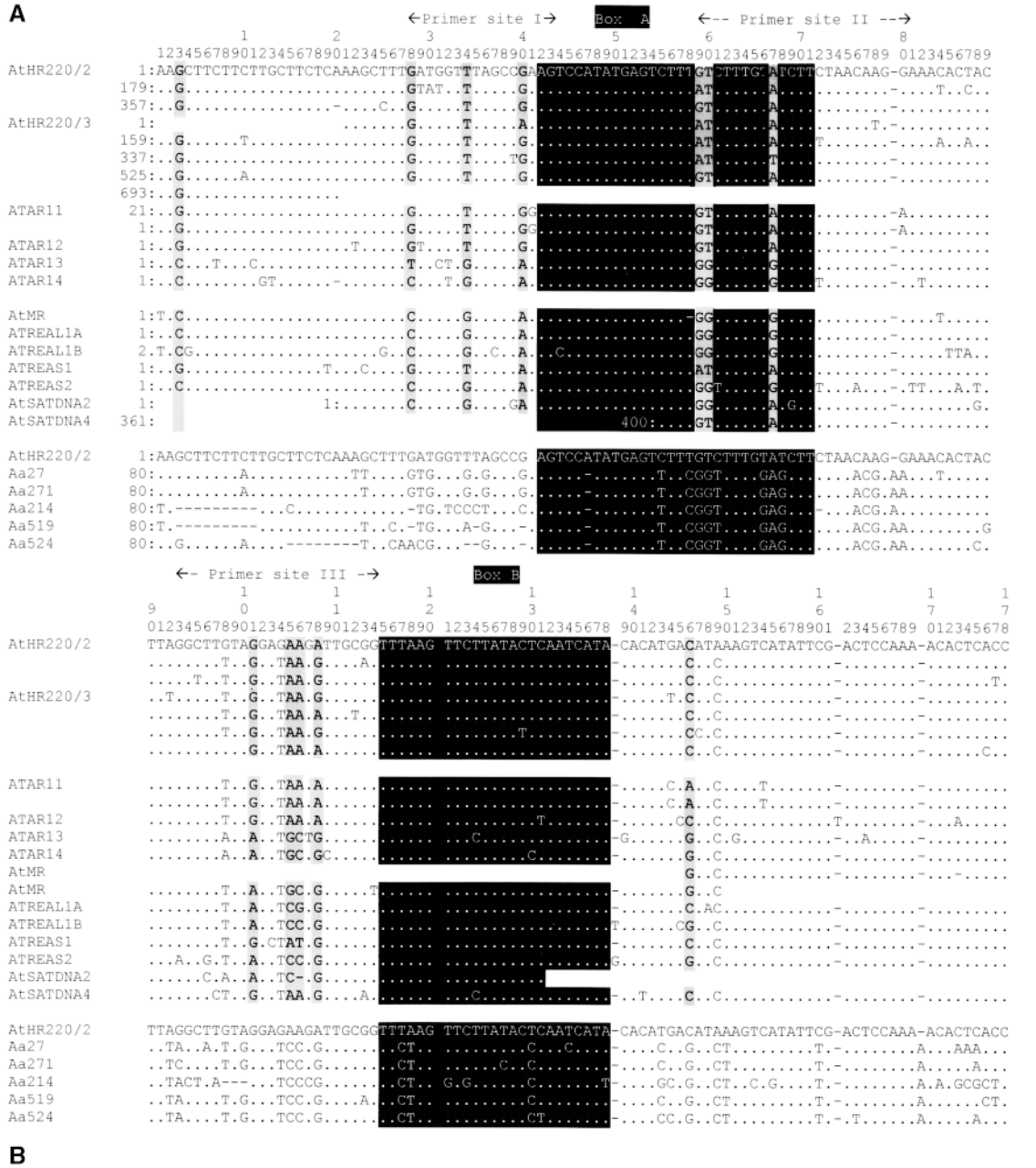
We designed oligonucleotide primers, shown in Figure 1B as primers Cen1 to Cen7, that would hybridize with three different regions, namely, I, II, and III, of AtCon that included alternative nucleotides. We considered that the presence of alternative bases at some positions indicated either preferential changes at these sites or mutation before amplification/homogenization events. The Cen primers were designed to investigate the long-range organization of AtCon variants by using polymerase chain reaction (PCR) with an annealing temperature of 55°C, >10°C above the melting temperature for the 13- and 16-mer primers but similar to the melting temperature for the 22- and 23-mer primers. Results from all primers and primer pairs were conclusive, except when primer Cen5 was used, perhaps because of the lower stringency, the duplication the five 3′ bases of Cen5 in some sequence variants (e.g., ATAR14), or duplication elsewhere in the genome. Attempting amplification from genomic DNA by using PCR with single primers gave no, or almost no, detectable products (Figure 2A), indicating that the repetitive motifs were organized in a head-to-tail and not a head-to-head orientation. PCR with primer pairs in which both primers were made in the same direction also gave no detectable products, but most pairs of primers with reverse orientation to each other (including nearly complementary pairs) gave products (Figure 2B). The length of the shortest product, between 50 and 180 bp, depended on and correlated with the position of the primer binding sites in the sequence; multiple longer products were also generated from amplification of the highly repetitive tandem arrays of AtCon. However, no products were detected with some of the reverse orientation primer pairs (e.g., Cen2 and Cen3; Cen2 and Cen7; Figure 2B), suggesting that the primers were not located nearby in the same tandem repeat block.

In Situ Hybridization and in Situ Primer Extension

In situ hybridization of the full-length clone of pAtMR1 showed strong and approximately equal hybridization to the centromeric regions of all five metaphase chromosome pairs from Arabidopsis (Figure 3A), as expected from previous results (Maluszynska and Heslop-Harrison, 1991; Murata et al., 1994). At the resolution of in situ hybridization to metaphase chromosomes, interruption of the arrays could not be detected. The method uses a labeled DNA probe several hundred base pairs long for detecting sites of homologous DNA sequences on chromosomes (normally with a stringency of 85%). Primed in situ extension involves annealing of primers (here 12 to 27 bp long) to chromosomes and enzymatic template-dependent extension from the annealed primers with incorporation of labeled nucleotides (Kipling et al., 1994; Koch et al., 1995). Thus, primed in situ extension exploits the DNA sequence specificity that is required for primers to anneal effectively in a similar fashion as does diagnostic PCR analysis. This technique has been used for the rapid analysis of chromosome-specific repetitive sequences in humans (Koch et al., 1995) and mouse (Kipling et al., 1994).

Here, we were able to use primed in situ extension as an effective method for determining the chromosomal locations of variants of AtCon. Pretreatment with dideoxy nucleotides for chain termination was helpful for increasing contrast between strongly labeled and other sites: this might be expected because the acids used in fixation and chromosome spreading and the DNases present in the crude enzymes used for chromosome preparation cause single- and double-stranded nicks in the DNA. After chromosomal denaturation, some sequences reanneal and provide starting points for polymerase extension. This is particularly true for highly repeated sequences. Termination of these unspecific DNA extension sites with dideoxy nucleotides, followed by washing, the addition of either no primer or an oligomer known not to be abundant in the *Arabidopsis* spp genome (e.g., the sequence at the centromeres of *Brassica* spp; Figure 1B), and extension with incorporation of labeled nucleotides gave only weak labeling of centromeric DNA sites (not shown).

Examples of flower bud metaphase and interphase chromosomes with in situ extension from primers Cen1 to



B

Primer	Position (bp)	Site in Fig 1A	Present in (bp)	Length	Sequence
Forward orientation (with respect to Fig. 1A):					
Cen1	93-108	III	ATMR	16	5′ GGC TTT TAA GAT GCG G 3′
Cen2	93-108	III	AtHR220/2, AtHR220/3	16	5′ GGC TTT TAG GAT AAG G 3′
Cen4	28-40	I	AtHR220/2, AtHR220/3, ATAR11	13	5′ GAT GGT TTA GCC G 3′
Cen5	93-114	III	ATAR11, ATAR12	22	5′ GGC TTT TAG GAT AAG ATT GCG G 3′
Cen8	42-69	Box A	Conserved	27	5′ GTC CAT ATG AGT CTT TGK CTT TGT RTC 3′
Reverse orientation:					
Cen3	108-93	III	ATREAL1B	16	5′ CCG GAT CTT AAA AGC C 3′
(Complement					*5′ GGC TTT TAA GAT CCG G 3′)*
Cen6	81-59	II	ATMR, ATREAL1A, ATREAL1B, ATAR13	23	5′ CCT TGT TAG GAA GAC ACA AAG CC 3′
(Complement					*5′ GGC TTT GTG TCT TCC TAA CAA GG 3′)*
Cen7	43-28	I	ATMR, ATREAL1A, ATREAS2	16	5′ CTT TGG CTA CAC CAT G 3′
(Complement					*5′ CAT GGT GTA GCC AAA G 3′)*
Cen9	76-51	Box A	Conserved	26	5′ GTT AGA AGA YAC AAA GMC AAA GAC TC 3′
(Complement					*5′ GAG TCT TTG KCT TTG TRT CTT CTA AC 3′)*
Brassica centromere sequence					
			BcKB6	12	5′ GAT GAA GAA GTT 3′

Figure 1. The AtCon Sequence Family and Primers.

Cen6, homologous to three regions of AtCon that include alternative bases (Figure 1B), are shown in Figures 3 and 4. Although there was some variability among cells, each of these primers showed on average two strong and a specific number of weaker signals of varying intensity at the centromeres of different groups of chromosomes. Most often, the area of the chromosome covered by the primed in situ extension signal was not as large as the signal generated by in situ hybridization with the whole AtCon sequence (e.g., cf. Figures 4A to 4F and 3A); in some cases it showed gaps. Cen7 and Cen8 were homologous to parts of box A; unlike Cen1 to Cen6, primer extension from these showed approximately equally strong fluorescence at all centromeric regions. It was normally possible to distinguish the three metacentric and the larger and smaller acrocentric chromosome pairs based on morphology. Digital images showing relative fluorescence of the hybridization sites of five to 15 metaphase chromosomes per primer were compared systematically with each other by using Adobe PhotoShop (see Methods). The average intensity of signal associated with suggested chromosome morphologies can be summarized as follows.

Cen1 (Figure 3E) showed two strong signals, not on the small acrocentric chromosome pair (one signal was often stronger than the other); three or four weak to medium sites were detected, and the signal was diffuse and extended over a large area.

Cen2 (Figures 3F and 3H) showed two strong signals on a metacentric chromosome pair and two or three medium signals; the remaining chromosomes showed at most a weak signal.

Cen3 (Figures 3C to 3D) showed two strong signals on metacentric chromosomes and six chromosomes with weak signals. Most signals extended over a large area and showed a clear gap in the middle.

Cen4 (Figures 4A to 4C) showed two strong signals on the larger acrocentric chromosome and four medium to weak signals. In extended prometaphase chromosomes (Figure 4B), the in situ extension signal was located between two major DAPI-positive heterochromatic blocks (compare the size with the in situ hybridization signal using the whole clone in Figure 3A).

Cen5 (Figures 4D and 4E) showed two strong sites on a metacentric chromosome pair, two medium to weak signals, and six chromosomes with weak to no signal.

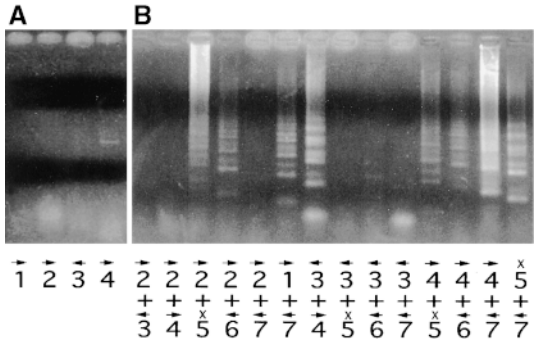


Figure 2. Amplification of Genomic Sequences from Arabidopsis by Using PCR.

(A) Single Cen primers.

(B) Primer pairs.

Primers and primer pairs used in the amplification reaction are as indicated under the lanes (1 to 7 for Cen1 to Cen7) and are described in Figure 1B. Arrows indicate the direction of primers with respect to Figure 1A. Results with forward orientation primer 5, indicated by x's, were not interpretable (see text). With single primers and primer pairs in the same orientation, no or very weak amplified products were seen, whereas most primer pairs with reverse orientation gave a series of strong, defined bands and a high molecular mass smear. Products correspond to expected lengths, depending on primer positions within the tandem repeat unit making up the arrays. The shortest products for Cen2 and Cen6 (2+6) are 166 and 344 bp; Cen1 and Cen7 (1+7), 126 and 304 bp; Cen3 and Cen4 (3+4), 80 and 258 bp; Cen4 and Cen6 (4+6) (52 bp not seen), 220 and 398 bp; and Cen4 and Cen7 (4+7), 194 and 372 bp. No products were found with reverse orientation primer pairs Cen2 and Cen3 (2+3) or Cen2 and Cen7 (2+7).

Figure 1. (continued).

(A) Sequence of the AtCon family of clones from the Arabidopsis centromere sequence. Dots indicate conserved nucleotides; dashes indicate gaps introduced to maintain alignment; multiple spaces are at the ends of sequenced regions (two HindIII sites, often used for cloning, are at base pairs 1 and 21) or where sequence similarity is not apparent. Hatching shows nucleotides at which there is no clear consensus nucleotide at a position, and the highly conserved boxes A and B are shown in black. The following sequences were analyzed from *A. thaliana* Columbia ecotype: AtHR220/2 and AtHR220/3 (this study); ATAR11, 12, 13, and 14, tandemly repeated sequences AR11, AR12, AR13, and AR14 (Simoens et al., 1988); AtMR, centromeric repetitive sequence AT53212 (Murata et al., 1994); ATREAL1A and 1B, repetitive DNA, AL1a and AL1b (Martinez-Zapater et al., 1986); ATREAS1 and 2, repetitive DNA AS1 and AS2 (Martinez-Zapater et al., 1986); and AtSATDNA2 and 4, DNA of a 180-bp satellite junction, DNA2 and DNA4 (Pelissier et al., 1996). Aa27, 271, 214, 519, and 524 are from *A. arenosa* centromeric sequence (Kamm et al., 1995). The five Aa sequences have been divided at a HindIII site and fused at the BamHI site (GGATCC) used for cloning.

(B) Primers used for genomic DNA amplification and in situ primer extension. Their positions with the repeat unit and sequence-of-origin are shown: K is G or T; and M is A or C (text within parentheses and in italics shows the complements of primers to facilitate comparison with the strand in **[A]**).

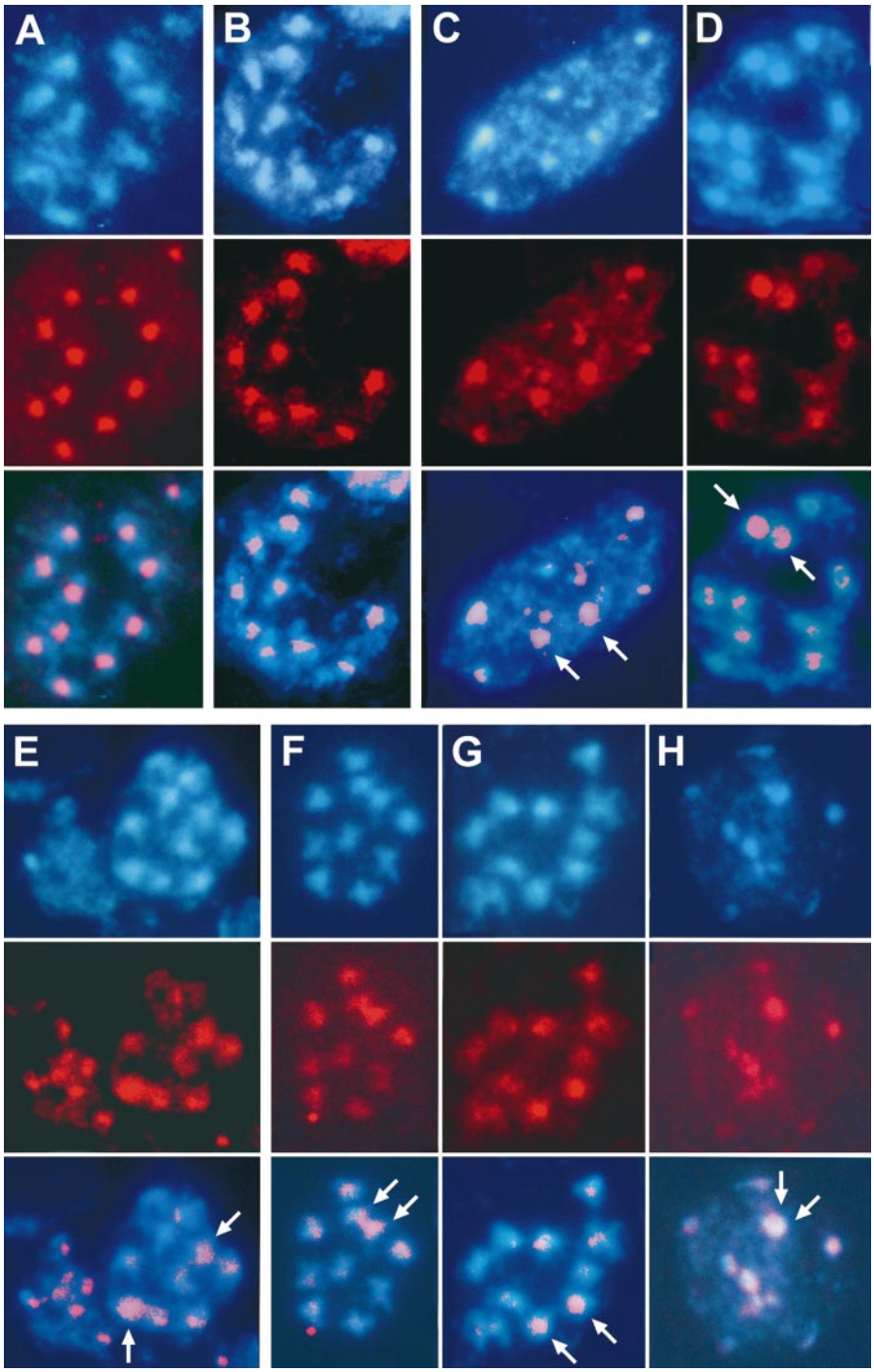


Figure 3. Fluorescence in Situ Hybridization and in Situ Primer Extension to Chromosome Preparations of Arabidopsis.

Blue DAPI staining after UV excitation is shown above the red Cy3 signal and an overlay of both images. Magnification is ×6000.
(A) pAtMR1, a member of the AtCon family, hybridized with the centromeres of all five metaphase chromosome pairs with approximately equal strength and extended over the large area of DAPI-positive centromeric heterochromatin.
(B) The conserved primer Cen8 primed labeled nucleotide incorporation in the centromeric regions of all chromosomes, which can be visualized by the bright red signal.

Cen6 (Figures 4F to 4I) showed two strong signals on acrocentric chromosomes (most likely the small acrocentric pair); two metacentric chromosomes had a medium signal, whereas two acrocentric chromosomes had a weak signal; the remaining four metacentric chromosomes showed very little signal. In a late zygotene nucleus (Figure 4I), centromeric regions were distinguished by bright DAPI staining along the axial core of the chromosomes, and most of the paired centromeric region of one axis was labeled strongly.

Cen7 and Cen8 (Figure 3B) showed strong and approximate signals at the centromeric regions of all five chromosome pairs.

At interphase, up to 10 labeled foci could be detected (e.g., Figure 4C), but often signals were close to each other or fused (e.g., Figures 4C, 4E, and 4H). It was also apparent that the primed in situ extension signal intensity did not always correspond to the size of DAPI chromocenters at interphase.

DISCUSSION

By using a combination of DNA sequence analysis and molecular and cytogenetic techniques, we show important and novel features of the organization, chromosomal distribution, and evolution of the major repetitive sequence, AtCon, present at the centromeres of Arabidopsis chromosomes. By using in situ hybridization, other researchers have shown that AtCon sequences are located on all five chromosome pairs of Arabidopsis (Maluszynska and Heslop-Harrison, 1991; Murata et al., 1994; see Figure 3A); however, because the stringency of hybridization is typically 85%, the differential distribution of minor variants has not been assayed. Analysis of 19 AtCon repeat units from *A. thaliana* and five units of the homologous sequence from *A. arenosa* demonstrated key features of AtCon (Figure 1A). First, the overall sequence divergence between the different sequenced units of median length of 178 bp, all isolated from the same ecotype Columbia, was very low at 5%. The variation between copies of the mouse centromeric satellite is similarly low (Kipling et al., 1994). The homogeneity is in contrast to the human α-satellite sequences, which show up to 40% variation; when many cloned α satellites are used directly as probes for in situ hybridization or DNA gel blot analysis, one or a small number of chromosomes have been detected (see Willard, 1985; Choo et al., 1991).

Second, within AtCon, we have found two highly (99%) conserved boxes (A and B; Figure 1B) and sites with a higher degree of divergence that are characterized by the preference for alternative bases (I, II, and III; Figure 1B). In situ primer extension with specific primers from the conserved region labeled all centromeres equally (Figure 3B). Primer extension with primers from regions I, II, and III showed strong labeling at the centromeres of different chromosome pairs, and a subset of chromosomes with a weaker signal, depending on the primer used (Figures 3 and 4), which indicates that different variants of AtCon are amplified on some chromosomes more than others and that chromosome-specific variants could exist. PCR analysis using single primers or same-direction primer pairs showed no major amplification products, confirming the view that the tandemly repeated monomers are present in only a head-to-tail organization. This contrasts with some subtelomeric sequences in rye, in which similar single-primer reactions generate major products because of the head-to-head sequence organization arising from chromatid-type breakage-fusion-bridge cycles in distal regions of the chromosomes (Vershinin et al., 1995).

The data from reverse primer pairs show that many primer pairs are present together on chromosomes. The lack of products with some Cen primer pairs indicates that some variants of the sequence are not present within a distance amplifiable by PCR or are not present together on a single chromosome. The latter interpretation would be consistent with the in situ primer extension data showing that some primers are amplified on particular chromosome pairs or groups.

AtCon sequences on each chromosome have either become homogenized independently or were distributed among the chromosomes and then amplified with specific variants on each chromosome or group of chromosomes. It has

Figure 3. (continued).

(C) and **(D)** An interphase and a metaphase nucleus, respectively, after in situ extension with primer Cen3. Two stronger signals are visible (arrows); several minor sites with slightly less fluorescence were detected.

(E) An interphase (left) and a metaphase nucleus (right) are shown after in situ extension with primer Cen1. Two strong sites (arrows) and three medium sites are visible at the centromeres of the metaphase chromosomes. The signal is diffuse and extends over much of the chromosomes. The interphase nucleus shows seven sites.

(F) and **(G)** show metaphase chromosome plates after in situ extension with primer Cen2; each showed two strong signals (arrows), possibly on metacentric chromosomes; there are three medium-strength sites.

(H) At interphase, the Cen2 extension signal corresponds to locations of the DAPI-positive chromocenters, but the weaker sites cover only part of the DAPI-positive sites (arrows); some loci are close together or fused.

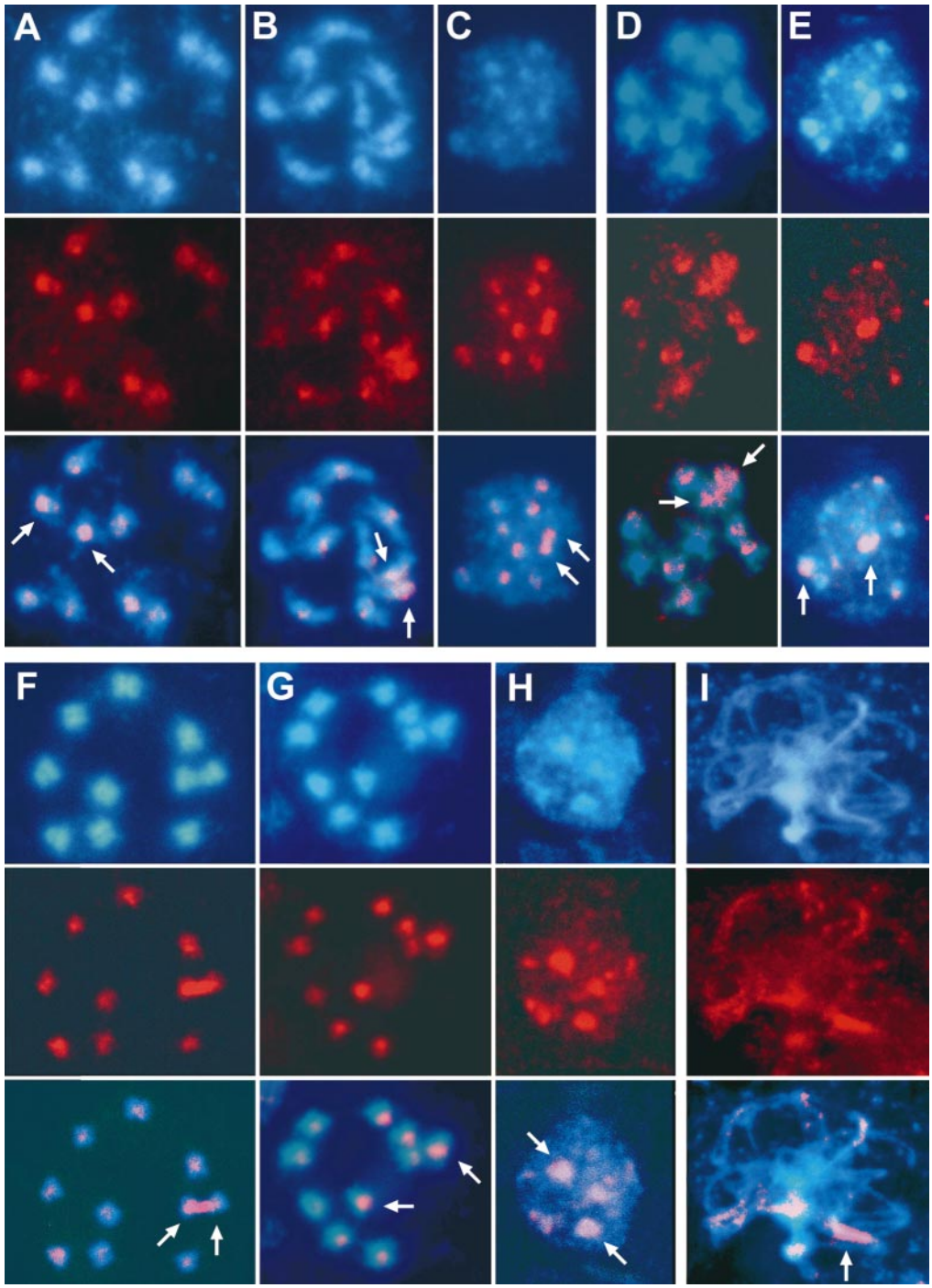


Figure 4. In Situ Primer Extension to Preparations of Arabidopsis Chromosomes.

(A) to **(C)** Metaphase, prometaphase, and interphase chromosomes, respectively, after in situ extension with primer Cen3. Two, probably acrocentric, chromosomes show strong labeling at their centromeres (arrows), four show medium labeling, and the remaining show weak to no labeling. At interphase, the signal colocalizes with the DAPI chromocenters, but on the extended prometaphase chromosomes, the signal is visible between two strongly stained DAPI regions that seem to flank the centromeric constriction.
(D) and **(E)** In metaphase, two strong signals (arrows) were detected with primer Cen5, most likely on a metacentric chromosome. The remaining signals are considerably weaker. At interphase, many DAPI chromocenters do not show in situ extension signal with primer Cen5.

been suggested that variants in satellite sequences can become amplified and spread through arrays, although the exact mechanisms underlying this process remain controversial (Dod et al., 1989); the processes attributed to molecular drive could lead to homogenization of variant sequences (Dover, 1982). The low sequence variation in mouse minor satellites (Kipling et al., 1994) has been taken to imply that the rate of exchange between nonhomologous chromosomes in mouse is particularly high relative to the rate between homologous chromosomes. Alternatively, sequence homogeneity may simply reflect common ancestry or slow accumulation of variants.

Based on results from bivariate flow cytometric analysis of plant chromosomes, Schwarzacher et al. (1997) have commented on the similar ratio of AT to GC base pairs in all the chromosomes of a species, in marked contrast to mammalian genomes, and suggest that there is relatively rapid, genome-wide sequence homogenization in plants. The presence of chromosome-specific variants of AtCon, perhaps like those for soybean (Morgante et al., 1997), indicates that intrachromosomal homogenization or amplification mechanisms are more rapid than are interchromosomal mechanisms for some sequences, although high-stringency in situ hybridization of different sequence variants in *Beta procumbens* shows that two variants may locate in different positions in single arrays (Schmidt and Heslop-Harrison, 1996).

Whether specific nucleotide positions are involved in the variation seen in AtCon (outside the conserved boxes) or whether variable positions are located at random is not known. Active processes of C→T transitional mutagenesis have been reported for fungi (Selker et al., 1993) and mammals (Steinberg and Gorman, 1992), which may be mediated by chromosome pairing or through cytosine methylation. In Neurospora, Centola and Carbon (1994) examined the centromere-specific repetitive DNA sequences and considered that the high sequence divergence, high AT content (65%), and predominantly transitional differences between sequences suggested operation of the repeat-induced point mutation (RIP) system (Selker et al., 1993). AtCon is also 65% AT rich; however, outside the conserved boxes, no nucleotide exchanges were significantly more frequent than any other. Scheid et al. (1994) found no RIP-like phenomena in repeated genes inserted into *A. thaliana*, and like them, we conclude that any RIP-like process in AtCon occurs with low frequency or is absent in the species or perhaps is confined to certain sequence types or boxes.

Many sequences reported at centromeres have relationships with transposable elements. Kipling and Warburton (1997) discuss the similarity of the CENP-B protein with transposase proteins and of the CENP-B binding site with the *Tigger2* terminal inverted repeat. In cereals, both Aragon-Alcaide et al. (1996) and Jiang et al. (1996) have reported centromeric sequences with high homology to retrotransposons. In Arabidopsis, AtCon is interspersed with *Ty1-copia*–like elements, as shown by both sequencing (Pelissier et al., 1996) and in situ hybridization with chromosome and DNA fiber spreads (Brandes et al., 1997b); the two AtSat sequences (isolated by Pelisser et al. [1996] and shown in Figure 1A) are adjacent to fragments of retroelements. Nevertheless, in many species, we know that centromeric and other heterochromatin is depleted in retrotransposons compared with the rest of the genome (Brandes et al., 1997a; Heslop-Harrison et al., 1997).

Our results showing the nonrandomness in sequence mutation sites and chromosomal dispersion indicate that, similar to centromeric satellite sequences in mammals, AtCon is functionally important. Box A corresponds to a box in the related family sequence AaKB27 from *A. arenosa* that shows complete conservation in five published sequences (Figure 1A), giving further support to a biological reason for conservation. Despite wide species conservation in proteins binding the sequence boxes in mammals, the boxes themselves are not well conserved, and considerable effort has been made to define the nucleotides with critical effects and those that influence the binding of centromere proteins, such as human CENP-B (Sugimoto et al., 1998) and yeast CDEs (see Clarke, 1990).

In AtCon, both boxes A and B show notable similarities to published binding boxes. Sugimoto et al. (1998) present the possible palindromic CENP-B consensus binding box YYYYGTTNNNAACRRRR as a hypothesis consistent with much of the data (although their new results did not support aspects of the idea). The 3′ end AtCon box A, from base 61, includes YYYYGT[RTCTTCT]AACRRRR. The motifs outside the brackets are identical to the flanking regions of the hypothesized box, show <99% conservation in AtCon (Figure 1A), and have the potential to form complex secondary structures or bind proteins in either orientation. Although

Figure 4. (continued).

(F) and **(G)** Two metaphase chromosome plates after in situ extension with primer Cen6. Two strong sites on the small acrocentric chromosomes that are often fused were detected (arrows). Four metacentric chromosomes (arrows) show very little signal, whereas the remaining chromosomes have a medium to weak signal.

(H) At interphase, the two strong signals (arrows) were often dispersed over a large area.

(I) In a late zygotene meiotic nucleus, Cen6 extension can be seen strongly over the whole of one paired centromeric region (arrow), whereas weaker sites are seen on other centromeres.

their function has not been tested, other sequences located at the centromeres of plants have been found to include CENP-B–like boxes (Aragon-Alcaide et al., 1996; Nagaki et al., 1998). In one variant of AtCon, AtREAL1B, all eight bases (RTCACRTG) of CDEI of yeast (Clarke, 1990; Wilmen et al., 1994) are present from the 3′ end of box B (position 138), and in the remaining clones, the six critical 3′ bases CACRTG are found. Outside this region, similarities to both CDEII and CDEIII are present in AtCon but might be expected because of the AT richness of both.

In conclusion, we were able to correlate sequence variations of AtCon with chromosome specificity, defined two highly conserved boxes, and analyzed sequence features. There are strong similarities in the conserved boxes to known centromere protein binding boxes in the animal and fungal kingdoms, although many aspects of large-scale plant genome organization differ markedly from mammals (such as the widespread occurrence of polyploidy or relative uniform AT/GC ratio of all plant chromosomes), so indiscriminate use of the mammalian centromere DNA model may not be fully justified. However, as in studies of human α-satellite sequences, we suggest that the AtCon sequences are important for centromere organization and function and might be involved in kinetochore protein binding and chromosome segregation. Furthermore, our study has implications for understanding modes of plant DNA sequence evolution and dispersion and for deducing functionality from conserved regions of highly repetitive DNA sequences. The existence of chromosome-specific variants may facilitate the characterization of the full sequence of the centromeric regions and provide chromosome-specific tags to assist chromosome identification.

METHODS

Plant Material

Arabidopsis thaliana ecotype Columbia was used for all experiments.

Clones, Sequence Analysis, Primers, and Polymerase Chain Reaction Analysis

Sequences with high homology and similar monomer length to the pAtMR1 sequence of Arabidopsis (Murata et al., 1994) were identified in the GenBank, EMBL, and DDBJ databases. Additional clones of a homologous trimer (pAtHR220/2) and tetramer (pAtHR220/3), which were cloned from a partial HindIII digest of genomic Arabidopsis DNA, as reported by Murata et al. (1994), were identified and sequenced in both directions by using an ALF (Pharmacia) sequencer. Sequence comparisons (Figure 1A) were made using the GenBank, EMBL, and DDBJ databases. The oligomers indicated in Figure 1B were synthesized for use as primers. *Brassica* spp sequences located at or near the centromere were used as controls: a 12-mer from *B. nigra* (a centromere-located sequence) and clone pBcKB1 from *B. campestris* (Harrison and Heslop-Harrison, 1995). For amplification by using the polymerase chain reaction (PCR), 500 nM of single or each of a pair of primers was added to 1 μg of genomic DNA isolated from leaves together with Taq polymerase (Takara Pharmaceuticals, Otsu, Japan) and PCR buffer, according to the manufacturer's instructions. After initial denaturation for 4 min at 94°C, amplification was for 30 cycles of 30 sec at 94°C, 30 sec at 55°C, and 30 sec at 72°C, with final primer extension at 72°C for 2 min. Products were separated on a 1% agarose gel by electrophoresis, stained with ethidium bromide, and analyzed under UV light.

In Situ Hybridization and in Situ Primer Extension along Chromosomes

Chromosome preparations were made from immature floral buds by using techniques modified from Murata et al. (1994). Briefly, buds were pretreated with 0.2 mM 8-hydroxyquinoline for 1 to 2 hr at room temperature followed by 1 to 2 hr at 4°C and fixed in 3:1 100% ethanol–glacial acetic acid for at least 4 hr. Buds were digested with a 2% pectinase–cellulase mixture for 90 to 120 min at 37°C and transferred to 45% acetic acid and subsequently to 60% acetic acid. Individual pistils or anthers were dissected, macerated, and squashed on a glass slide in a drop of 60% acetic acid. For in situ hybridization, preparation of biotinylated clone pAtMR1 (40 to 60 ng per slide) in 50% formamide and 2 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M Na-citrate), combined denaturation of the probe and chromosomal DNA at 80 to 85°C for 10 min by using a modified thermal cycler, reannealing at 37°C, and stringent washing (20% formamide in 0.1 × SSC at 40°C) was as given by Maluszynska and Heslop-Harrison (1991).

Primer extension was modified from the method of Kipling et al. (1994). Chromosome preparations on slides were fixed in freshly prepared 4% paraformaldehyde in water, denatured, and subjected to template-dependent extension of annealed regions of double-stranded DNA into single-stranded regions. When two rounds of extension were used, the first round included one base as a dideoxynucleoside-5′-triphosphate (ddNTP) in the extension solution and extended from sites of nicking and self-annealing ("self-priming") until a terminating ddNTP was incorporated. Round 2 was used for specific extension with addition of a synthetic primer and labeled base.

The primer extension buffer consisted of 5 mM Tris-HCl, pH 8.3, 25 mM KCl, 0.005% gelatine, 0.005% Tween 20, 0.005% Triton X-100, 0.05% Nonidet P-40, and 0.75 mM $MgCl_2$. For extension round 1, each slide preparation was covered with 40 μL of buffer containing 1.5 units of Taq polymerase and 0.1 mM of each nucleoside; one was ddNTP (ddCTP, ddGTP, or ddATP, depending on the experiment), and the other three were monodeoxyribonucleoside-5′-triphosphates (dNTPs). After covering the slides with plastic cover slips, we heated them to 85°C for an 8-min denaturation, left them at 37°C for 30 min of reannealing, and then heated them to 72°C for a 30-min extension/termination in a modified thermal-cycling instrument (Heslop-Harrison et al., 1991). Slides were then washed in 2 × SSC and dehydrated through an ethanol series before air drying.

For round 2, the labeling mixture consisted of 40 μL of buffer containing 1.5 units Taq polymerase, 5 ng of Cen primer (Figure 1B), 2 μM biotin-11-dUTP, and 40 μM each of dCTP, dGTP, and dATP. After we covered the slides with plastic cover slips, they were primer annealed at 55°C for 40 min, followed by 72°C for 40 min for primer extension. Slides were usually left in a humid chamber at 4°C overnight at this stage. Single-round extension using the labeling mixture and primer as described above was performed with denaturation at

85°C for 5 min and annealing at 42°C for 45 min, followed by extension at 55°C for 40 min. Controls for both sets of experiments included no primer, Cen8 and Cen9, found in the conserved boxes, and the Brassica centromere sequence (Figure 1B) primer, which is not known to be homologous to Arabidopsis.

Detection of biotin-labeling sites after in situ hybridization or primer extension was performed with streptavidin–Cy3 (Sigma) 1:400 diluted in 5% (w/v) BSA in 4 × SSC and 0.2% Tween 20, as previously described (Maluszynska and Heslop-Harrison, 1991). Preparations were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; 2 μg/mL) and mounted in an antifade solution, *p*-phenilendiamine (in 1 mg/mL water) diluted 1:9 in 1,4-di-azobicyclo-(2,2,2)-octane. Preparations were analyzed with a Zeiss (Oberkochen, Germany) epifluorescence microscope with suitable filters for DAPI and Cy3 and photographed on Fujicolor SuperHG 400 ASA print film (Tokyo, Japan). Color figures and overlays were prepared from digitized images of the film negatives by using Adobe (Mountain View, CA) PhotoShop, with only those processing functions that could be applied equally to all pixels of the image being used. The "image adjust level" function, giving a histogram of the intensity of fluorescence for the pixels, was used to compare and quantify signals. Between five and 15 metaphase chromosomes from one to three slides were quantified after extension with each of six different primers, and other metaphase and interphase chromosomes were photographed and examined.

ACKNOWLEDGMENTS

J.S.H.-H. is grateful for a visiting professorship at Okayama University, and T.S. thanks the Biotechnology and Biological Sciences Research Council for support. We thank the Japanese Society for the Promotion of Science, the British Council, and the Royal Society for their support of this ongoing collaboration.

Received July 20, 1998; accepted October 21, 1998.

REFERENCES

Aragon-Alcaide, L., Miller, T., Schwarzacher, T., Reader, S., and Moore, G. (1996). A cereal centromeric sequence. Chromosoma **105,** 261–268.

Bevan, M., et al. (1997). Objective: The complete sequence of a plant genome. Plant Cell **9,** 476–478.

Brandes, A., Heslop-Harrison, J.S., Kamm, A., Kubis, S., Doudrick, R.L., and Schmidt, T. (1997a). Comparative analysis of the chromosomal and genomic organization of *Ty1-copia*–like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol. Biol. **33,** 11–21.

Brandes, A., Thompson, H., Dean, C., and Heslop-Harrison, J.S. (1997b). Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. Chromosome Res. **5,** 238–246.

Brinkley, B.R., Ouspenski, I., and Zinkowski, R.P. (1992). Structure and molecular organization of the centromere–kinetochore complex. Trends Cell Biol. **2,** 15–21.

Centola, M., and Carbon, J. (1994). Cloning and characterization of centromeric DNA from *Neurospora crassa.* Mol. Cell. Biol. **14,** 1510–1519.

Choo, K.H., Vissel, B., Nagy, A., Earle, E., and Kalitsis, P. (1991). A survey of the genomic distribution of α satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. **19,** 1179–1182.

Clarke, L. (1990). Centromeres of budding and fission yeasts. Trends Genet. **6,** 150–154.

Clarke, L., Baum, M., Marschall, L.G., Ngan, V.K., and Steiner, N.C. (1993). Structure and function of *Schizosaccharomyces pombe* centromeres. Cold Spring Harbor Symp. Quant. Biol. **58,** 687–695.

Copenhaver, G.P., Browne, W.E., and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. Proc. Natl. Acad. Sci. USA **95,** 247–252.

Dod, B., Mottez, E., Desmarais, E., Bonhomme, F., and Roizes, G. (1989). Concerted evolution of light satellite DNA in genus *Mus* implies amplification and homogenization of large blocks of repeats. Mol. Biol. Evol. **6,** 478–491.

Dover, G.A. (1982). Molecular drive: A non-Darwinian model of evolution. Nature **299,** 111–117.

Goldberg, I.G., Sawhney, H., Pluta, A.F., Warburton, P.E., and Earnshaw, W.C. (1996). Surprising deficiency of CENP-B binding-sites in African-green monkey α-satellite DNA—Implications for CENP-B function at centromeres. Mol. Cell. Biol. **16,** 5156–5168.

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

Harrison, G.E., and Heslop-Harrison, J.S. (1995). Centromeric repetitive DNA in the genus *Brassica.* Theor. Appl. Genet. **90,** 157–165.

Hegemann, J.H., and Fleig, U.N. (1993). The centromere of budding yeast. Bioessays **15,** 451–460.

Heslop-Harrison, J.S., Schwarzacher, T., Anamthawat-Jónsson, K., Leitch, A.R., Shi, M., and Leitch, I.J. (1991). In situ hybridization with automated chromosome denaturation. Technique **3,** 109–115.

Heslop-Harrison, J.S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A.V., Alkhimova, E.G., Kamm, A., Doudrick, R.L., Schwarzacher, T., Katsiotis, A., Kubis, S., Kumar, A., Pearce, S.R., Flavell, A.J., and Harrison, G.E. (1997). The chromosomal distributions of Ty*1-copia* group retrotransposable elements in higher plants and their implications for genome evolution. Genetica **100,** 197–204.

Jiang, J., Nasuda, S., Dong, F., Scherrer, C.W., Woo, S.-S., Wing, R.A., Gill, B.S., and Ward, D.C. (1996). A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc. Natl. Acad. Sci. USA **93,** 14210–14213.

Kamm, A., Schmidt, T., and Heslop-Harrison, J.S. (1994). Molecular and physical organization of highly repetitive undermethylated DNA from *Pennisetum glaucum.* Mol. Gen. Genet. **244,** 420–425.

Kamm, A., Galasso, I., Schmidt, T., and Heslop-Harrison, J.S. (1995). Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species. Plant Mol. Biol. **27,** 853–862.

Kipling, D., and Warburton, P.E. (1997). Centromeres, CENP-B and *Tigger* too. Trends Genet. **13,** 141–145.

Kipling, D., Wilson, H.E., Mitchell, A.R., Taylor, B.A., and Cooke, H.J. (1994). Mouse centromere mapping using oligonucleotide probes that detect variants of the minor satellite. Chromosoma **103,** 46–55.

Koch, J., Hindkjaer, J., Kolvraa, S., and Bolund, L. (1995). Construction of a panel of chromosome-specific oligonucleotide probes (PRINS-primers) useful for the identification of individual human chromosomes in situ. Cytogenet. Cell Genet. **71,** 142–147.

Leach, C.R., Donald, T.M., Franks, T.K., Spiniello, S.S., Hanrahan, C.F., and Timmis, J.N. (1995). Organization and origin of a B chromosome centromeric sequence from *Brachycome dichromosomatica.* Chromosoma **103,** 708–714.

Lee, C., Wevrick, R., Fisher, R.B., Ferguson-Smith, M.A., and Lin, C.C. (1997). Human centromeric DNAs. Hum. Genet. **100,** 291–304.

Maluszynska, J., and Heslop-Harrison, J.S. (1991). Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana.* Plant J. **1,** 159–166.

Martinez-Zapater, J.M., Estelle, M.A., and Somerville, C.R. (1986). A highly repeated DNA sequence in *Arabidopsis thaliana.* Mol. Gen. Genet. **204,** 417–423.

Morgante, M., Jurman, I., Shi, L., Zhu, T., Keim, P., and Rafalski, J.A. (1997). The STR120 satellite DNA of soybean: Organization, evolution and chromosomal specificity. Chromosome Res. **5,** 363–373.

Murata, M., Ogura, Y., and Motoyoshi, F. (1994). Centromeric repetitive sequences in *Arabidopsis thaliana.* Jpn. J. Genet. **69,** 361–370.

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., and Okazaki, T. (1992). Centromere protein B assembles human centromeric α-satellite at the 17-bp sequence, CENP-B box. J. Cell Biol. **166,** 585–596.

Nagaki, K., Tsujimoto, H., and Sasakuma, T. (1998). A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. Chromosome Res. **6,** 295–302.

Pelissier, T., Tutois, S., Tourmente, S., Deragon, J.M., and Picard, G. (1996). DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. Genetica **97,** 141–151.

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

Scheid, O.M., Afsar, K., and Paszkowski, J. (1994). Gene inactivation in *Arabidopsis thaliana* is not accompanied by an accumulation of repeat-induced point mutations. Mol. Gen. Genet. **244,** 325–330.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D., and Dean, C. (1995). Physical map and organization of *Arabidopsis thaliana* chromosome. Science **270,** 480–483.

Schmidt, T., and Heslop-Harrison, J.S. (1996). High resolution mapping of repetitive DNA by *in situ* hybridization: Molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens.* Plant Mol. Biol. **30,** 1099–1114.

Schwarzacher, T., Wang, M.L., Leitch, A.R., Miller, N., Moore, G., and Heslop-Harrison, J.S. (1997). Flow cytometric analysis of the chromosomes and stability of a wheat cell-culture line. Theor. Appl. Genet. **94,** 91–97.

Selker, E.U., Fritz, D.Y., and Singer, M.J. (1993). Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora.* Science **262,** 1724–1728.

Simoens, C.R., Gielen, J., Van Montagu, M., and Inzé, D. (1988). Characterization of highly repetitive sequences of *Arabidopsis thaliana.* Nucleic Acids Res. **16,** 6753–6766.

Steinberg, R.A., and Gorman, K.B. (1992). Linked spontaneous CG$\rightarrow$TA mutations at CpG sites in the gene for protein kinase regulatory subunit. Mol. Cell. Biol. **12,** 767–772.

Sugimoto, K., Shibata, A., and Himeno, M. (1998). Nucleotide specificity at the boundary and size requirement of the target sites recognized by human centromere protein B (CENP-B) *in vitro.* Chromosome Res. **6,** 133–140.

Sunkel, C.E., and Coelho, P.A. (1995). The elusive centromere: Sequence divergence and functional conservation. Curr. Opin. Genet. Dev. **5,** 756–767.

Tyler-Smith, C., Corish, P., and Burns, E. (1998). Neocentromeres, the Y chromosome and centromere evolution. Chromosome Res. **6,** 65–71.

Uzawa, S., and Yanagida, M. (1992). Visualization of centromeric and nucleolar DNA in fission yeast by fluorescence in situ hybridization. J. Cell Sci. **101,** 267–275.

Vershinin, A., Schwarzacher, T., and Heslop-Harrison, J.S. (1995). The large scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. Plant Cell **7,** 1823–1833.

Willard, H.F. (1985). Chromosome-specific organization of human α satellite DNA. Am. J. Hum. Genet. **37,** 524–532.

Wilmen, A., Pick, H., Niedenthal, R.K., Sen-Gupta, M., and Hegemann, J.H. (1994). The yeast centromere CDEI/Cpf1 complex: Differences between in vitro binding and in vivo function. Nucleic Acids Res. **22,** 2791–2800.

Zachgo, E.A., Wang, M.L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolan, M., and Goodman, H.M. (1996). A physical map of chromosome 2 of *Arabidopsis thaliana.* Genome Res. **6,** 19–25.